

7^e Colloque National BRG –
STRASBOURG

LES RESSOURCES GÉNÉTIQUES
À L'HEURE DES GÉNOMES

*CHARACTERISING GENETIC RESOURCES
IN THE GENOMIC ERA*

Les Actes du



n°7

Comité de lecture (*Editorial board*)

Sigrid AUBERT, Amadou BA, Philippe BARET, Julien BERTHAUD, Philippe BERTIN, Christian BIEMONT, Elisabeth BLESBOIS, François BONHOMME, Pierre BOUDRY, Pierre CAPY, Serge CASAREGOLA, André CHARRIER, Eléonore CHARVOLIN, Yuna CHIFFOLEAU, Jean-Marie CORNUET, Grégory DECHAMP-GUILLAUME, Johann DETILLEUX, Christine DILLMAN, Françoise FRIDLANSKY, Gustavo GANDINI, Sylvain GLÉMIN, Françoise GREHAND, Eugénie HEBRARD, Sophie HUBERT, Louise JOUANIN, Frédéric LANTIER, Frédérique LE ROUX, Marc-Henry LEBRUN, François LEFÈVRE, Nathalie MACHON, Thierry NOËL, Jean-Louis PHAM, Daniel PRAT, Edwige QUILLET, Pierre-Yves RESCAN, Juliette RIQUET, Sylvain SANTONI, Pierre SAUMITOU-LAPRADE, Jean-Luc SOUCIET, Hélène TORDJAMN, Michel TROMMETTER, Étienne VERRIER.

Comité éditorial (*Editorial secretary*)

Eléonore CHARVOLIN, Françoise FRIDLANSKY, Frédérique MARIE et Armelle SAÏHI.

Mise en page

Frédérique MARIE.

Remerciements

Les travaux ont été réalisés avec le soutien des ministères en charge de la Recherche, de l'Agriculture et de l'Écologie et du Développement Durable, de l'INRA, de l'IRD et du Cirad.

Le colloque a bénéficié du concours financier de la Mairie et Communauté Urbaine de Strasbourg, de l'Université Louis Pasteur-Strasbourg I, d'Alsace BioValley, de l'INRA de Colmar, de la Région Alsace, du Conseil Général du Bas-Rhin, de l'INRA [Direction « Plante et Produits du Végétal », Départements de Génétique animale, de Microbiologie, de Génétique et Amélioration des Plantes], du Cirad, de l'IRD, du Ministère de l'Écologie et du Développement Durable.

Nous remercions également l'Université Louis Pasteur à Strasbourg pour son accueil et son organisation.

Enfin, un grand merci à Françoise FRIDLANSKY pour son aide précieuse.

SOMMAIRE

Introduction.....	9
I – Restitution de l'appel à propositions.....	13
Test de la performance des gènes en copie unique pour obtenir des phylogénies robustes : exemple en utilisant les génomes complets de champignons <i>Gabriela Aguilera, Sylvain Marthey, Hélène Chiapello, Marc-Henri Lebrun, François Rodolphe, Elisabeth Fournier, Annie Gendrault-Jacquemard, Tatiana Giraud.....</i>	15
Modélisation multi-agents des réseaux d'échanges de semences pour la conservation de la biodiversité agricole <i>Didier Bazile, Géraldine Abrami, Souleymane Dembele, Harouna Coulibaly, Christophe Le Page, Mathieu Dionnet, Jacques Chantereau, Céline Boué, Mattea Orsini, François Bousquet, Jean-Louis Pham, Kadiatou Sangare, Gilles Bezançon</i>	29
Analyse comparative du polymorphisme des gènes <i>eIF4E</i> chez les plantes : conservation <i>versus</i> sélection positive exercée par les virus <i>Carole Caranta, Jacques David, Carine Charron, Stéphane De Mita, Pascale Sanchez, Sylvain Santoni, Alberto Cenci, Joëlle Ronfort, Audrey Weber, Anne-Françoise Adam-Blondon, Dominique Brunel, Nathalie Chantret</i>	51
Diversité et évolution d'une famille multigénique impliquée dans la biosynthèse de l'amidon chez les graminées, et cas particulier de la domestication du maïs <i>Jonathan Corbi, Marilyne Debien, Agnès Rousselet, Martine Le Guilloux, Jean-Louis Prioul, Catherine Damerval, Maud Tenaillon, Domenica Manicacci</i>	65
Diversité et évolution des déshalogénases bactériennes : détection bioinformatique et perspectives de recherche <i>Maude David, Françoise Bringel, Marco Pagni, Niamb Gilmartin, Hasna Boubakri, Thierry Nadalig, Pascal Simonet, Timothy Vogel, Stéphane Vuilleumier</i>	83
Étude de QTL liés à la trypanotolérance bovine dans une population métisse ouest-africaine par analyse d'association <i>Guignighaza Kossigan Dayo, Sophie Thevenon, Jean-Paul Poivey, Issa Sidibe, Zakaria Bengaly, André Eggen, Didier Boichard, Mathieu Gauthier</i>	95
Étude des complémentarités entre gestion dynamique à la ferme et gestion statique en collection : cas de la variété de blé Rouge de Bordeaux <i>Elise Demeulenaere, Christophe Bonneuil, François Balfourier, Alain Basson, Jean-François Berthelot, Vincent Chesneau, Henri Ferté, Nathalie Galic, Guy Kastler, Jean Koenig, Florent Mercier, Joël Payement, Alain Pommart, Bernard Ronot, Yves Rousselle, Nicolas Supiot, Hélène Zabaria, Isabelle Goldringer</i>	117

Diversité agricole et patrimoine dans le moyen Rio Negro (Amazonie brésilienne)	
<i>Laure Emperaire, Pascale de Robert, Juliana Santilli, Ludvine Eloy, Lúcia van Velthem, Esther Katz, Cláudia Lopez, Anne-Elisabeth Laques, Manuela Carneiro Da Cunha, Mauro Almeida</i>	139
Détection par EcoTILLING du polymorphisme nucléotidique de gènes de résistance chez le peuplier	
<i>Patricia Faivre-Rampant, Aloïs Bresson, Isabella Paolucci, Erwan Cailleux, Marion Dalmais, Florence PIRON, Arnaud Dowkin, Vanina Guerin, Abdelhafid Bendhamane, Catherine Bastien, Véronique Jorge</i>	155
Apports du séquençage multiloci à la phylogénie et à la taxonomie de deux genres majeurs de bactéries phytopathogènes : <i>Pseudomonas</i> et <i>Xanthomonas</i>	
<i>Marion Fischer-Le Saux, Olivier Pruvost, Emilie Fargier, Nathalie Ab-You, Sophie Bonneau, Carolina Gonzalez, Lionel Gagnevin, Valérie Verdier, Charles Manceau</i>	171
Diversité nucléotidique pour le gène <i>PHANTASTICA</i> (<i>PHAN</i>) le long d'un gradient altitudinal chez le frêne commun (<i>Fraxinus excelsior</i> L.) et chez le frêne oxyphylle (<i>Fraxinus angustifolia</i> Valh)	
<i>Nathalie Frascaria-Lacoste, Juan Fernandez-Manjarres, Joly Basak, Pierre Gérard, Paola Bertolino, Jean Dufour, Corinne Cruaud, Arnaud Couloux, Sophie Massot</i>	187
Analyse et exploitation de la diversité génétique des polykétides synthases de type I dans l'ADN metagénomique d'un sol	
<i>Aurélien Ginolbac, Sandrine Demaneche, Cyrille Jarrin, Patrick Robe, Fabrice Lefèvre, Renaud Nalin, Guy Perrière, Timothy M. Vogel, Pascal Simonet</i>	201
Un modèle de variabilité fonctionnelle chez les arbres forestiers : le gène <i>CCR</i> d'eucalyptus	
<i>Jean-Marc Gion, Frédéric Mortier, Eric Mandrou, Paulo Ricardo Hein Gherardi, Tristan Costecalde, Gilles Chaix, Marie Étienne, Pierre Sivadon, Jacqueline Grima-Pettenati, Emilie Villar, Aubin Saya, Brigitte Pollet, Catherine Lapierre, Philippe Vigneron</i>	215
Analyse du polymorphisme des α 1,2fucosyltransférases chez le lapin de garenne et recherche d'une relation avec la sensibilité à la maladie hémorragique virale	
<i>Patrice Guillon, Nathalie Ruvoën-Clouet, Béatrice Le Moullac-Vaidye, Stéphane Marchandean, Jacques Le Pendu</i>	239
Polymorphisme de gènes impliqués dans l'acquisition et la gestion de l'énergie chez l'huître creuse <i>Crassostrea gigas</i> : relations avec des caractères d'intérêt aquacole	
<i>Arnaud Huvet, Dario Moraga, Elise David, Anne-Leila Meistertzheim, Arnaud Tanguy, Alain Van Wormboudt, Jeanne Moal, Fanny Jeffroy, Christopher Sawage, Sylvie Lapègue, Pierre Boudry</i>	253

Spécialisation de la tique des oiseaux marins et diversité des bactéries du complexe <i>Borrelia burgdorferi sensu lato</i> , agents de la maladie de Lyme: effets en cascade dans les systèmes à vecteur <i>Karen D. McCoy, David Duneau, Thierry Boulmier</i>	277
Tragédie des anti-communautaires et gestion collective dans les biotechnologies <i>Yann Ménière, Michel Trommetter, Henry Feyt, Catherine Potvin</i>	293
Analyse de séquences multi-loci (MLSA) dans le genre <i>Bradyrhizobium</i> <i>Angèle N'Zoué, Gisèle Laguerre, Farida Boulila, Chinna swamy Appunu, Nourredine Sebbane, Raoul Rivas, Anne Willems, Lionel Moulin, Philippe de Lajudie</i>	319
Caractérisation de populations à partir de l'analyse de régions chromosomiques potentiellement sélectionnées en comparaison à la diversité neutre : étude conjointe de lignées expérimentales de poules et des quatre sous-populations ovines de race Lacaune <i>Isabelle Palhière, Valérie Loynnyck, Magali Sancristobal, Denis Laloé, Yves Amigues, Bertrand Bed'hom, Bernard Bibé, Pieter Bijma, Loys Bodin, Daïa Boudarene, Claude Chevalet, Sophie Leroux, Katayoun Moazami-Gondarzi, Philippe Mulsant, Frédérique Pitel, Bertrand Servin, Etienne Verrier, Florence Vignoles</i>	335
Démasquage des gènes spécifiques d'une espèce génomique du complexe <i>Agrobacterium tumefaciens</i> par AFLP et multicapteur à ADN <i>Perrine Portier, Denis Costechareyre, Daniel Muller, David Chapulliot, Christine Oger, Guy Perrière, Laurent Guéguen, Vincent Daubin, Xavier Nesme</i>	351
Identification des gènes candidats de la tolérance au zinc chez la pseudométallophyte <i>Arabidopsis halleri</i> par l'intégration des données de génomique comparative et de transcriptomique disponibles chez <i>Arabidopsis</i> <i>Nancy Roosens, Zaigham Shahzad, Eric Lacombe, Hélène Frérot, Anne Creach, Pierre Berthomieu, Pierre Saumitou-Laprade</i>	373
La Cryobanque Nationale Lapin : un outil de gestion dynamique des ressources génétiques cynicoles <i>Pascal Salvetti, Daniel Allain, Gérard Bolet, Jacques Hurtaud, Samuel Boucher, Thierry Joly</i>	391
Déperdition ancienne et restauration moderne de la diversité génétique du blé <i>Anne-Céline Thuillet, Pierre Roumet, Amandine Bordat, Christine Tollon, Gérard Poux, Sylvain Santoni, Jacques David</i>	405
Interactions génomiques en régime de sélection et taille de population <i>Michel Veuille, Lionel Brazier, Florent Meli</i>	423

Contribution des éléments transposables à la création de diversité et à l'adaptation du maïs - une étude pilote <i>Tatiana Zerjal, Christine Dillmann, Alain Charcosset, Karine Alix, Corinne Mhiri, Marie-Angèle Grandbastien, Maud Tenaillon</i>	439
II – Communication en séances plénières.....	457
Le séquençage du génome de la vigne dans un projet public franco-italien <i>A-F Adam-Blondon</i>	459
Diversité et homogénéité génétique de l'espèce canine : potentiel en génétique médicale <i>Catherine André, Christophe Hitte</i>	463
Analyse fonctionnelle de micro-organismes et de communautés complexes par des approches (méta)protéomiques <i>Florence Arsène-Ploetz, Florence Goulben-Chollet, Bertrand Chaumande, Stéphanie Weiss, Christine Carapito, Jessica Cleiss, Sébastien Gallien, Christine Schaeffer, Alain Van Dorselaer, Philippe N Bertin</i>	469
Multiplicité des facteurs conduisant à un choix de conservation <i>François Boulineau</i>	475
Qui était LUCA? Portrait du Last Universal Common Ancestor <i>Céline Brochier</i>	477
Le mildiou de la vigne : diversité génétique, introductions et durabilité des résistances <i>François Delmotte, Pere Mestre</i>	485
La gestion des informations relatives aux races à petits effectifs par les systèmes nationaux d'information génétique (SNIG) : contribution et problèmes rencontrés <i>C. Dos, C. Danchin-Burge</i>	487
Siregal, un système d'information multi-espèces sur les ressources génétiques végétales françaises <i>Sophie Durand</i>	495
Modèles économiques de la valeur d'une ressource génétique <i>A.Fadlaoui, P. Baret,</i>	499
La connaissance du processus de domestication nous éclaire t-elle sur les choix de conservation ? <i>Paul Gepts</i>	501

Mesure et impact de la stratification génétique intra et inter races chez le chien <i>Christophe Hitte, Catherine André</i>	503
La place des projets territoriaux dans les prises de décision sur le devenir des races animales locales <i>Adeline Lambert-Derkimba, François Casabianca, Etienne Verrier</i>	509
Caractérisation et utilisation de la diversité génétique de la vigne <i>L. Le Cunff, M. Di Vecchi Staraz, R. Bacilieri, S. Nicolas, V. Laucon, T. Lacombe, A. Doligez, A-F Adam-Blondon, J-M Boursiquot, P. This</i>	517
Métagénomique et diversité génétique et métabolique <i>Denis Le Paslier</i>	519
Whole genome sequencing of the fungal plant pathogens <i>Botrytis cinerea</i> and <i>Sclerotinia sclerotiorum</i> <i>Marc-Henri Lebrun</i>	521
Diversité moléculaire et structure du génome chez les levures <i>S. cerevisiae</i> œnologiques <i>JL Legras, B. Blondin, S. Casaregola, S. Dequin</i>	523
Une information hors norme : les savoirs traditionnels <i>Anne Luxereau</i>	525
Génomique et évolution des Actinobactéries <i>Normand P, Alloisio N, Pujic P, Vallenet D, Medigue C, Berry A, Santos C, Tavares F, Daubin V</i>	527
De la collecte de données à l'information pour les décideurs : la production d'indicateurs de biodiversité <i>Dominique Richard, Eléonore Charvolin</i>	529
La mesure de la diversité génétique : quels outils, quelles méthodes, quel futurs ? <i>Joëlle Ronfort</i>	531
Nouveaux outils moléculaires pour décrire la biodiversité <i>Pierre Taberlet</i>	533
Les développements technologiques dans l'analyse des génomes <i>Patrick Wincker</i>	535

CONTENTS

Introduction.....	9
I –.....	13
Assessing the performance of single-copy orthologs for recovering robust fungal phylogenies <i>Gabriela Aguilera, Sylvain Marthey, Hélène Chiapello, Marc-Henri Lebrun, François Rodolphe, Elisabeth Fournier, Annie Gendraud-Jacquemard, Tatiana Giraud.....</i>	15
Modelling the seed system with agent-based models <i>Didier Bazile, Géraldine Abrami, Souleymane Dembele, Harouna Coulibaly, Christophe Le Page, Mathieu Dionnet, Jacques Chantereau, Céline Boué, Mattea Orsini, François Bousquet, Jean-Louis Pham, Kadiatou Sangare, Gilles Bezançon</i>	29
Comparative analysis of the DNA polymorphism at <i>eIF4E</i> genes in plants: conservation <i>versus</i> positive selection exerted by Potyviruses <i>Carole Caranta, Jacques David, Carine Charron, Stéphane De Mita, Pascale Sanchez, Sylvain Santoni, Alberto Cenci, Joëlle Ronfort, Audrey Weber, Anne-Françoise Adam-Blondon, Dominique Brunel, Nathalie Chantret</i>	51
Diversity and evolution of a multigenic family involved in starch biosynthesis in grasses, and the case of maize domestication <i>Jonathan Corbi, Marilyne Debieu, Agnès Rousselet, Martine Le Guilloux, Jean-Louis Prioul, Catherine Damerval, Maud Tenaillon, Domenica Manicacci</i>	65
Diversity and evolution of bacterial dehalogenases: Detection by bioinformatic tools and research perspectives <i>Maude David, Françoise Bringel, Marco Pagni, Niamb Gilmartin, Hasna Boubakri, Thierry Nadalig, Pascal Simonet, Timothy Vogel, Stéphane Vuilleumier</i>	83
QTL linked to bovine trypanotolerance in West African crossbred population <i>Guiguighaza Kossigan Dayo, Sophie Thevenon, Jean-Paul Poivey, Issa Sidibe, Zakaria Bengaly, André Eggen, Didier Boichard, Mathieu Gauthier</i>	95
Complementary aspects of dynamic management on farm and static conservation in genebank based on a case study: the wheat variety 'Rouge de Bordeaux' <i>Elise Demeulenaere, Christophe Bonneuil, François Balfourier, Alain Basson, Jean-François Berthelot, Vincent Chesneau, Henri Ferté, Nathalie Galic, Guy Kastler, Jean Koenig, Florent Mercier, Joël Payement, Alain Pommart, Bernard Ronot, Yves Rousselle, Nicolas Supiot, Hélène Zabaria, Isabelle Goldringer</i>	117

Agriculture diversity and heritage in the mid-Rio Negro region (Brazilian Amazon)	
<i>Laure Emperaire, Pascale de Robert, Juliana Santilli, Ludvine Eloy, Lúcia van Velthem, Esther Katz, Cláudia Lopez, Anne-Elisabeth Laques, Manuela Carneiro Da Cunha, Mauro Almeida</i>	139
EcoTILLING for identification of allelic variation in poplar resistance genes	
<i>Patricia Faivre-Rampant, Aloïs Bresson, Isabella Paolucci, Erwan Cailloux, Marion Dalmais, Florence PIRON, Arnaud Dowkim, Vanina Guerin, Abdelhafid Bendhamane, Catherine Bastien, Véronique Jorge</i>	155
Contribution of Multilocus sequence analysis to the phylogeny and taxonomy of two major groups of plant pathogenic bacteria : <i>Xanthomonas</i> and <i>Pseudomonas</i>	
<i>Marion Fischer-Le Saux, Olivier Pruvost, Emilie Fargier, Nathalie Ab-You, Sophie Bonneau, Carolina Gonzalez, Lionel Gagnevin, Valérie Verdier, Charles Manceau</i>	171
Nucleotide diversity at the <i>PHANTASTICA</i> gene (<i>PHAN</i>) along an altitudinal gradient for common ash (<i>Fraxinus excelsior</i> L.) and narrow-leaved ash (<i>Fraxinus angustifolia</i> Valh.)	
<i>Nathalie Frascaria-Lacoste, Juan Fernandez-Manjarres, Jolly Basak, Pierre Gérard, Paola Bertolino, Jean Dufour, Corinne Cruaud, Arnaud Couloux, Sophie Massot</i>	187
Analysis and exploitation of type I polyketide synthase genetic diversity in soil metagenomic DNA	
<i>Aurélien Ginolbac, Sandrine Demaneche, Cyrille Jarrin, Patrick Robe, Fabrice Lefèvre, Renaud Nalin, Guy Perrière, Timothy M. Vogel, Pascal Simonet</i>	201
CCR gene in <i>Eucalyptus</i> : a model of functional variability in forest trees	
<i>Jean-Marc Gion, Frédéric Mortier, Eric Mandrou, Paulo Ricardo Hein Gherardi, Tristan Costecalde, Gilles Chaix, Marie Étienne, Pierre Sivadon, Jacqueline Grima-Pettenati, Emilie Villar, Aubin Saya, Brigitte Pollet, Catherine Lapierre, Philippe Vigneron</i>	215
Analysis of wild rabbit α 1,2fucosyltransferases polymorphism and search of a relationship with sensitivity to viral hemorrhagic disease	
<i>Patrice Guillon, Nathalie Ruvoën-Clouet, Béatrice Le Moullac-Vaidye, Stéphane Marchandeanu, Jacques Le Pendu</i>	239
Polymorphism of nutrition and energy metabolism related-genes in the cupped oyster <i>Crassostrea gigas</i> : implication in important traits for aquaculture	
<i>Arnaud Huvet, Dario Moraga, Elise David, Anne-Leila Meistertzheim, Arnaud Tanguy, Alain Van Wormboudt, Jeanne Moal, Fanny Jeffroy, Christopher Sauvage, Sylvie Lapègue, Pierre Boudry</i>	253
Specialisation of the seabird tick and diversity of Lyme borreliosis bacteria: cascading host effects in vector-borne systems	
<i>Karen D. McCoy, David Duneau, Thierry Boulmier</i>	277

The tragedy of anticommons and collective mangement in plant biotechnology <i>Yann Ménière, Michel Trommetter, Henry Feyt, Catherine Potvin</i>	293
Multilocus sequence analysis in <i>Bradyrhizobium</i> <i>Angèle N'Zoué, Gisèle Laguerre, Farida Boulila, Chinmaswamy Appunu, Nourredine Sebbane, Raoul Rivas, Anne Willems, Lionel Moulin, Philippe de Lajudie</i>	319
Characterization of populations by detecting genomic regions supposed to be under selection, compared to the neutral diversity: example of chicken experimental lines and the four breeding populations in Lacaune sheep breed <i>Isabelle Palbière, Valérie Løynyck, Magali Sancristobal, Denis Laloé, Yves Amigues, Bertrand Bed'hom, Bernard Bibé, Piter Bijma, Loys Bodin, Daïa Boudarene, Claude Chevalet, Sophie Leroux, Katayoun Moazami-Goudarzi, Philippe Mulsant, Frédérique Pitel, Bertrand Servin, Etienne Verrier, Florence Vignoles</i>	335
Unmasking species specific genes in <i>Agrobacterium tumefaciens</i> species G8 by AFLP and microarray <i>Perrine Portier, Denis Costechareyre, Daniel Muller, David Chapulliot, Christine Oger, Guy Perrière, Laurent Guéguen, Vincent Daubin, Xavier Nesme</i>	351
Genetic Zn tolerance in <i>Arabidopsis thaliana</i> <i>Nancy Roosens, Zaigham Shahzad, Eric Lacombe, Hélène Frérot, Anne Creach, Pierre Berthomieu, Pierre Saumitou-Laprade</i>	373
The French Rabbit Cryobank: a tool for management of rabbit genetic resources <i>Pascal Salvetti, Daniel Allain, Gérard Bolet, Jacques Hurtaud, Samuel Boucher, Thierry Joly</i> ..	391
Loss and restoration of wheat genetic diversity <i>Anne-Céline Thuillet, Pierre Roumet, Amandine Bordat, Christine Tollon, Gérard Poux, Sylvain Santoni, Jacques David</i>	405
Genomic interactions under selection regime <i>Michel Veuille, Lionel Brazier, Florent Meli</i>	423
How did transposable elements contribute to maize diversity and adaptation? – a pilot study <i>Tatiana Zerjal, Christine Dillmann, Alain Charcosset, Karine Alix, Corinne Mhiri, Marie-Angèle Grandbastien, Maud Tenaillon</i>	439

Introduction

Au cours de la vie d'un homme ou d'une femme, il y a des moments de joie et des moments de peine. Au cours de la vie, d'un ou d'une responsable, il y a des moments pour l'action et d'autres pour la réflexion. Lors de la réalisation d'un programme de recherche, il y a des périodes fructueuses de certitude et d'autres toutes aussi positives d'inquiétude. Si le doute torture, la conviction parfois égare. Dans tous les cas, il y aura un espace de décision. Aussi paradoxal que cela puisse paraître, cet espace est ouvert à tous. Il ne tient compte ni du chemin emprunté, ni de votre origine, de votre expérience ou de votre savoir. Vous participez à la décision et le rôle de chacun est primordial.

Si vous vous en souvenez, lors de notre colloque de Lyon en 2004, j'avais osé vous parler d'un petit livre d'Italo Calvino, les "Villes invisibles" et de la force de l'imagination et de la concertation. A La Rochelle, en 2006, avec les "p'tits vélos jaunes", je reprenais un thème qui m'est très cher : celui de l'ouverture des ressources génétiques vers des domaines plus larges. Face aux risques pris, il y aura toujours une solution nouvelle et, sans doute, encore plus enrichissante. Dans l'angle d'une image, surgira toujours une petite bicyclette qui apportera outre la surprise, la réflexion et la hauteur nécessaire à une nouvelle dynamique et à de nouveaux progrès. Etre dans la réalité n'est pas faire preuve d'un optimisme béat. Nous sommes dans une période de désordre. C'est une énorme force d'ouvertures et d'opportunités.

Deux problématiques ont émergé ces dernières années au sein de la communauté des ressources génétiques dont chacun des acteurs de terrain devine l'importance des enjeux et implications. D'une part, les problèmes pratiques, interrogations et réflexions se multiplient sur les droits, obligations et responsabilités qui s'attachent aux différentes formes du végétal, de l'animal ou du micro-organisme. Comment considérer les règles qui s'appliquent au matériel génétique selon qu'il est "produit" (grain, animal ou ferment), "matériel de reproduction" (semence, sperme, embryon ou la culture) ou "ressource génétique" (variété, race ou souche) ?

La complexité des relations homme, ressources et environnement impose d'autre part, une réflexion au plus profond de la ressource. Les initiatives foisonnent, notamment dans les champs de la génomique et des ressources biologiques, mais l'opportunité individuelle prime souvent sur la cohérence nationale, au risque de fragiliser la démarche collective de gestion décentralisée des Collections nationales. Les pouvoirs publics peuvent-ils se passer de coordination entre ces divers niveaux ?

Pour répondre à ces problématiques, dans le domaine des ressources génétiques animales, les groupes pilotes forment un ensemble opérationnel qui

peut répondre à de multiples sollicitations. Ils peuvent, par leurs activités, répondre aux multiples demandes provenant des acteurs de terrain et favoriser ainsi l'émergence de nouveaux groupes pilotes (chien, cheval, abeille, poissons) et d'une nouvelle approche des espèces sauvages, élevées et relâchées (poissons, gibiers, espèces invasives). Cette posture a déjà permis d'aborder prioritairement une gestion intégrée des trois niveaux de la biodiversité.

Le développement d'un ensemble cohérent, aujourd'hui encore incomplet même si nous incluons le travail auprès de la Cryobanque Nationale, concourt à la mise en œuvre du Plan d'Action Mondial validé par la France en 2007. Plus directement, il permet de maintenir les engagements français pris auprès du réseau international de gestion des ressources zoogénétiques (ERFP), dont la France a assuré le fonctionnement pendant six années. Cette participation a été non seulement le relais à l'établissement de programmes de recherche européens mais aussi la source de nouvelles réflexions nécessaires à la mise en place concertée de règlements européens (sanitaires, agro-environnementaux, de gestion partagée).

Dans le domaine des ressources génétiques végétales, l'ensemble des concertations réalisées a débouché sur une politique d'organisation souple et décentralisée de la conservation par des réseaux associant, sur une base d'espèce ou de groupe d'espèces, des partenaires publics et privés, acceptant de mettre en commun, de maintenir, de caractériser et de diffuser une partie de leur matériel ne faisant plus l'objet d'une protection intellectuelle et qui n'est pas non plus engagé dans un processus de sélection. Ces Collections nationales portent plus spécifiquement sur un matériel diversifié et original, dont la responsabilité de conservation incombe à la France, dans l'idée d'une répartition des efforts à l'échelle européenne voire mondiale.

Cette approche dynamique des ressources phytogénétiques a permis de voir non seulement l'intégration de nouvelles espèces importantes, comme la pomme de terre mais aussi la justesse d'une démarche intégrant toutes les dimensions : réseaux et Collections nationales, bases de données, aspects juridiques et économiques relatifs à la gestion des ressources génétiques en France. Elle a permis d'entrevoir la mise en place d'une démarche qualité (en lien avec l'initiative CRB), d'une coopération européenne, d'une contribution à l'élaboration et à la défense d'une position française dans les négociations internationales cohérente avec l'action nationale et respectueuse des intérêts de l'ensemble des acteurs impliqués.

Au cours de ces dernières années, des efforts importants ont été entrepris pour fédérer par grands thèmes le monde des microorganismes (intérêt laitier, cidrerie, rhumerie, œnologie, panification, malterie, etc.). Les enjeux sont importants. Avec une participation active dans la mise en place des

Centres de Ressources Biologiques et l'établissement de normes de qualité, les responsables ressources microbiennes montrent le chemin des futurs développements pour l'ensemble des ressources. L'approche interrègles trouve alors sa pleine signification dans un domaine très demandeur de structuration et d'actions.

Pour les ressources génétiques, cette "gestion", terme souvent mal compris qui englobe l'inventaire, la caractérisation, l'utilisation, la valorisation, la conservation, la protection et l'encadrement, ne se résume pas à une simple liste de races, de variétés ou de souches mais à une intégration de son approche dans tous les processus de décision, et notamment dans ceux, oh combien importants, se situant tout au long des programmes de recherche.

Depuis 15 années, la totalité des 200 projets de recherche présentés au cours des 7 colloques nationaux "ressources génétiques" a permis de bien voir non seulement les avantages d'une intégration des diverses sciences impliquées dans la recherche mais également la nécessité de prendre en compte, au cours de ces programmes, en amont comme en aval de leur réalisation, la gestion des ressources génétiques. Il n'y aura pas de recherche en biologie sans collection de ressources génétiques. Il n'y aura pas de bon résultat en biologie sans apport significatif à la gestion des ressources génétiques. Dans ces deux cas, il n'y aura pas de progrès sans l'ambition de connaître la diversité génétique dans son ensemble et dans ses dynamiques.

C'est avec cette volonté que nous abordons ce colloque de Strasbourg. Les premiers appels à propositions de recherche étaient fortement orientés sur les méthodologies d'organisation et de conservation des ressources génétiques en laissant toutefois une large place aux inventaires de la diversité, notamment dans le domaine des micro-organismes. Progressivement, nous avons mis l'accent de ces appels sur des recherches touchant à l'étude des équilibres entre les conservations *in situ* et *ex situ*. Le thème relatif à la diversité de gène d'intérêt ou de fonctions connues focalisait toujours de nombreuses demandes. Depuis quatre années, nous avons cherché à attirer les projets relatifs à l'appropriation, l'accessibilité et l'échange des ressources et des informations. Cette approche a traduit notre volonté de développer la prise en compte des Sciences Humaines et Sociales, sans que cela soit exclusif de l'apport de ces disciplines dans les autres rubriques de appels à propositions. Enfin, nous avons voulu laisser place à la surprise et à l'imagination en soutenant des démarches scientifiques innovantes.

En conclusion, je reprendrai une phrase de notre document publié en juin 2006 – Les Ressources Génétiques à l'orée de temps nouveaux : "*les véritables ressources du futur sont l'aptitude du vivant à générer cette diversité génétique dans laquelle des tris nouveaux seront faits*". Les avancées de la biologie et le

développement des biotechnologies ouvrent de nouvelles perspectives en termes d'exploitation et de valorisation des ressources génétiques. Il est également clair que les systèmes de production évoluent sur la base d'un triptyque environnement – itinéraire technique – ressource génétique. Par ailleurs, le concept de ressource génétique est porteur de certaines dualités : production / gestion durable, système anthropisé / dynamique naturelle. Dans un contexte de changement global, il y a donc plus que jamais nécessité d'intégrer les dimensions biologique, écologique et socio-économiques qui soulève en elle-même de nouvelles questions de recherche.

Nous voici au cœur d'un problème qui se veut vaste et enthousiasmant. Nous sommes aujourd'hui à Strasbourg, au cœur de l'Europe. C'est un symbole pour un défi que la France ne pourra pas relever seule.

Dominique PLANCHENAULT

13 octobre 2008

I

RESTITUTION
DE L'APPEL À PROPOSITIONS

Test de la performance des gènes en copie unique pour obtenir des phylogénies robustes : exemple en utilisant les génomes complets de champignons

Gabriela AGUILETA⁽¹⁾, Sylvain MARTHEY⁽²⁾, Hélène CHIAPELLO⁽²⁾,
Marc-Henri LEBRUN⁽⁴⁾, François RODOLPHE⁽²⁾,
Elisabeth FOURNIER⁽³⁾, Annie GENDRAULT-JACQUEMARD⁽²⁾,
Tatiana GIRAUD^{(1)*}

⁽¹⁾UMR8079, Univ Paris-Sud/CNRS, Ecologie, Systématique et Evolution, ,
91405 Orsay cedex, France;

⁽²⁾INRA, Mathématiques, Informatique et Génomes, Bâtiment 233
Domaine de Vilvert 78350 Jouy en Josas cedex France

⁽³⁾UMR BGPI, TA A 54/K, Campus International de Baillarguet,
34398 Montpellier cedex 5, France

⁽⁴⁾UMR 5240 CNRS-UCB-INSA, Bayer Cropscience,
Microbiologie Adaptation Pathogénie, 14, rue Pierre Baizet, 69009 Lyon, France

Abstract: Assessing the performance of single-copy orthologs for recovering robust fungal phylogenies. Fungal phylogenies involving non-model species are based on a few genes, mostly chosen using historical or practical criteria. Because gene trees are sometimes incongruent with species trees, the resulting phylogenies may not accurately reflect the evolutionary relationships among species. The increase in fungal genome sequences now provides large numbers of genes that could be used for building phylogenies. However, for practical reasons only a few genes can be sequenced in a wide range of species. Here, we asked how many and which genes, among the single-copy genes common to most fungal genomes, were required to recover an accurate and well-supported phylogeny. An automated procedure was developed to extract single-copy orthologous genes from complete fungal genomes using a Markov Clustering Algorithm (MCL-Tribe). After the analysis of 21 complete, publicly available fungal genomes reliably annotated, 246 single-copy orthologous gene clusters were identified. A website called FunyBase is available with open access to the clusters of orthologous sequences.

We then inferred maximum likelihood gene trees using the individual aligned orthologous sequences and constructed a reference tree from the concatenated alignments. The same topology as the reference tree was recovered for two genes, individually, that together provided high bootstrap values. The performance of individual genes was highly variable. Both gene size and the total number of variable

* Correspondance et tirés à part : Tatiana.Giraud@u-psud.fr

sites were highly correlated and were good predictors of the phylogenetic performance of individual genes. All the genes longer than 700 bp yielded good topological scores. However, the two best performing genes had intermediate gene sizes and there was a great variability in the phylogenetic performance for a given gene size, indicating that other factors have an impact on the phylogenetic performance of single genes. The categories of the putative functions of the best performing genes according to the Gene Ontology were significantly different from those of the 246 clusters of orthologs. Noteworthy, the proteins most frequently used in fungal phylogenies, such as tubulins and elongation factor EF1 were not found in the list of the best performing genes.

We expect that the best-performing genes identified here should be very useful for phylogenetic studies in fungi, at least at a large taxonomic scale, and allow lowering sequencing costs, as well as improving phylogenetic accuracy. The approach developed here will be further useful for other taxonomic groups when more complete genomes become available. The genes we ranked as best performing may not be suited for lower taxonomic scales. For lower-level phylogenetics, the FunyBase allows rapidly mining single-copy genes present in all fungal genomes, with given level of variability, and will therefore be useful for finding suitable genes at the desired taxonomic scale.

Paralogs/ orthologs/ Ascomycota/ Basidiomycota/ systematics

Résumé : Les phylogénies moléculaires sont généralement basées sur les séquences de quelques gènes. Cependant, les phylogénies basées sur des gènes uniques ne reflètent pas toujours bien les relations de parenté entre espèces. La disponibilité croissante en génomes complets de champignons fournit de nombreux gènes potentiellement utilisables pour des phylogénies, dont seulement quelques uns peuvent cependant être séquencés chez un grand nombre d'espèces. Nous avons recherché quels étaient les meilleurs gènes pour construire des phylogénies fiables chez les champignons. Nous avons développé une procédure automatisée pour rechercher les gènes en copie unique et communs à tous les génomes fongiques complets publiés. A partir de 21 génomes fongiques complets et bien annotés, nous avons extrait 246 groupes d'orthologues en copie unique, qui sont disponibles dans une base de données en ligne, FunyBase. Les phylogénies de chaque gène ont ensuite été inférées et comparées à l'arbre inféré en utilisant tous les orthologues concaténés. Deux gènes donnaient une topologie complètement identique à celle de l'arbre de référence, et avec des supports très forts quand ils étaient concaténés. Ces gènes devraient donc être très utiles pour réaliser des phylogénies fiables chez les champignons. L'approche développée ici devrait être intéressante pour d'autres groupes taxonomiques quand davantage de génomes seront disponibles.

Paralogues/ orthologues/ Ascomycota/ Basidiomycota/ systématique

1. INTRODUCTION

Les champignons constituent un des règnes principaux de la diversité eukaryote. Approximativement 80 000 espèces ont été décrites mais le nombre réel d'espèces a été estimé à 1,5 million [10]. Les champignons jouent des rôles écologiques essentiels dans pratiquement tous les écosystèmes, par leurs rôles de saprophytes, pathogènes, mutualistes et symbiotes. Les champignons ont également un fort impact sur les activités humaines, impliquant les microbes pathogènes d'humains et de cultures, et les espèces utilisées dans la transformation des produits alimentaires et les biotechnologies. Les deux groupes principaux qui ont été traditionnellement identifiés parmi les champignons vrais sont les Ascomycètes, incluant les levures et les champignons filamenteux, avec plusieurs espèces modèles importantes (*Neurospora crassa* et *Saccharomyces cerevisiae* par exemple) et les Basidiomycètes, incluant les champignons comestibles et les rouilles. Jusqu'à présent, des gènes employés pour établir des phylogénies de champignons étaient choisis sur la base de critères historiques ou pratiques, en utilisant les mêmes gènes que dans des études précédentes d'autres champignons ou les gènes qui étaient disponibles dans les ensembles d'espèces considérés. Cependant, les arbres construits à partir de gènes uniques ne reflètent pas toujours parfaitement les relations de parenté entre les espèces. Généralement, l'utilisation de plusieurs gènes pour des phylogénies augmente leur résolution et leur robustesse. La disponibilité récente de nombreux génomes complets de champignons fournit un grand nombre de gènes qui pourraient être employés pour des phylogénies robustes. De fait, des génomes fongiques complets ont été employés avec succès pour établir des phylogénies bien supportées [7], [19], [20]. Cependant, pour reconstruire des phylogénies d'espèces fongiques dont le génome complet n'est pas séquencé, seul un nombre limité de fragments d'ADN peut pratiquement être séquencé. Il serait donc utile de déterminer si certains gènes ont une meilleure puissance phylogénétique que d'autres, et combien sont nécessaires pour obtenir une phylogénie fiable.

Le but de cette étude était donc d'évaluer quels gènes permettaient d'établir les phylogénies fongiques les plus fiables et robustes parmi les orthologues en copie unique qui étaient communs à la plupart des génomes fongiques, et combien de gènes étaient nécessaires pour obtenir une phylogénie précise et bien soutenue.

2. MATÉRIELS ET MÉTHODES

2.1. Recherche d'orthologues et génomes utilisés

Nous avons développé une procédure automatisée pour rechercher les gènes en copie unique et communs à tous les génomes fongiques complets publiés, d'après la méthode décrite par Dujon *et al.*, [6]. Nous avons utilisé un algorithme de regroupement de Markov (MCL) pour rechercher les groupes d'orthologues communs à 30 génomes fongiques complets (tabl. I). Les mauvaises annotations de certains génomes faisaient chuter fortement le nombre de groupes d'orthologues communs, à cause de protéines prédites comme deux protéines distinctes, ou à cause d'introns oubliés. Nous n'avons donc gardé finalement que les 23 génomes fongiques les mieux annotés (tabl. I). En outre, parce que nous avons besoin d'un arbre de référence fiable pour effectuer les comparaisons topologiques, nous avons enlevé deux espèces, *Aspergillus oryzae* et *Stagonospora nodorum*, car leurs positions respectifs dans la phylogénie étaient faiblement soutenus, même avec les données des génomes complets. Les comparaisons de topologie d'arbre décrites ci-dessous ont été donc effectuées en utilisant 21 génomes (tabl. I).

2.2. Phylogénies et comparaisons de topologies

Les alignements ont été réalisés avec ClustalW [24] sur la totalité des ORF. Les régions où l'alignement était ambigu ont été enlevées grâce au programme Gblocks [4]. Le logiciel ProTest [1] a été utilisé pour déterminer quel était le modèle de substitution d'acides aminés qui convenait le mieux pour chaque alignement d'orthologues, et les phylogénies de chaque gène ont été inférées par maximum de vraisemblance à l'aide du logiciel PHYML [9] en réalisant des bootstraps sur 100 réplicats. Un arbre a également été inféré en utilisant les gènes concaténés. Cet arbre retrouvait les relations phylogénétiques communément admises, et a donc été considéré comme l'arbre « vrai » de référence, auquel a été comparé chaque arbre individuel en utilisant un indice topologique. La topologie de chacun des 246 arbres obtenus a donc été comparée à celle de l'arbre de référence obtenu par la concaténation (fig. 1). La congruence des phylogénies a été évaluée en utilisant la méthode de congruence topologique développée par Nye *et al.* [18].

Tableau 1: Génomes complets utilisés pour la recherche d'orthologues. Les 23 espèces retenues pour construire l'arbre de référence sont indiquées par un astérisque et les 21 espèces retenues pour les comparaisons d'arbres sont indiquées par 2 astérisques.

Species	Source	Number of proteins	Rd case	Online database	Taxonomy
<i>Aspergillus fumigatus</i> **	AGD	4 726	« 2.1 »	http://agldmbs.ch/ashbya_gosspii/	Hemi-Ascomycete
<i>Aspergillus nidulans</i> **	NCHI	9 923	07/10/2006	ftp://ftp.ac.biology.gov/genom/es/fungi/	Ascomycete
<i>Aspergillus nidulans</i> **	BROAD	9 541	27/10/2003	http://www.broad.mit.edu/annotation/fungi/fgi/	Ascomycete
<i>Aspergillus oryzae</i> *	NITE	12 074	21/12/2005	http://www.nitego.jp/index-e.html	Ascomycete
<i>Boryspis cinerea</i>	BROAD	16 448	05/10/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Ascomycete
<i>Candida lusitana</i> **	NCHI	5 181	08/10/2005	ftp://ftp.ac.biology.gov/genom/es/fungi/	Hemi-Ascomycete
<i>Candida lusitana</i> **	BROAD	5 941	12/01/2006	http://www.broad.mit.edu/annotation/fungi/fgi/	Hemi-Ascomycete
<i>Chaetomium globosum</i>	BROAD	11 124	24/06/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Ascomycete
<i>Coccidioides immitis</i> **	BROAD	10 457	23/01/2006	http://www.broad.mit.edu/annotation/fungi/fgi/	Basidiomycete
<i>Cryptococcus neoformans</i> **	NCHI	6 475	26/10/2005	ftp://ftp.ac.biology.gov/genom/es/fungi/	Ascomycete
<i>Cryptococcus neoformans</i> **	BROAD	11 640	11/10/2003	http://www.broad.mit.edu/annotation/fungi/fgi/	Hemi-Ascomycete
<i>Fusarium graminearum</i> **	NCHI	5 331	04/08/2005	ftp://ftp.ac.biology.gov/genom/es/fungi/	Ascomycete
<i>Kluyveromyces fragilis</i> **	BROAD	12 841	08/12/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Ascomycete
<i>Magnaporthe oryzae</i> **	BROAD	10 020	28/06/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Basidiomycete
<i>Nerepiza cerasae</i> **	JGI	10 048	« 1.0 »	http://genome-igf-psf.org/	Zygomycete
<i>Phanerochaete chrysosporium</i> **	BROAD	17 467	29/11/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Hemi-Ascomycete
<i>Rhizopus oryzae</i>	MIT	9 424	15/12/2004	ftp://genome-	Basidiomycete
<i>Saccharomyces bayanus</i> *	WashU	4 077	17/12/2004	ftp://genome-	Hemi-Ascomycete
<i>Saccharomyces cerevisiae</i> *	NCHI	5 873	23/01/2006	ftp://genome-	Hemi-Ascomycete
<i>Saccharomyces cerevisiae</i> *	WashU	2 968	04/10/2003	ftp://genome-	Hemi-Ascomycete
<i>Saccharomyces kluyveri</i>	WashU	3 768	08/07/2003	ftp://genome-	Hemi-Ascomycete
<i>Saccharomyces kluyveri</i>	MIT	9 057	15/12/2004	ftp://genome-	Hemi-Ascomycete
<i>Saccharomyces kluyveri</i>	MIT	8 955	15/12/2004	ftp://genome-	Hemi-Ascomycete
<i>Saccharomyces kluyveri</i>	NCHI	5 093	04/08/2005	ftp://ftp.ac.biology.gov/genom/es/fungi/	Ascho-scromycete
<i>Sclerotinia sclerotiorum</i> **	BROAD	14 522	03/10/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Ascomycete
<i>Stagonospora nodorum</i> **	BROAD	16 597	13/06/2005	http://www.broad.mit.edu/annotation/fungi/fgi/	Ascomycete
<i>Tetrahymena rex1</i> **	JGI	9 997	« 1.2 »	http://genome-igf-psf.org/	Basidiomycete
<i>Ustilago maydis</i> **	BROAD	6 522	20/11/2003	http://www.broad.mit.edu/annotation/fungi/fgi/	Hemi-Ascomycete
<i>Yarrowia lipolytica</i> **	NCHI	6 666	04/08/2005	ftp://ftp.ac.biology.gov/genom/es/fungi/	Hemi-Ascomycete
MEAN		8 890			

2.3. Tests statistiques

Les tests statistiques ont été réalisés avec le logiciel JMP (institut 1995 de SAS). Comme les congruences topologiques transformées en Arcsin-racine carrée ne devaient pas significativement de la normalité (test de Shapiro-Wilk W), nous avons réalisé une ANOVA pour examiner l'effet du nombre de sites variables, de la taille des gènes et du pourcentage de sites variables sur les congruences topologiques. Le nombre de sites variables et la taille des gènes étant corrélés ($r=0,959$; $P<0,00001$), une seule de ces variables a été examinée à la fois.

3. RÉSULTATS

Pour les 21 génomes retenus, 246 groupes d'orthologues en copie unique ont été trouvés. Les alignements avaient des longueurs de 111 à 2197 acides aminés (longueur moyenne : 569 aa, longueur médiane : 482 aa). L'identité moyenne parmi les 246 groupes d'orthologues allait de 24 % à 84 % (moyenne : 51 %, médiane 51 %). Ces groupes d'orthologues sont disponibles dans une base de données en ligne, FunnyBase (http://genome.jouy.inra.fr/funnybase/funnybase_result.cgi). Les groupes d'orthologues peuvent être téléchargés pour l'ensemble des données contenant 30 génomes fongiques ou pour l'ensemble des données comprenant les 21 génomes retenus par la suite. Pour chaque groupe d'orthologues, l'identité moyenne, le pourcentage de sites variables, le modèle de substitution d'acide aminé retenu, les scores topologiques et l'arbre obtenu sont disponibles.

L'arbre de référence obtenu en concaténant les 246 groupes d'orthologues en copie unique est présenté en figure 1. Les arbres obtenus pour chacun des 246 alignements ont été comparés à l'arbre de référence en calculant des congruences topologiques sur une échelle de 0 à 100. La distribution des congruences topologiques était proche d'une Gaussienne biaisée vers les faibles congruences (fig. 2). Seuls deux gènes ont donné exactement la même topologie que l'arbre de référence. La plupart des nœuds dans les phylogénies dont la congruence était proche de l'arbre de référence avaient des supports supérieurs à 70 %.

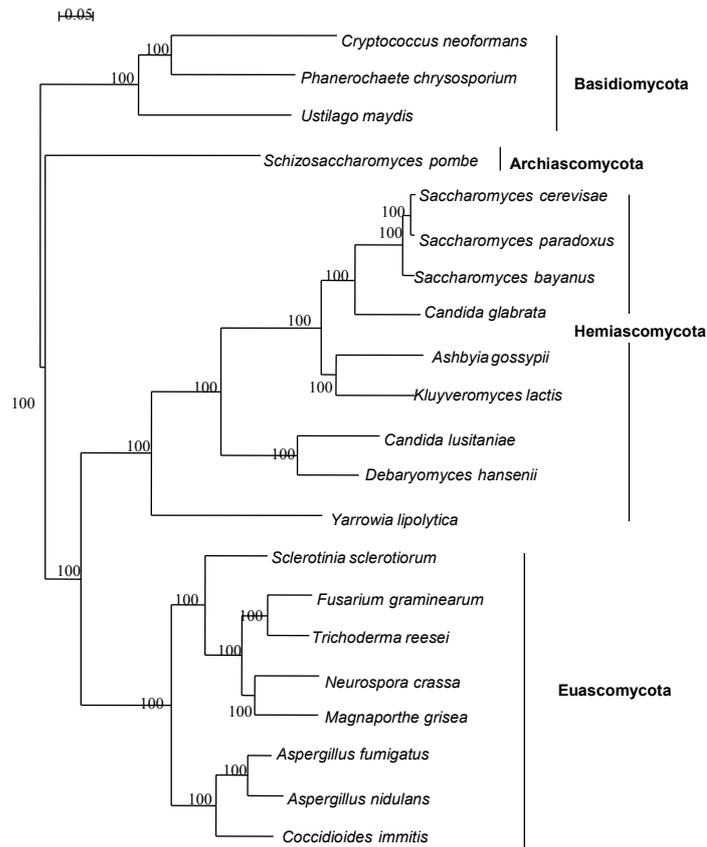


Figure 1: Phylogénie non racinée construite à partir des 21 génomes retenus, en utilisant les 246 groupes d'orthologues en copie simple.

Nous avons également construit des phylogénies en concaténant les 2, 3, 5, 10, 15, 20 et 25 gènes qui avaient fourni les topologies les plus proches de celle de l'arbre de référence. Concaténer les 2 gènes avec la meilleure performance phylogénétique était suffisant pour obtenir une topologie identique à celle de l'arbre de référence, avec tous les supports supérieurs à 70 %.

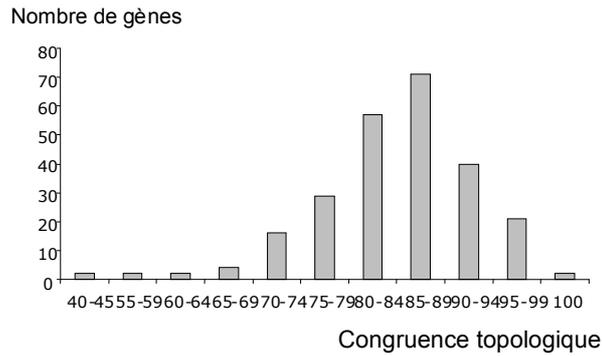


Figure 2: Distribution de la congruence topologique des 246 phylogénies de gènes individuels.

Les gènes les plus généralement utilisés pour établir des phylogénies de champignons (β -tubuline, facteur d'élongation EF1- α , actine, chitinases, polymérase d'ARN, déshydrogénases, et histones) n'ont pas été trouvés dans la liste des gènes avec les meilleurs pouvoirs phylogénétiques. Certains de ces gènes n'étaient en fait même pas présents dans les groupes d'orthologues en copie unique communs aux 21 génomes, car ils avaient des paralogues dans certains génomes.

Nous avons calculé les scores topologiques pour le facteur EF1- α d'élongation, la β -tubuline et la γ -tubuline. Seules 20 espèces avaient des orthologues d'EF1 α parmi notre jeu de 21 d'espèces, et nous avons choisi les gènes avec la plus grande similitude parmi les paralogues dans quelques espèces pour la β -tubuline et la γ -tubuline. Les scores topologiques étaient faibles pour ces trois gènes, même lorsqu'ils ont été concaténés : 80,3 % pour EF1- α , 79,4 % pour la β -tubuline, 84,6 % pour la γ -tubuline et 85,2 % pour l'arbre résultant de leur concaténation. Parmi les 246 orthologues en simple copie, les pourcentages de gènes qui avaient une meilleure performance phylogénétique que la β -tubuline, la γ -tubuline et EF1- α étaient respectivement de 49,2 %, 77,6 %, et 53,6 %.

Tableau II : Gènes donnant les topologies les plus proches de celles de l'arbre de référence, classés par leur score topologique et leur annotation dans *Saccharomyces cerevisiae*.

Gène (nom dans Funybase)	Score topologique	Annotation
MS277	100,0	Protéine impliquée dans le traitement du pré-ARNr 20S dans le cytoplasme
MS456	100,0	Composant du complexe hexamérique MCM qui aide à amorcer la réplication de l'ADN
FG805	96,3	Sous-unité catalytique de la NatB N-terminal acetyltransférase qui catalyse l'acétylation des résidus amino-terminaux de certaines protéines
FG832	96,3	Facteur d'initiation de traduction eIF-5
MS413	96,3	Composant du facteur de transcription TFIIH

Certaines catégories de fonction GO (Gene Ontology, [3]) étaient surreprésentées parmi les 59 gènes donnant les topologies les plus proches de l'arbre de référence, en comparaison avec les 246 groupes d'orthologues. Ces catégories correspondaient aux catégories régulation, organelle, et membrane (fig. 3; test de Chi², P=0,004). Selon l'annotation disponible dans *Saccharomyces cerevisiae* (tabl. II), les deux gènes donnant les topologies identiques à celle de l'arbre de référence codaient : (i) une protéine impliquée dans le traitement du pré-ARNr 20S dans le cytoplasme ; et (ii) un composant du complexe hexamérique MCM qui aide à amorcer la réplication de l'ADN.

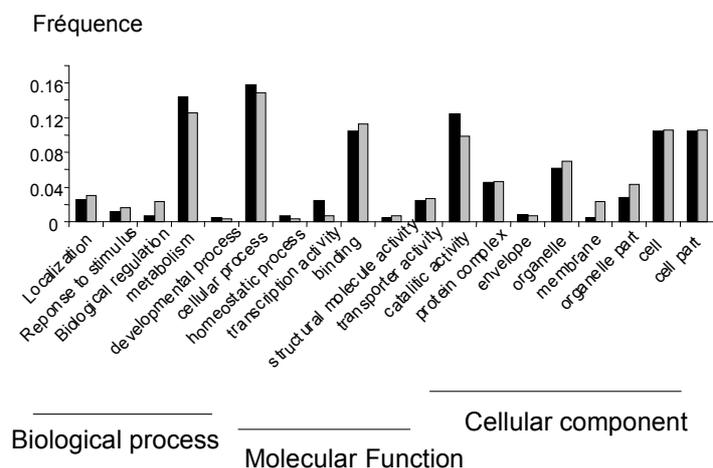


Figure 3: Proportions des fonctions putatives suivant la classification GO (Gene Ontology), parmi les 246 groupes d'orthologues en copie simple (en noir) et parmi les 59 gènes ayant la meilleure performance phylogénétique (en gris).

Les ANOVAs ont révélé que la taille des gènes (c'est-à-dire la longueur d'alignement) et le nombre de sites variables avaient un fort impact sur la congruence topologique. Cependant, les deux gènes donnant la même topologie que l'arbre de référence avaient des nombres de sites variables et des longueurs intermédiaires (fig. 4). Tous les gènes dont l'alignement était plus long que 700 paires de bases avaient des congruences topologiques supérieures à 88 %. Cependant, il y avait une grande variabilité dans la performance phylogénétique pour une taille donnée de gène, et des congruences topologiques supérieures à 90 % pouvaient être obtenues pour n'importe quelle taille de gène.

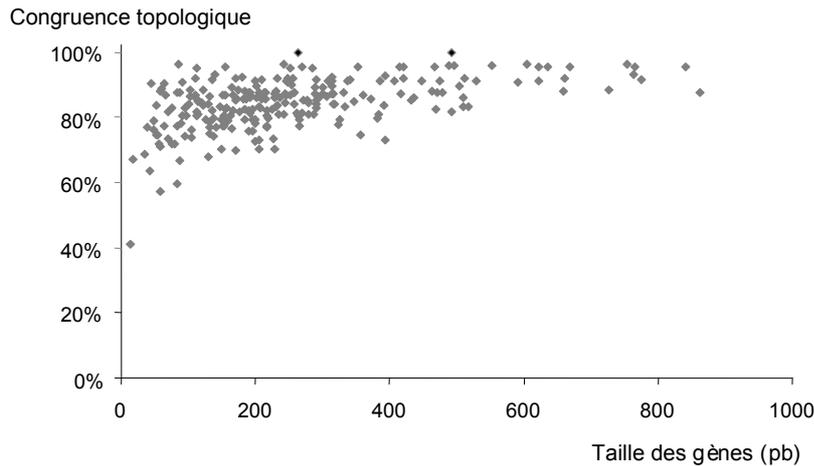


Figure 4: Congruence topologique en fonction de la taille des gènes. Les deux gènes donnant une congruence topologique de 100% avec l'arbre de référence sont indiqués en noir.

4. DISCUSSION

Nous avons trouvé 246 groupes d'orthologues en simple copie communs aux 21 espèces retenues, ce qui représente seulement 2,5 % du nombre moyen de protéines trouvées dans les espèces fongiques considérées. Ce faible nombre de gènes en simple copie partagés par tous les génomes fongiques est en accord avec des études précédentes : 531 groupes d'orthologues en simple copie avaient été trouvés communs à 25 génomes eukaryotes [13], 153 communs à 42 champignons [7] et 854 communs à 17

Ascomycètes [19]. Ceci confirme la difficulté de trouver des gènes orthologues conservés entre taxons distants [13].

La phylogénie des espèces fongiques obtenue ici en utilisant tous les clusters d'orthologues communs aux 21 espèces retenues avait la même topologie que les arbres inférés dans des études précédentes suivant des méthodes similaires ou des méthodes dites de « superarbre » [7], [19], [13]. Dans ces études précédentes, les positions de *Stagonospora nodorum* et d'*Aspergillus oryzae* étaient également problématiques. Les faibles pouvoirs de résolution pour ces deux espèces peuvent être dus à des spéciations rapides qui ont empêché l'accumulation de synapomorphies [21], [22], à des duplications, pertes de paralogues différentielles, ou transferts horizontaux de gènes. *A. oryzae* a en effet été sujet à une expansion à grande échelle de son génome, en raison de duplications et/ou de transferts horizontaux [16]. Les positions difficiles des espèces *S. nodorum* et *A. oryzae* confirment que quelques clades sont réfractaires à la résolution phylogénétique même en utilisant des génomes complets. Nous avons donc enlevé *S. nodorum* et *A. oryzae* de notre ensemble de données pour obtenir un arbre fiable de référence auquel comparer les arbres construits à partir de chacun des 246 groupes d'orthologues.

Une grande variabilité a été obtenue pour la performance phylogénétique des différents gènes : certains ont donné des topologies très différentes de l'arbre de référence, avec parfois pourtant des supports élevés. Seuls deux gènes ont produit exactement la même topologie que l'arbre de référence. Les gènes généralement utilisés pour inférer des phylogénies fongiques n'étaient pas inclus parmi les gènes donnant les topologies les plus proches de celle de l'arbre de référence. Le choix des gènes pour construire des phylogénies est donc crucial, et nous avons identifié ici de bons candidats pour construire des phylogénies fongiques à une large échelle taxonomique. Nos résultats montrent que deux gènes bien choisis étaient suffisants pour obtenir un arbre phylogénétique robuste qui reflétait correctement les relations de parenté entre les espèces.

Des études précédentes sur des taxons différents avaient également montré qu'il y avait une variabilité élevée dans la performance phylogénétique des gènes et que des phylogénies correctes pouvaient être obtenues en utilisant seulement quelques uns des meilleurs gènes, pour des bactéries [12], pour des levures [20] et pour des vertébrés [5], [8], [17], [23], [25].

Les gènes que nous avons identifiés comme les gènes ayant les meilleures performances phylogénétiques devraient être utiles pour construire des phylogénies de champignons à large échelle taxonomique, en économisant les coûts de séquences et en améliorant l'exactitude des phylogénies, en particulier

pour construire l'arbre de la vie. Des nœuds basaux chez les champignons sont en effet encore mal soutenus [11]. La taille des gènes et le nombre de sites variables étaient fortement corrélés et étaient de bons prédicteurs de la performance phylogénétique des différents gènes. Cependant, les deux gènes aux meilleures performances phylogénétiques avaient des tailles intermédiaires et il y avait une grande variabilité dans la performance phylogénétique pour une taille donnée de gène, indiquant que d'autres facteurs sont importants pour la performance phylogénétique.

Les gènes classés ici parmi les gènes aux meilleurs pouvoirs phylogénétiques peuvent ne pas convenir à des échelles taxonomiques plus fines. Néanmoins, ce sont les clades basaux qui posent le plus de problèmes de résolution [11], [15] et sont ceux pour lesquels nos gènes devraient être les plus utiles [14]. Pour des échelles taxonomiques plus fines, le site internet FunyBase permet de rechercher rapidement des gènes en copie unique et communs à la plupart des génomes fongiques, en fonction de leurs degrés de divergence, et sera donc utile pour trouver les gènes appropriés à l'échelle taxonomique désirée.

Les résultats de cette étude ont été publiés dans un article en anglais [2].

REMERCIEMENTS

Nous remercions Gilles Fisher, Bernard Dujon et Francis Martin pour avoir partagé des données non publiées et Joëlle Amsellem, Emmanuel Quevillon, Pierre Nicolas, Damien de Vienne, Sébastien Ollier et Manuela López-Villavicencio pour des discussions sur le sujet. Cette étude a été financée par le Bureau des Ressources Génétiques (BRG 2005-2007), une ACI Jeunes Chercheurs (2003-2006) et un ANR Blanc (06-0201). G. A. a bénéficié d'une bourse post-doctorale du CNRS et d'une bourse post-doctorale de l'Université Paris Sud.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Abascal F., Zardoya R., Posada D., 2005. - ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, 21: 2104-2105.
- [2] Aguilera G., Marthey S., Chiapello H., Lebrun M.-H., Rodolphe F., Fournier E., Gendault-Jacquemard A., Giraud T. 2008 Assessing the performance of single-copy genes for recovering robust phylogenies. *Syst. Biol.* in press.
- [3] Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H. *et al.*, 2000. - Gene Ontology: tool for the unification of biology. *Nature Genet.*, 25: 25-29.
- [4] Castresana J., 2000 -Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.*, 17: 540-552.

- [5] Cummings M. P., Otto S. P., Wakeley J., 1995. - Sampling properties of DNA-sequence data in phylogenetic analysis. *Mol. Biol. Evol.*, 12: 814-822.
- [6] Dujon B., *et al.*, 2004. - Genome evolution in yeast. *Nature*, 430: 35-44.
- [7] Fitzpatrick D.A., Logue M., Stajich J., Butler G., 2006. - A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* 6: 99.
- [8] Graybeal A., 1994. - Evaluating the phylogenetic utility of genes - a search for genes informative about deep divergences among vertebrates. *Systematic Biol.*, 43: 174-193.
- [9] Guindon S., Gascuel O., 2003. - A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biol.*, 52: 696-704.
- [10] Hawksworth D. L., 1991. - The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycol. Res.*, 95: 641-655.
- [11] James T., *et al.*, 2006. - Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, 443: 818-822.
- [12] Konstantinidis K. T., Ramette A., Tiedje J. M., 2006. - Toward a more robust assessment of intraspecies diversity using fewer genetic markers. *App. Environ. Microbiol.*, 72: 7286-7293.
- [13] Kuramae E. E., Robert V., Snel B., Weiss M., Boekhout T., 2006. - Phylogenomics reveal a robust fungal tree of life. *FEMS Yeast Res.*, 6: 1213-1220.
- [14] Liu Y. J., Hall B. D., 2004. - Body plan evolution of ascomycetes, as inferred from an RNA polymerase II phylogeny. *Proc. Natl. Acad. Sci. USA*, 101: 4507-4512.
- [15] Lutzoni F., Kauff F., Cox C. J., McLaughlin D., Celio G. *et al.*, 2004. - Where are we in assembling the fungal tree of life, classifying the fungi, and understanding the evolution of their subcellular traits? *Am. J. Bot.*, 91: 1446-1480.
- [16] Machida M., *et al.*, 2005. - Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, 438: 1157-1161.
- [17] Mueller R. L., 2006. - Evolutionary Rates, Divergence Dates, and the Performance of Mitochondrial Genes in Bayesian Phylogenetic Analysis. *Systematic Biol.* 55: 289-300.
- [18] Nye T., Lio P., Gilks W., 2005. - A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics*, 22: 117-119
- [19] Robbertse B., Reeves J. B., Schoch C. L., Spatafora J. W., 2006. - A phylogenomic analysis of the Ascomycota. *Fungal Genetics and Biology*, 43: 715-725.
- [20] Rokas A., Williams B. L., King N., Carroll S. B., 2003. - Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 425: 798-804.
- [21] Rokas A., Carroll S. B., 2006. - Bushes in the tree of life. *Plos Biol.*, 4: 1899-1904.
- [22] Rosenberg N. A., Nordborg M., 2002. - Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genet.*, 3: 380-390.
- [23] Springer M., DeBry R., Douady C, Amrine H., Madsen O., de Jong W. *et al.*, 2001. - Mitochondrial *versus* nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.*, 18: 132-143.
- [24] Thompson J. D., Higgins D. G., Gibson T. J., 1994. - Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence

- Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.*, 22: 4673-4680.
- [25] Zardoya R., Meyer A., 1996. - Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.*, 13: 933-942.

Modélisation multi-agents des réseaux d'échanges de semences pour la conservation de la biodiversité agricole

Didier BAZILE^{(1)*}, Géraldine ABRAMI⁽²⁾, Souleymane DEMBELE⁽³⁾,
Harouna COULIBALY⁽³⁾, Christophe LE PAGE⁽⁴⁾, Mathieu DIONNET⁽⁵⁾,
Jacques CHANTEREAU⁽⁶⁾, Céline BOUÉ⁽⁶⁾, Mattea ORSINI⁽¹⁾,
François BOUSQUET⁽¹⁾, Jean-Louis PHAM⁽⁷⁾, Kadiatou SANGARE⁽³⁾,
Gilles BEZANÇON⁽⁸⁾

⁽¹⁾CIRAD, UPR 47 "GREEN", 34398 Montpellier, France

⁽²⁾CIRAD, UMR G-Eau, 34398 Montpellier, France

⁽³⁾IER, LaboSEP, Bamako, Mali

⁽⁴⁾CIRAD, UPR 47 "GREEN", Bangkok, Thailand

⁽⁵⁾CIRAD, UPR G-Eau, 34398 Montpellier, France

⁽⁶⁾CIRAD, UPR Agrobiodiversité en savane, 34398 Montpellier, France

⁽⁷⁾IRD, UMR DIA-PC, BP 64501, 34394 Montpellier, France

⁽⁸⁾IRD, Niger, BP 11416, Niamey, Niger

Abstract: Modelling the seed system with agent-based models. This study deals with cereal varieties local management and biodiversity preservation in Mali. The State is not able to provide the adequate seeds to the farmers which supplying mainly relies on seeds exchanges within existing social networks. This informal seed system operates a collective management of many traditional varieties. Our research aims at improving the knowledge on this traditional seeds system. We are particularly interested in identifying the potential and existing roles of the different stakeholders for biodiversity preservation. We use a methodology based on the concept of companion modeling. On the basis of initial hypotheses extracted from an extensive set of surveys, we have produced several successive models using Role-Playing Games and Agent-Based Systems. The process has evolved through the rendering and questioning of these models in specific workshops. The main result of this process is a global and generic agent-based model that will be used to simulate with the stakeholder new regulations for the collective management of cereals biodiversity.

Mali/ Niger/ Biodiversity/ Agriculture/ Collective management/ Agent-based models

Résumé : Notre étude traite de la gestion locale des variétés de céréales au Mali dans un objectif de conservation de la biodiversité. Pour pallier aux insuffisances de

* Correspondance et tirés à part : didier.bazile@cirad.fr

l'Etat, un système informel basé sur les réseaux sociaux existants s'est développé pour approvisionner les paysans en semences. Il permet la gestion collective d'un grand nombre de variétés traditionnelles. Notre recherche cherche à améliorer les connaissances sur ce système semencier traditionnel et notamment sur le rôle des différents acteurs en jeu dans le maintien de la biodiversité. La méthodologie repose sur les concepts de la modélisation d'accompagnement. Sur la base d'hypothèses initiales issues d'un travail d'enquête sur la gestion paysanne des variétés, nous avons produit une série de modèles remis en question au cours d'ateliers successifs utilisant les jeux de rôles et les systèmes multi-agents. Le résultat de ce travail est un modèle qui servira à simuler, avec l'ensemble des acteurs, de nouvelles régulations dans le cadre de la gestion collective de la biodiversité des céréales.

Mali/ Niger/ Biodiversité/ Agriculture/ Gestion collective/ Modèle multi-agents

1. INTRODUCTION

Notre recherche se rattache au concept d'agrobiodiversité pour décrire les rapports entre la diversité biologique, environnementale et socioculturelle [11], [12]. Nous examinons les nombreuses manières dont les paysans exploitent la biodiversité pour la production agricole. La prise en compte exhaustive des parcelles cultivées d'un village permet de recenser l'ensemble des cultivars présents mais c'est au final la multiplicité des stratégies des paysans qui dessine la richesse du patrimoine biologique du village et l'inscrit dans la logique de diversité culturelle propre aux sociétés agricoles des pays du Sud. C'est donc en analysant comment et pourquoi les agriculteurs choisissent leur portefeuille de variétés que nous pouvons comprendre les mécanismes de gestion en jeu et chercher à les améliorer.

Dans un contexte quasi non marchand, la recherche montre l'importance du lien entre les échelles spatiales pour appréhender les dynamiques locales (paysan, communauté) autour de la circulation et de la gestion des semences. La présente étude explicite les liens existants entre le paysan (en tant qu'entité individuelle) et son environnement (au sens large) puis, formalise les règles de choix et les stratégies en expliquant les changements dans les variétés semées.

Après avoir restitué le contexte et les objectifs de ce travail, nous présentons la démarche générale de modélisation d'accompagnement qui sous-tend notre méthodologie, ainsi que les outils que nous utilisons : les jeux de rôles et les systèmes multi-agents. Nous détaillons ensuite l'évolution de notre compréhension du système aux différentes étapes du processus de recherche : formalisation des hypothèses issues d'enquêtes de terrain dans un modèle multi-agents expert initial ; désagrégation de la représentation

dans trois jeux de rôles portant sur les différents modules de décision qui permettent de caractériser les liens du paysan à son environnement ; intégration des nouvelles hypothèses issues des ateliers de jeux dans un modèle multi-agents global à portée générique. Nous concluons alors sur la manière dont ce modèle, après avoir été validé et paramétré par les paysans, nous permet d'explorer et de construire de nouvelles règles de gestion des semences via la simulation des dynamiques d'évolution des variétés au sein des villages.

2. MATÉRIEL ET MÉTHODES

2.1. Contexte et objectifs

La diversité des ressources phytogénétiques constitue la base biologique de la sécurité alimentaire mondiale car elle fournit des moyens de subsistance à tous les habitants de la planète [13], [17]. Ces ressources sont la matière première la plus importante pour le sélectionneur et l'intrant le plus essentiel à l'agriculteur ; elles sont donc indispensables à la durabilité des systèmes agricoles. Beaucoup de ressources phytogénétiques en agriculture sont le résultat d'une intervention de l'homme et plus récemment les sélectionneurs ont exploité cette diversité avec des effets remarquables sur l'amélioration des variétés. La viabilité des ressources génétiques en agriculture nécessite une gestion humaine, active et continue. Ainsi, depuis la Convention de Rio sur la Biodiversité [15], le maintien de la diversité des pratiques humaines apparaît de plus en plus comme une solution pour la conservation d'une diversité élevée dans la biosphère. Ces variétés continuent d'évoluer en permanence sous l'effet de pressions de sélection humaine et environnementale, de flux de semences au sein des réseaux d'échanges et de choix formés par des identités socioculturelles différentes. Ces processus participent à la conservation dynamique du patrimoine génétique et ainsi à la création de la nouvelle variabilité génétique d'importance globale. Au Sahel, les agriculteurs ont essentiellement recours à des systèmes semenciers informels, conçus autour de leur propre production de semences [2]. Même s'ils sont méconnus, ils participent à plus de 90% à l'approvisionnement de proximité en semences de céréales. Ils sont aujourd'hui confrontés au renforcement formalisé, habituellement commercial, des systèmes nationaux de production des semences [3]. Les paysans ont maintenu leur système local parce qu'il est en mesure de leur apporter plus de sécurité et de productivité que le système semencier national grâce à un fonctionnement susceptible d'entretenir une plus grande diversité. Néanmoins les paysans peuvent accéder aux variétés du Système Semencier National dès lors qu'elles répondent à leurs besoins variétaux et objectifs d'approvisionnement en semences.

Actuellement, la gestion des semences par le paysan et l'amélioration des variétés par les chercheurs ont beaucoup de difficultés à trouver des points de rencontre. Les différents acteurs impliqués dans la conservation de l'agrobiodiversité (chercheurs, développeurs et paysans) ont pourtant tous besoin d'intégrer les savoirs scientifiques et locaux dans une représentation partagée du système semencier. Ceci est indispensable pour bâtir un programme commun cohérent avec les différents acteurs pour que le maintien *in situ* de la diversité repose sur les valeurs socio-économiques et culturelles. L'objectif général de cette recherche est de renforcer notre capacité commune (paysans et chercheurs) à évaluer l'effet des pratiques de gestion des semences sur le maintien de la diversité variétale. De multiples acteurs participent à la fourniture des semences ; ils incluent notamment la famille, les organisations paysannes et ONG locales, les distributeurs privés, la recherche et le secteur semencier public. Ces acteurs, leurs comportements et leurs multiples interactions constituent le « système semencier ».

La présente étude porte sur les systèmes locaux de gestion des semences de sorgho au Mali sur une zone allant de Gao à Sikasso. Le sorgho y est la principale céréale alimentaire et les paysans cultivent majoritairement des variétés locales, rustiques et adaptées aux conditions agroclimatiques difficiles de la zone soudano sahélienne. Dans le sud de la zone d'étude, où le climat est plus favorable à l'intensification des cultures, le sorgho est concurrencé par le maïs et on observe une érosion variétale. Dans ce contexte, le taux d'adoption des variétés améliorées issues de la recherche reste marginal et l'approvisionnement en semence se fait essentiellement de manière informelle.

2.2. Méthodes et outils

La méthodologie générale repose sur le développement, via une démarche de modélisation d'accompagnement, d'un modèle du système semencier capable de représenter la diversité des situations agricoles présentes au Mali. Ce modèle doit permettre de caractériser les circuits d'approvisionnement en semences et l'accessibilité des paysans au sein de la communauté à la diversité des variétés. Le souci de développer un modèle multi-agents intégrant dès sa conception des éléments susceptibles d'augmenter la généralité et l'adaptabilité du modèle a été pris en compte en travaillant à plusieurs échelles et sur différents sites. Les sites d'étude du Mali s'étalent sur tout le gradient nord-sud de la zone de culture du sorgho pluvial (de 16°N à 11°N). Les enquêtes portent sur une trentaine de village avec des études approfondies sur 5 sites pilotes entre Gao et Sikasso (fig.1). Les différents ateliers organisés au cours de cette recherche ont permis de faire interagir entre eux les paysans de ces 5 villages puis d'étendre les

échanges avec des paysans de 4 régions agricoles du Niger où des enquêtes complémentaires ont été conduites.

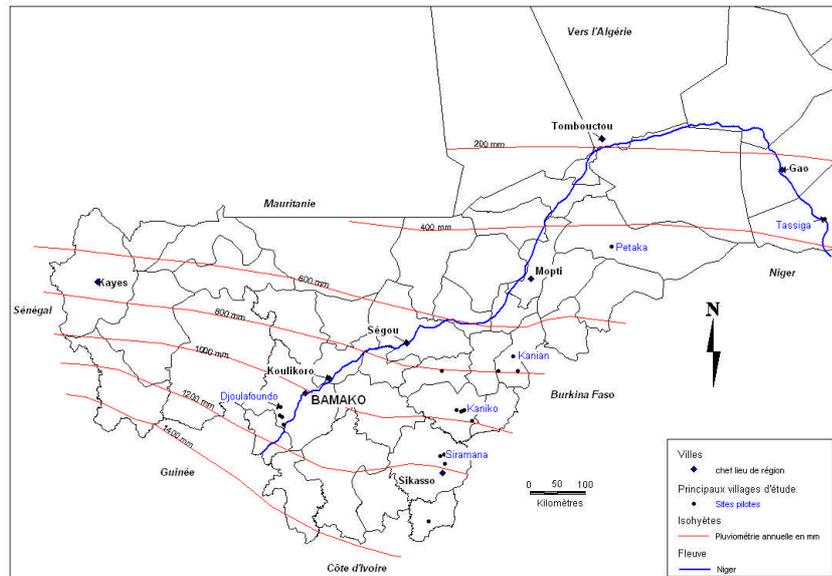


Figure 1 : Localisation au Mali des principaux sites d'étude.

2.2.1. La modélisation d'accompagnement

La modélisation d'accompagnement est une démarche mise au point par un collectif de chercheurs issus de différentes disciplines [4], [9]. Elle vise à faciliter des processus collectifs de décision sur la gestion de ressources naturelles en s'appuyant sur la formalisation des points de vue et des critères subjectifs auxquels se réfèrent les différentes parties prenantes. Cette formalisation s'opère dans différents modèles construits au cours de cycles itératifs permettant la remise en question continue des représentations. Un cycle de modélisation d'accompagnement comporte typiquement trois étapes : 1. synthèse des connaissances existantes, 2. formalisation dans un modèle, 3. simulation et confrontation aux parties prenantes.

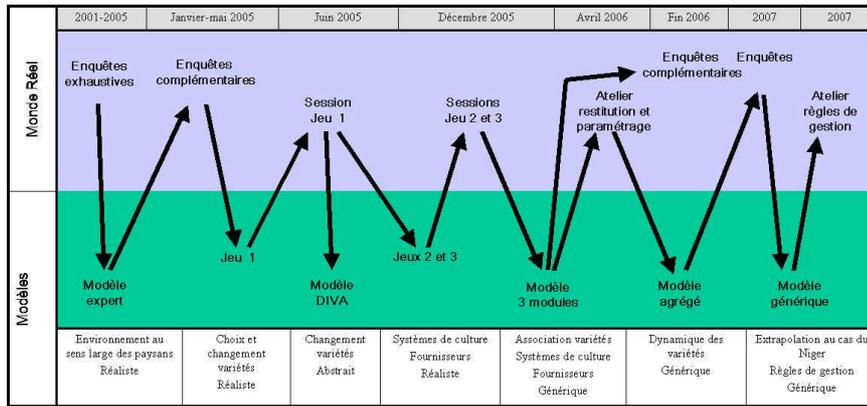


Figure 2 : Démarche itérative de la modélisation d'accompagnement développée au Mali.

La figure 2 retrace les étapes du processus initié où nous avons utilisé les systèmes multi-agents (SMA) et les Jeux de Rôles (JdR) :

- mobilisation des données d'enquêtes obtenues à l'échelle de l'exploitation agricole sur une trentaine de villages pour construire l'architecture d'un modèle multi-agents expert (octobre 2004) ;
- identification des lacunes pour mener des enquêtes complémentaires et formalisation des hypothèses établies dans un premier JdR (jeu 1) ;
- mise en oeuvre du Jeu 1 lors d'un atelier en juin 2005 : construction de nouvelles hypothèses sur les stratégies paysannes d'association de variétés (module 1) ;
- identification de deux autres modules qui permettent de caractériser les liens entre le paysan et son environnement par rapport à la gestion des variétés : influence des systèmes de culture sur le choix des variétés (module 2) et importance des aspects sociaux et spatiaux dans les échanges de semences (module 3). Formalisation des hypothèses correspondantes dans deux JdR (jeux 2 et 3) ;
- mise en oeuvre des jeux 2 et 3 lors d'ateliers en décembre 2005 : révision des hypothèses ;
- généralisation dans un SMA comportant 3 modules correspondant aux 3 JdR.
- mises en situation du SMA au cours d'un atelier en avril 2006 : analyse et discussion des similarités et des différences propres à chacune des 5 situations agricoles retenues.

- mise en œuvre du JdR « Seed-Div » dans un atelier sur les règles de gestion en avril 2007 afin de travailler sur l'émergence de nouvelles institutions favorables à la conservation et à un accès partagé à la diversité variétale.

2.2.2. Les enquêtes

Un important travail d'enquêtes a été conduit durant 4 années au Mali avec, pour objectif principal, la caractérisation des écosystèmes à base de sorgho. L'identification des variables constituant l'environnement des systèmes de cultures à base de sorgho a été conduite autour de trois axes : 1- l'analyse socio-économique des agroécosystèmes céréaliers ; 2- l'analyse agronomique de la culture du sorgho ; et 3- l'analyse de la biodiversité du sorgho et des facteurs relatifs à son évolution. Cette approche classique a été couplée avec la création d'un système spatialisé d'information sur l'environnement du sorgho intégrant la caractérisation bioclimatique du pays et les résultats d'un zonage agroécologique. Une cartographie détaillée des sites pilotes à l'échelle des territoires villageois a été conduite avec pour objectif de préciser la place de la diversité variétale du sorgho dans les stratégies d'assolement. L'analyse des modes de culture du sorgho s'appuie alors sur le choix des variétés en fonction des facteurs de l'environnement au sens large des géographes incluant la diversité des exploitations. L'approche spatiale permet de bien prendre en compte l'abondance des différentes variétés au-delà d'une simple liste représentant une richesse variétale.

2.2.3. Les jeux de rôles

Le jeu de rôle correspond à la mise en scène d'une situation complexe dans un espace contrôlé. C'est un modèle qui sert de médiateur avec une réalité parfois difficile à formuler pour un paysan lors des enquêtes et, dans ce cas, l'aspect ludique du jeu favorise les échanges d'information. L'objectif principal de la première série de jeux de rôles mis en place au Mali est d'améliorer notre connaissance [ou représentation du système] en la partageant pour la remettre sans cesse en question (tabl. I):

- le Jeu 1 plaçait les paysans en situation dans des exploitations agricoles virtuelles soumises à des contraintes climatiques accentuées. Il a permis d'identifier des stratégies archétypales d'association de variétés mais aussi un manque de connaissances sur les différentes modalités d'approvisionnement en semences et sur l'influence de la succession des cultures sur les choix variétaux ;
- le Jeu 2 s'est alors intéressé spécifiquement aux systèmes de cultures et aux rotations ;

- le Jeu 3 s'est focalisé sur les contraintes spatiales et sociales lors de l'approvisionnement en semences. Différents types de fournisseurs ont pu être associés à des modes de recherche spécifiques.

Tableau I : Principales caractéristiques des jeux de rôles mis en place au Mali.

Jeux de rôles	J1 Système de production	J2 Système de culture	J3 Fournisseurs
Question posée	Assurer la sécurité alimentaire de la famille	Définir l'assolement de l'exploitation en respectant la succession des cultures	S'adapter au changement climatique en changeant, si nécessaire, ses variétés
Objectif général	Analyser les comportements de gestion des variétés de sorgho selon les types d'exploitations	Comprendre le choix des variétés de mils et de sorghos dans le temps et l'espace	Caractériser les modalités d'accès aux semences en fonction des fournisseurs
Joueurs	12 joueurs (diversité des systèmes de production, liens familiaux) Exploitation identique à la leur.	15 joueurs (3 tailles d'exploitations sur 3 sols contrastés) Exploitation « modèle » correspondant à la taille de la leur et sur un seul sol.	10 joueurs (diversité des types de systèmes de production, liens familiaux) 3 rôles: chef de village, organisation paysanne et détenteur des savoirs traditionnels. Les joueurs jouent leur rôle et leur exploitation.
Espace	Tables pour les groupes familiaux. Maquette du village comme espace collectif.	3 tables pour chaque groupe de 5 exploitations cultivant sur un même sol	2 tables réunissant sur un même sol 5 exploitations. 1 table pour les types de fournisseurs extérieurs au village. Chaises en cercles pour chaque groupe familial.
Déroulement du temps	Année agricole découpée en 3 pas de temps : annonce de la saison climatique, récolte et sélection des semences, saison sèche	Pas de temps annuel avec annonce de l'année climatique	Année agricole scindée en 4 pas de temps : annonce de la saison climatique, récolte et sélection des semences, échanges familiaux, saison sèche. Carte d'évènements
Date réalisation	Juin 2005	Décembre 2005	Décembre 2005

Pour faciliter l'appropriation par les paysans de l'environnement virtuel proposé, nous avons toujours inclus des éléments de contexte du monde réel : structure de l'exploitation agricole, famille, surface, appellations locales, etc... On parle alors de réalité explicite : le jeu offre une représentation réaliste des acteurs, des ressources et des configurations spatiales rencontrées par les paysans dans la réalité. Le paysan joue son

exploitation avec les caractéristiques qu'il lui connaît : surface cultivée, équipement, population active, bouches à nourrir, cultures, etc. Le JdR Seed-DIV se place dans une autre perspective avec un objectif nouveau de servir de support au dialogue entre des paysans issus de régions différentes. Afin de faciliter les échanges, le jeu était volontairement très abstrait afin de que les participants dialoguent collectivement autour d'un système générique sur les stratégies à mettre en place permettant le maintien en commun d'une diversité variétale.

2.2.4. Les SMA

Les SMA sont issus du domaine de l'intelligence artificielle distribuée qui cherche à résoudre les problèmes par la coordination d'agents hétérogènes indépendants. Les agents perçoivent leur environnement et les autres agents, et s'en construisent des représentations. Utilisant des capacités de raisonnement plus ou moins évoluées, ils agissent alors en modifiant l'environnement ou en communiquant avec les autres agents. Les SMA sont adéquats pour la simulation de formes de coordination observables sur le terrain, en particulier dans le domaine de la gestion des ressources renouvelables car ils sont bien adaptés à l'étude de problèmes intégrant des dynamiques naturelles et sociales [10]. La modélisation multi-agents permet de simuler les changements de règles de gestion et de comprendre les effets superposés des stratégies individuelles et des règles collectives. C'est dans ce cadre que nous avons utilisé les modèles comme outils de médiation et de partage de la connaissance pouvant servir de support aux questionnements, l'objectif étant de réfléchir à des techniques, règles ou attitudes favorables au maintien de la diversité génétique *in situ* et notamment de déterminer l'échelle la plus adéquate pour définir des règles relatives à la conservation *in situ* des variétés de céréales.

3. RÉSULTATS

La formalisation du système semencier dans des systèmes multi-agents s'est construite selon une démarche itérative en boucles où trois modèles multi-agents ont été produits (tabl. IIa et IIb):

- un premier modèle « expert » synthétisait l'ensemble des hypothèses issues des enquêtes ;
- un second modèle abstrait formalisait les comportements des joueurs lors du Jeu 1 ;
- un troisième modèle générique reprend dans des modules indépendants les modalités de prise de décision identifiées dans les trois ateliers de jeux.

Tableau IIa : Niveau de réalisme, paramètres et dynamiques pour les entités naturelles dans chacun des 3 modèles produits.

	Modèle EXPERT	Modèle DIVA	Modèle à 3 MODULES
CLIMAT			
Paramètres	∅	Facteur d'impact sur le rendement (entre 0 et 1)	Date d'arrivée des pluies Cumul de précipitation
Dynamiques	∅	Généré aléatoirement	Lecture de séries climatiques
Données	∅	∅	Séries climatiques historiques Appréciation année paramétrée par les utilisateurs
CULTURE			
Paramètres	Type de culture (coton / maïs / sorgho / mil / autres / jachère) Parcelle (surface, localisation)	1 seule culture : le sorgho	Type de culture (coton / maïs / sorgho / mil / autre) Surface dans l'exploitation
VARIETE			
Paramètres	Nom des variétés de la région Probabilité initiale de présence	5 variétés abstraites Rendement maximal 4 caractéristiques abstraites	Cycle: précoce, moyen, tardif Nom de 3 variétés par type
Dynamiques	∅	Rendement selon climat	Matrice de gain f(type/climat)
Données	Statistiques issues des enquêtes*	Génération aléatoire des caractéristiques Fonction théorique d'impact du climat	Caractérisation qualitative de l'écologie des variétés de sorgho

*Les statistiques utilisées dans le modèle expert sont basées sur 3 zones géographiques : San, Koutiala et Sikasso

D'un modèle expert complexe car très réaliste, les modèles suivants ont gagné en abstraction en affinant les hypothèses et en se limitant aux variables déterminantes du système pour le choix des variétés de sorgho.

Tableau IIb : Niveau de réalisme, paramètres et dynamiques pour les entités sociales et spatiales dans chacun des 3 modèles produits.

	Modèle Expert	Diva	Modèle à 3 Modules
TERRITOIRE			
Paramètres	Texture, Pente et Profondeur des sols	Territoire uniforme	2 types de sol
Dynamiques	Priorité de défriche	Ø	Priorité de défriche
Données	Statistiques issues des enquêtes*	Ø	Paramétrage par les utilisateurs
EXPLOITATION			
Paramètres	Nombre de champs (1 à 3) Taille des champs (4 à 13 ha) Voisinage (distance seuil) Groupe Familial (3 à 5)	Nombre de parcelles (5) Voisinage (territoire entier) Groupe Familial (3 à 5 exploitations)	Taille de l'exploitation (G/P) Niveau d'équipement Voisinage (distance seuil) Groupe Familial (3 à 5)
Dynamiques	Assolement, rotations	Rendement moyen	Attribution des stratégies
Données	Statistiques issues d'enquêtes*	Fonctions théoriques	Paramétrage par les utilisateurs
AGRICULTEUR			
Paramètres	Nombre de variétés (1 à 3) Mode d'essai (direct, échantillon, progressif) Choix initial des variétés	4 préférences pour chacune des caractéristiques des variétés Facteur d'impact des préférences Stratégie de remplacement (actif ou passif) Stratégie de recherche (idéotype ou pas)	Stratégie d'association de variétés Système de culture Stratégie de recherche de variété Nombre de variétés Base de connaissance variétés
Dynamiques	Essai d'une nouvelle variété tous les 10 ans avec 75% de chance de réussite Recherche de variétés dans le voisinage et/ou le Groupe Familial	Calcul de la satisfaction (préférences et rendements) Recherche d'une variété selon ses caractéristiques et /ou son poids (taux de présence et proximité)	Calcul d'un assolement selon climat et stratégies Recherche de variété dans le voisinage et/ou le Groupe Familial selon stratégie de recherche de variété
Données	Statistiques issues des enquêtes*	Fonctions théoriques caractérisant des types de stratégies observées	Caractérisation qualitative de stratégies observées

*Les statistiques utilisées dans le modèle expert sont basées sur 3 zones géographiques : San, Koutiala et Sikasso

3.1. Des enquêtes de terrain... à la formalisation du modèle expert

Le sorgho est cultivé du sud du Mali jusqu'à la fin de la bande sahélienne sous environ 500 mm de pluviométrie annuelle avec un risque de sécheresse qui est intégré aux stratégies paysannes. Partout, les agriculteurs ont gardé de multiples variétés locales qui ont comme caractéristiques communes leur adaptation aux conditions écologiques locales et le calage de leur cycle avec le microclimat de la zone géographique [16]. C'est le photopériodisme qui permet l'ajustement du cycle de la plante à la durée probable de la saison des pluies en permettant une grande souplesse dans les dates de semis. Les exploitations agricoles se différencient en fonction de leur taille (surface et main d'œuvre disponible) et du niveau d'équipement qui détermineront la possibilité de semer plusieurs variétés de sorgho chaque année sur différents sols.

L'élaboration de référentiels locaux de la culture du sorgho nous a permis de préciser les pratiques paysannes relatives à la culture du sorgho pour prévoir l'aire de diffusion des variétés locales ou améliorées [6], [14]. Malgré la sécheresse, les variétés photopériodiques occupent toujours une place prépondérante dans les systèmes de culture où la gestion du risque l'emporte sur les besoins d'intensification. C'est pourquoi dans la connaissance de l'agrobiodiversité, le nombre de variétés gérées dans l'exploitation et l'identification de la dynamique de remplacement, constituent un pas essentiel pour l'identification de nouveaux modes de conservation [7].

Le modèle premier multi-agents dit «expert» a été développé pour mettre en relation les nombreux résultats issus des enquêtes et représenter au sens large l'environnement des paysans : les différentes entités du territoire et la structure sociale des paysans [8]. Ce modèle expert a servi d'intégrateur pour les diverses disciplines et permis de préciser les paramètres d'entrée nécessaires pour caractériser le système semencier. La création d'un état initial de référence «modulable» permet de paramétrer les spécificités environnementales ou structurelles des exploitations agricoles des différentes zones géographiques du Mali. Néanmoins, le modèle expert reste statique et il est nécessaire d'implémenter des règles de dynamique pour pouvoir réaliser des simulations sur des scénarios simplifiés de circulation des variétés autour des cas suivants : pas d'introduction extérieure, accès par proximité spatiale des champs, accès aux pools de variétés des groupes familiaux et connaissance des personnes ressources.

3.2. Des lacunes du modèle expert au premier jeu de rôle

Après avoir identifié les lacunes du modèle expert sur le remplacement des variétés, des enquêtes spécifiques ont été conduites pour hiérarchiser les

critères de choix des variétés et les modalités du changement. La curiosité et la volonté d'améliorer la productivité poussent certains paysans à tester régulièrement de nouvelles variétés sans pour autant noter d'échec sur leurs précédentes cultures. Le climat est le plus souvent à l'origine du changement de variété mais des modifications de la stratégie agricole de l'exploitation peuvent également être impliquées ; par exemple le développement de la culture du maïs dans le sud du Mali. Enfin, l'introduction d'une nouvelle variété ne prend pas forcément la forme d'un test sur plusieurs années.

Lors du premier jeu de rôle, en fonction de la date d'arrivée des pluies, les paysans devaient choisir des variétés de sorgho pour atteindre un objectif de production correspondant aux besoins alimentaires de leur famille. En fonction des rendements obtenus, calculés à partir d'abaques, les paysans étaient libres de produire de la semence sur leur parcelle ou de rechercher de nouvelles semences à l'extérieur. L'objectif de production est contraint par la quantité de travail disponible pour déterminer l'assolement décrit en termes de surfaces par culture et de variétés semées à des dates précises. Le premier JdR permet de faire ressortir deux résultats essentiels qui ont alimenté notre démarche de modélisation d'accompagnement par la suite :

- le comportement du paysan pour la décision de changement de variété peut reposer sur un processus d'évaluation propre ou sur un processus d'imitation des voisins ;
- l'hypothèse de trois stratégies distinctes d'association des variétés au sein de l'exploitation sur laquelle a été construite un module du troisième SMA.

3.3. Le modèle abstrait DIVA

Le SMA « DIVA » (pour DIVersité Variétale) représente la gestion paysanne de la biodiversité dans un modèle de décision simple, déconnecté de l'environnement réel. Il permet de tester les conséquences à un niveau global de la somme de décisions individuelles relatives au choix de variétés selon deux processus d'évaluation ou d'imitation. La portée théorique du modèle repose sur la question de la diffusion d'une innovation, la variété dans notre cas, grâce à la communication entre les individus. L'échange d'information est possible au sein d'un réseau géographique, correspondant à l'environnement physique du paysan et d'un réseau social, correspondant à l'univers social propre à chaque paysan. Ce réseau social permet à des individus géographiquement éloignés d'échanger des informations qu'ils ne pourraient pas ou rarement avoir avec leur entourage proche. Il est contraignant, dans le sens où il est prioritaire sur la distance qui sépare deux individus. Deux comportements paysans archétypaux peuvent être décrits:

- Les paysans passifs fonctionnent sur un mode d'imitation où les variétés introduites sont celles qui sont le plus cultivées autour d'eux.
- Les paysans actifs sont sans cesse à la recherche d'un « idéotype » de variété. Le paysan prend alors un risque pour introduire une variété dans son exploitation sans référence sur celle-ci dans son voisinage. Les agents de DIVA doivent donc évaluer le niveau de production de leurs parcelles à chaque pas de temps et lorsque les résultats sur leur(s) variété(s) semée(s) ne leur conviennent pas, ils choisissent une nouvelle variété en fonction de leurs préférences.

La stratégie passive d'imitation (fig. 3) correspond à une situation stable dans laquelle aucune variété ne semble prendre le dessus sur les autres. A l'opposé, la stratégie active d'évaluation aboutit très rapidement à un déséquilibre dans lequel une variété domine (fig. 3). La diversité globale du système chute très vite et seules deux variétés se maintiennent dans le système avec l'une qui représente 80% des surfaces semées. Nous pouvons souligner la puissance de l'outil de modélisation qui, avec des règles très simples de décisions des agents dans leur environnement, donne la possibilité de produire des situations qui seraient très difficilement prévisibles et caractérisables par une approche mathématique classique. Dans notre cas, nous ne cherchons pas à valider la réalité des règles introduites dans le modèle mais plutôt à les utiliser pour discuter de façon théorique du type de comportement qu'elles illustrent. Ainsi, on s'intéresse plus à l'impact d'une introduction sur la biodiversité au niveau global qu'au fait qu'une variété prenne le dessus sur les autres.

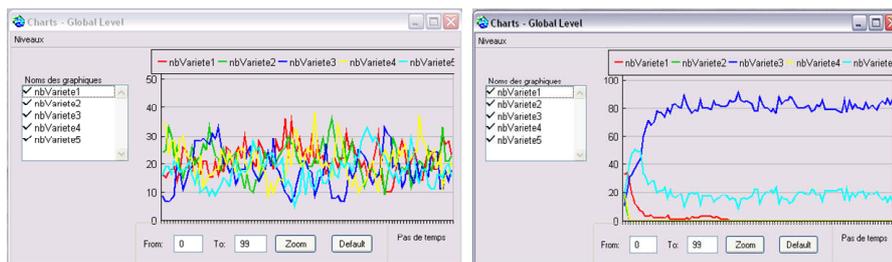


Figure 3 : Analyse de sensibilité du modèle DIVA sur 2 scénarios :

à gauche : 100 % des paysans suivent la stratégie passive d'imitation ;

à droite : 100 % des paysans suivent la stratégie active d'évaluation.

3.4. Une décomposition du système en 3 modules

La désagrégation de notre représentation du système semencier permet d'analyser progressivement les règles dans des sous-systèmes en éliminant volontairement au départ certaines relations. Chaque contexte de décision

particulier a fait l'objet du développement d'un module indépendant et le modèle final est structuré en trois modules indépendants [1] :

- Les stratégies d'association de variétés ;
- La composante système de culture en fonction de la structure de l'exploitation;
- La quête du fournisseur ;

Les hypothèses issues du jeu 1 permettent de formaliser les dynamiques du module « stratégies d'association de variétés ». Dans le jeu 2, les paysans devaient établir leur assolement annuel en tenant compte des rotations et des contraintes climatiques. Lors du jeu 3, les paysans devaient chercher des semences en fonction d'évènements marqués, alors que certains d'entre eux représentaient des rôles particuliers dans le village et que les animateurs jouaient le rôle de fournisseurs extérieurs. Les résultats des jeux de rôles 2 et 3 alimentent les modules « système de culture » et « fournisseur ».

4. DISCUSSION

4.1. Complémentarité des approches : du jeu au SMA

Les différents jeux ont permis d'améliorer notre connaissance sur le système semencier et, d'implémenter immédiatement les nouveaux résultats dans le modèle en construction permanente. Sans acquérir nécessairement de connaissances nouvelles, le jeu permet de bien remettre les hypothèses et résultats déjà acquis sous forme de relations cohérentes. Ainsi, la gestion des variétés selon les types d'exploitation est déterminée par deux facteurs principaux : la surface et le niveau d'équipement. C'est en effet la capacité de travail qui influence en premier lieu le calendrier agricole de l'exploitation et les choix variétaux qui en découlent (JdR2). De la même façon, le JdR3 a permis de valider notre structuration sociale des exploitations avec les personnes ressources, les groupes familiaux, les petits et gros producteurs mais il a surtout mis l'accent sur l'environnement spatial et temporel qui favorise ces échanges.

En soumettant à la fois nos outils à la critique et nos résultats à la réfutation, nous nous engageons réellement dans une co-construction sur la base d'un partenariat équitable. Ces échanges lors des étapes de debriefing des jeux nous ont permis de spécifier trois cas particuliers relatifs à la recherche d'une variété qui déterminent des cheminements différents pour l'accès à la semence recherchée :

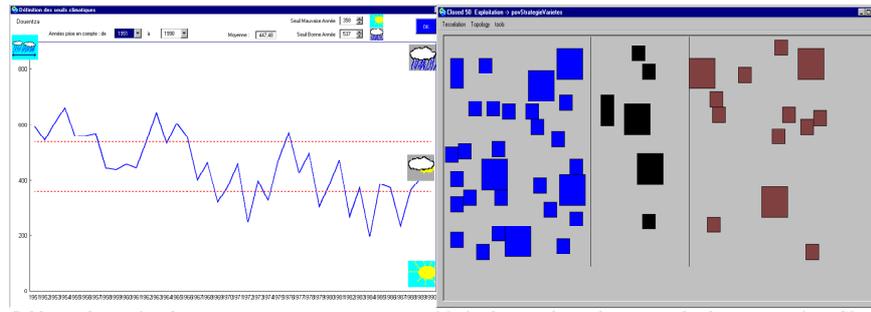
- recherche de la quantité nécessaire d'une variété pour semer toute sa parcelle ;

- recherche d'un type de variété précis connu sur village (même si peu de semence) ;
- recherche d'une variété nouvelle (idéotype) non connue qu'il faudra expérimenter.

4.2. La construction du modèle autour des 3 modules du système

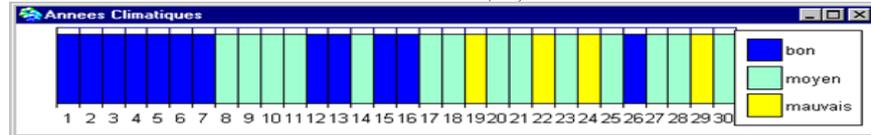
L'objectif du 3^e modèle SMA est de pouvoir travailler indépendamment sur chacun des trois aspects de la prise de décision des paysans. Les trois modules peuvent fonctionner ensemble, ou bien séparément. Dans ce cas, là, un comportement est défini par défaut dans le modèle pour les modules non utilisés. Ce modèle présente un niveau d'abstraction supérieur au modèle expert pour faire discuter des paysans issus de différentes régions. Cette abstraction se traduit par une qualification qualitative des paramètres descriptifs du modèle (climat, types de variétés, gains, types de sols, taille et équipement des exploitations). Une simulation permet d'observer la dynamique d'occupation des variétés dans un village avec un pas de temps annuel. Le climat simulé est construit à partir de séries climatiques historiques sur les régions d'origine des paysans et caractérisées qualitativement par leurs soins en début de session. Selon l'année climatique, les agents paysans déterminent leur assolement (module système de culture ; par défaut valeur moyenne sur la région) et les variétés qu'ils sèment (module association de variétés; par défaut choix d'une variété moyenne). Des gains traduisent le niveau de rendement obtenu. Selon les gains obtenus et leur stratégie définie de renouvellement des semences, les agents paysans peuvent changer de variétés. Ils recherchent alors des semences dans le village ou bien s'adressent à des fournisseurs extérieurs.

Dans la perspective d'utilisation rapide des modèles multi-agents pour la mise en place d'un dialogue entre acteurs de la gestion des RPG, la caractérisation génétique a été passée sous silence pour afficher une caractérisation des variétés davantage agromorphologique que génétique, plus en accord avec la représentation des variétés que se construit le paysan. Néanmoins, la dimension de la diversité génétique des variétés paysannes devra être considérée de façon plus approfondie dans le futur au sein de nos modèles pour analyser réellement la dynamique *in situ* de cette diversité au travers des pratiques et stratégies paysannes de gestion.



Calibrage des années climatiques

Matérialisation des exploitations selon les trois stratégies *S1* avec une seule variété précoce (*S1P*), moyenne (*S1M*) ou tardive (*S1T*)



Succession des années climatiques sur les 30 années de la simulation

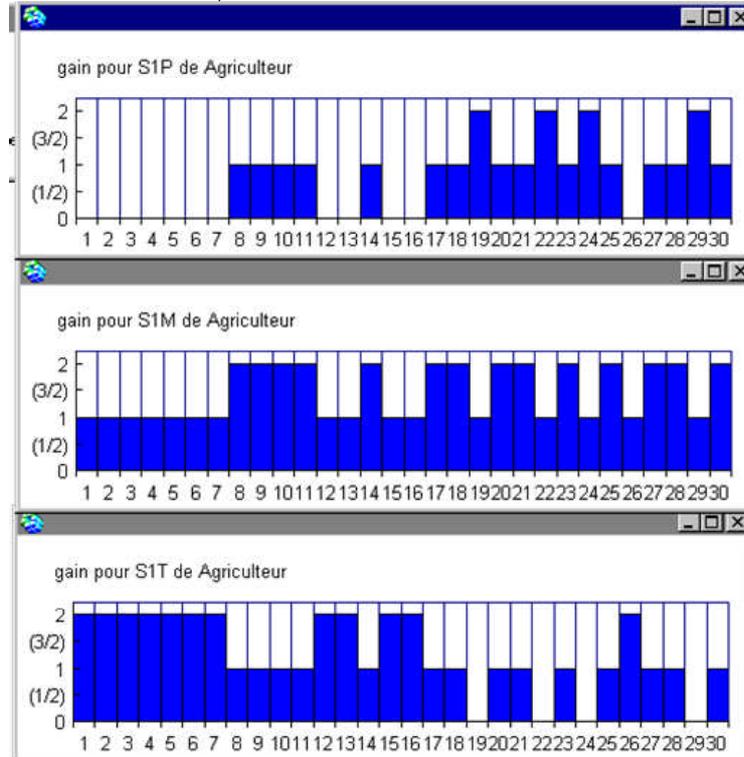


Figure 4 : Calibrage des saisons des pluies et simulation des gains sur 30 ans (Douentza).

4.3. L'appropriation du modèle par les paysans

La construction d'un modèle sur la base de modules mis en situation dans des JdR facilite son appropriation par les paysans qui participent au paramétrage du modèle pour leur région. A titre d'exemple, le module « association des variétés » a été construit en simplifiant la réalité du choix des variétés de sorgho à trois types, précoce, moyen et tardif, en fonction du calage de leur maturité avec la fin de saison des pluies qui peut être bonne, moyenne ou mauvaise. C'est sur ce dernier point que le paysan travaille avec nous sur le calibrage du modèle pour sa région. L'ordinateur dispose de séries climatiques sur 40 ans qui permettent de présenter une courbe des précipitations annuelles. Le travail se fait spécifiquement sur une zone (fig. 4) et les paysans d'autres régions réagissent s'ils ont un mode de raisonnement différent. Les paramètres de caractérisation de la saison climatique pris en compte par le paysan pour faire son choix de variété sont principalement : la date de début et de fin de saison des pluies et le cumul pluviométrique annuel. Ces indicateurs sur l'année climatique de 5 régions agricoles du Mali ont été décrits par les paysans afin de distinguer les bonnes, moyennes et mauvaises années pour la production agricole. Ces paramètres ont ensuite été appliqués, grâce au modèle, à des séries climatiques sur 30 ans (1969-1998) pour tester via des simulations la réponse du modèle aux choix des paysans (fig. 4). Cette simulation du gain pour une stratégie donnée permet de visualiser immédiatement des résultats sur un nombre d'année défini et de tester la robustesse du modèle, puis d'explorer et de comparer des stratégies entre elles.

Les jeux 2 et 3 s'appuyaient sur plusieurs résultats d'enquêtes : d'une part le morcellement d'une exploitation sur plusieurs sols augmente le nombre de variétés de sorgho cultivées dans l'exploitation et, d'autre part, la taille et le niveau d'équipement sont les déterminants de la structure des exploitations qui influencent en premier lieu le choix des variétés. Les facteurs de production étant ainsi fixés, l'ajustement se fait sur le choix des cultures et des variétés pour établir un assolement donné en fonction de l'année climatique. Même si les critères qui fixent les seuils entre grandes et petites exploitations, équipées ou non, diffèrent d'une région à l'autre, ce schéma reste valable quelle que soit la zone climatique. Grâce à l'articulation du SMA en modules, il est possible d'appliquer une stratégie particulière à une structure d'exploitation donnée (fig. 5). La simulation sur un certain nombre d'années permet alors de suivre les résultats de chaque agriculteur dans une stratégie particulière ou de se placer à l'échelle du village pour visualiser l'effet de la diversité des stratégies à un niveau plus global. C'est à ce niveau que la dynamique des variétés est suivie et analysée au travers des rôles joués par chaque type de paysan. Le modèle utilisé comme support de discussion permet de traiter chaque cas d'utilisation d'une variété de façon pratique

avec les paysans. En simulant sa réponse à la série climatique de la zone paramétrée par les paysans pour discuter ensemble, chercheurs et paysans débattent des meilleures variétés mais surtout des combinaisons possibles de variétés pour améliorer globalement la production et limiter le risque pour le paysan. La figure 5 prend en compte la typologie des exploitations d'un village moyen et analyse comment les contraintes structurelles de l'exploitation déterminent des choix variétaux spécifiques intégrés dans l'assolement et les rotations.

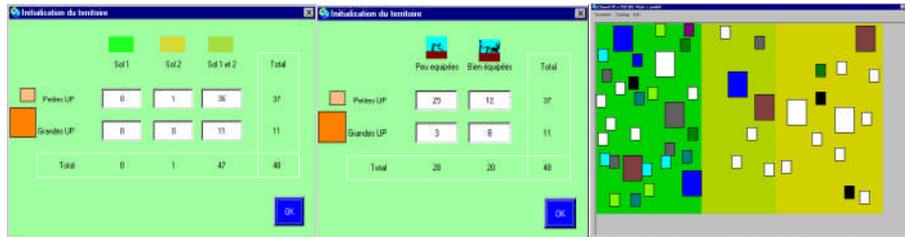


Figure 5 : Distribution des structures d'exploitations dans un village et, implémentation des stratégies d'association de variétés (cas de San).

5. CONCLUSION ET PERSPECTIVES : VERS L'ÉMERGENCE DE NOUVELLES RÈGLES DE GESTION

L'organisation des différentes sessions de jeux de rôles montre qu'il s'agit d'un outil bien adapté pour valider des hypothèses ou des résultats acquis par ailleurs et permettre de les discuter le cas échéant pour formuler de nouvelles hypothèses. La synthèse des données collectées et sa formalisation dans un jeu sont apparues comme une étape nécessaire pour partager notre représentation du système semencier avant d'aborder le modèle. Une bonne connaissance du terrain et de ses protagonistes facilite l'organisation du jeu mais aussi les modalités de son observation qui ne prennent sens qu'à la lumière des connaissances précédemment acquises sur le terrain. Le JdR permet une bonne mise en situation des paysans pour partager notre représentation de leur système informel d'échange de variétés. Les paysans apprécient beaucoup l'aspect de mise en scène de leurs pratiques dans le jeu de rôle qui favorise par la suite des discussions plus étroitement connectées à la réalité. Même si des contraintes ont été soulevées sur la mise en œuvre des jeux, il apparaît qu'ils demeurent dans notre cas un passage obligé pour passer à l'étape suivante et aborder le modèle multi-agents de façon sereine. En effet, les modèles informatiques isolés sont souvent illisibles et considérés comme des « boîtes noires » par des non-spécialistes c'est pourquoi il est difficile de communiquer autour d'eux. Leur usage conjoint

avec les jeux de rôle permet, d'une part, de rendre intelligible des modèles conceptuels lourds sous une forme ludique et, d'autre part, d'accélérer le déroulement des jeux (reproduction des résultats sur le long terme) ou d'explorer de nouveaux scénarios en effectuant des simulations [5].

Enfin, la finalité à moyen terme de la simulation multi-agents est de l'utiliser pour favoriser l'émergence de nouvelles règles favorables au maintien de la diversité des variétés de céréales. Le jeu « Seed-DIV » testé en avril 2007 avec des paysans issus de 5 régions du Mali et de 4 régions du Niger a montré la pertinence du modèle pour questionner de nouvelles règles de gestion. Le JdR a alors permis la confrontation de deux modèles de conservation, la banque de semence villageoise et la coopérative semencière. L'adhésion des paysans à notre démarche permet maintenant d'envisager un réel accompagnement de l'évolution des filières semencières informelles par la modélisation en prenant en compte de façon explicite un objectif de maintien de la diversité variétale *in situ*.

REMERCIEMENTS

L'application développée a été initié par le financement obtenu auprès du BRG† grâce à des études portées par le CIRAD‡ au sein de trois projets: le projet « *Agrobiodiversité du sorgho au Mali et au Burkina Faso* » financé par le FFEM§ (2002-2006), le projet « *Modélisation multi-agents des réseaux d'échange de semences pour améliorer la conservation in situ des variétés locales de céréales en Afrique de l'Ouest* » financé par le BRG (2005-2007) et le projet « *Vers une autonomie des agriculteurs sahéliens dans la gestion et l'exploitation de la diversité agricole, pour l'amélioration de leurs stratégies de subsistance* » financé par le FIDA** (2005-2007). Le Cirad a initié la démarche avec des chercheurs de l'Institut d'Economie Rurale en impliquant des ONG et des Organisations Paysannes (OP) locales. Mais le succès de ce projet tient surtout à la participation des paysans du Mali et du Niger. Que tous soient ici vivement remerciés de leur aide et du temps passé à échanger sur les modalités d'échanges et de conservation des semences.

† Bureau des Ressources Génétiques

‡ Centre de coopération Internationale en Recherche Agronomique pour le Développement

§ Fond Français pour l'Environnement Mondial

** Fonds International pour le Développement Agricole

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Abrami, G., Bazile, D., Le Page, C., Dembele, S., Dionnet, M., 2005. Preparing a framework for participatory modelling of farmers seed systems in Mali: varieties selection and exchange. *In: ESSA (Ed.), The third annual conference of the European Social Simulation Association*. ESSA, Koblenz-Landau University on September 5-9, 2005, 7 p.
- [2] Almekinders C.J.M., Louwaars, N.P. and Bruijn, G.H.d., 1994. Local seed systems and their importance for an improved seed supply in developing countries. *Euphytica*, 78: 207-216.
- [3] Almekinders C.J.M., Louwaars, N.P. 2002. The importance of the farmer's seed systems in a functional national seed sector. *Journal of new Seeds*, 4/1-2: 15-33.
- [4] Antona M. *et al*, 2003. Our Companion Modelling Approach. *Journal of Artificial Societies and Social Simulation*, 6-1, <http://jasss.soc.surrey.ac.uk/6/2/1.html>.
- [5] Barreteau O., 2003. The joint use of role-playing games and models regarding negotiation processes: characterization of association. *Journal of Artificial Societies and Social Simulation*, 6-2, <http://jasss.soc.surrey.ac.uk/6/2/3.html>.
- [6] Bazile D., Soumare M., 2003. La valorisation de l'agrobiodiversité pour assurer une production agricole durable. Le cas du sorgho dans la zone cotonnière de Koutiala au Mali. *In: ICRISAT and IPALAC (Ed.), International Symposium for Sustainable Dry land Agriculture Systems*, December 2-5, 2003. ICRISAT, Niamey (Niger), pp. 261-269.
- [7] Bazile D., Soumare M., 2004. Pratiques paysannes de gestion de la diversité variétale en réponse à la diversité écosystémique. Le cas du sorgho [*Sorghum bicolor* (L) Moench] au Mali. Étude Originale. *Cahiers Agricultures*, 13: 480-487.
- [8] Bazile D., *et al*, 2005. Perspectives of modelling the farmer's seed system for *in situ* conservation of sorghum varieties in Mali. *In: J.B. Cunha and R. Morais (Editeurs), EFITA/WCCA 2005 Joint Conference: 5th Conference of the European Federation for Information Technology in Agriculture, Food and Environment and 3rd World Congress on Computers in Agriculture and Natural Resources*. Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal, 25-28 July 2005, pp. 39-46.
- [9] Bousquet F. *et al*, 2002. Multi-agent systems and role games: collective learning processes for ecosystem management. *In M.A. JANSSEN (Ed.), Complexity and ecosystem management*, Edward Elgar Publishing, pp. 248-285.
- [10] Bousquet F. et Le Page, C., 2004. Multi-agent simulations and ecosystem management: a review. *Ecological Modelling*, 176: 313-332.
- [11] Brookfield H. (Ed.), 2001. *Exploring agrobiodiversity*. Issues, Cases, and Methods in Biodiversity Conservation Series. Columbia University Press, New York (USA), 348 p.
- [12] Brookfield H., *et al*, M. (Ed.), 2002. *Cultivating biodiversity*. ITDG Publishing and United Nations University, London (UK), 292 p.
- [13] FAO, 1999. *The state of the world of Plant genetic resources for food and agriculture (PGRFA)*. FAO, Rome (I), 511 p.
- [14] Soumare M., 2004. *Contribution à la prévision de l'aire de diffusion de variétés de sorgho au Mali. Couplage entre Modèle de Croissance des Cultures et Système d'Information Géographique*. DEA Géographie, Université de Nanterre, Paris X, 92 p.

- [15] UNEP, 1993. *Convention on Biological Diversity*. Text and Annexes. CBD/94/1. UNEP/CBD, Montreal, 34 p.
- [16] Vaksman M., Tradore S.B. et Niangado O., 1996. Le photopériodisme des sorghos africains. *Agriculture et Développement*, 9: 13-18.
- [17] Wood D. et Lenne J.M. (Editeurs), 1999. *Agrobiodiversity: Characterization, utilization and Management*. CABI Publishing, Wallingford (UK), 490 p.

Analyse comparative du polymorphisme des gènes *eIF4E* chez les plantes : conservation *versus* sélection positive exercée par les virus

Carole CARANTA^{(1)*}, Jacques DAVID⁽²⁾, Carine CHARRON⁽¹⁾,
Stéphane DE MITA⁽²⁾, Pascale SANCHEZ⁽¹⁾, Sylvain SANTONI⁽²⁾,
Alberto CENCI⁽²⁾, Joëlle RONFORT⁽²⁾, Audrey WEBER⁽³⁾,
Anne-Françoise ADAM-BLONDON⁽³⁾, Dominique BRUNEL⁽⁴⁾,
Nathalie CHANTRET⁽²⁾

⁽¹⁾ INRA-UR1052, Unité de Génétique et Amélioration des Fruits et Légumes,
Domaine Saint Maurice, BP94, 84143 Montfavet Cedex, France

⁽²⁾ UMR1097-UMR INRA, U Montpellier II, ENSAM, IRD - Diversité et
Adaptation des Plantes Cultivées, Domaine de Melgueil, 34130 Mauguio, France

⁽³⁾ UMR1165-UMR INRA, CNRS, UEVE - Recherche sur les Génomes Végétaux,
2 rue G. Crémieux, BP5708, 91057 Evry Cedex, France

⁽⁴⁾ INRA-CNG, 2 rue G. Crémieux, CP5701, 91057 Evry, France

Abstract: Comparative analysis of the DNA polymorphism at *eIF4E* genes in plants: conservation *versus* positive selection exerted by Potyviruses. According to data obtained on the key role of the translation initiation factor eIF4E in plant-RNA virus interactions, the aim of this project is to analyse and compare the polymorphism of *eIF4E* genes in five plant families infected by RNA viruses and for which a role of eIF4E in the viral infectious cycle was demonstrated; the final objective being to detect selection pressure exerted by viruses. Overall, *eIF4E* and *eIF(iso)4E* genes were subjected to a strong purifying selection in all species except pepper and grapevine for which the *eIF4E1* gene is submitted to positive selection. In pepper, functional analysis show that amino acids under positive selection correspond to those involved in virus resistance. The fact that this positive selection is not conserved among genus from a same plant family suggest that the biological features of interaction with viruses "*in natura*" led to distinct pattern of evolution for *eIF4E* genes. These results together with those showing a conserved position of amino acid mutations in eIF4E resistance proteins from unrelated species, suggest a convergent mode of evolution of some *eIF4E* genes toward resistance.

Eukaryotic translation initiation factor 4E (eIF4E) / Virus resistance / Polymorphism / Molecular evolution / Selection

Résumé : A partir des connaissances acquises sur le rôle généraliste des facteurs d'initiation de la traduction eIF4E dans les interactions plante-virus à ARN, ce

* Correspondance et tirés à part : carole.caranta@avignon.inra.fr

projet vise à décrire et comparer le polymorphisme des gènes *eIF4E* chez 5 familles végétales infectées par des virus à ARN pour lesquels un rôle de eIF4E dans le cycle infectieux a été démontré et ce, avec l'objectif de détecter des indices de sélection liés à l'interaction avec les virus. Globalement, les gènes *eIF4E* et *eIF(iso)4E* sont soumis à une forte sélection purificatrice, excepté chez le piment et la vigne, chez lesquels le gène *eIF4E1* apparaît soumis à de la sélection positive. Chez le piment, les analyses fonctionnelles montrent que les acides aminés sous sélection positive correspondent à ceux impliqués dans la résistance. Cette pression de sélection n'est pas conservée entre les genres d'une même famille, suggérant que les caractéristiques biologiques d'interaction avec les virus « *in natura* » conduisent à des profils différents d'évolution des gènes *eIF4E*. Ces résultats associés à ceux montrant la conservation de la position des substitutions en acides aminés dans les protéines eIF4E de résistance chez des espèces très éloignées suggèrent l'existence d'une évolution convergente de certains gènes *eIF4E* vers la résistance.

Facteur d'initiation de la traduction 4E (eIF4E) / Résistance aux virus / Polymorphisme / Evolution moléculaire / Sélection

1. INTRODUCTION

Les ressources génétiques constituent le réservoir de diversité nécessaire à l'adaptation des variétés aux changements des usages et à l'évolution des climats et des milieux. Les allèles de résistance aux bioagresseurs sont depuis toujours des cibles de choix dans l'évaluation des collections de ressources génétiques. L'approche traditionnelle est le plus souvent empirique, les géniteurs intéressants étant repérés par leur capacité de tolérance ou de résistance. Cette approche a été révolutionnée par les progrès réalisés dans la connaissance des mécanismes moléculaires d'interaction entre les agents pathogènes et leurs plantes hôtes, par l'accès aux gènes impliqués dans la résistance et par la description de leur polymorphisme [8], [9]. Par ailleurs, la génétique des populations et la génétique évolutive ont fait de grands progrès dans les concepts utilisés pour détecter des traces spécifiques de la sélection à partir du polymorphisme de séquence des gènes [17], [19].

Ces dernières années, de nombreuses études ont montré le rôle des facteurs d'initiation de la traduction eIF4E (eukaryotic Initiation Factor 4E) dans la résistance récessive de nombreuses espèces végétales aux virus à ARN, définissant ainsi une nouvelle catégorie de gène de résistance [8], [11].

eIF4E est un élément clé de la cellule eucaryote: *via* une interaction directe avec la coiffe des ARN, c'est le premier facteur qui initie la traduction des ARNm. Il interagit également avec d'autres facteurs eIF pour permettre la circularisation de l'ARN et le recrutement de la sous-unité 40S du ribosome. Chez les plantes, eIF4E appartient à une petite famille multigénique : 5 gènes ont été identifiés chez *Arabidopsis thaliana*, dont 3

codent pour des protéines correspondant à l'isoforme eIF4E, un pour eIF(iso)4E et un pour la protéine nCBP (new Cap Binding Protein). Le rôle biologique de cette redondance génique reste encore à déterminer.

La majorité des résistances récessives contrôlées par eIF4E sont efficaces vis-à-vis de virus à ARN de la famille des *Potyviridae*, dont le génome se caractérise entre autre par la présence d'une protéine nommée VPg liée à l'extrémité 5' de l'ARN viral. Cette protéine (ou son précurseur, la NIa) interagit avec certaines isoformes de eIF4E et des substitutions en acides aminés dans sa région centrale sont responsables du contournement des résistances médiées par eIF4E [8], [11]. Du côté de l'hôte, la résistance résulte d'un petit nombre de substitutions en acides aminés dans eIF4E et la position de la plupart de ces substitutions « signatures » des allèles de résistance est conservée entre espèces et restreinte à deux régions adjacentes, localisées en surface de la protéine, nommées région I et II. Dans certains cas, la résistance résulte de la combinaison de mutations dans les protéines eIF4E1 et eIF(iso)4E [15]. Le mécanisme sous-jacent aux résistances contrôlées par eIF4E est encore assez mal connu, mais des résultats récents montrent que les substitutions identifiées dans les protéines de résistance modifient/abolissent l'interaction entre eIF4E et la VPg [3], [6].

Dans ce contexte, l'objectif est de décrire et comparer le polymorphisme des gènes codant pour eIF4E chez 5 familles végétales incluant espèces modèles et espèces cultivées, afin de détecter des traces de sélection liées à l'interaction avec les virus et, le cas échéant, déterminer les caractéristiques à l'origine de la variation observée. Cette étude a été réalisée chez des Solanacées (*Capsicum* spp. et *Lycopersicon* spp.), des Triticées (blés diploïdes et apparentées), des Légumineuses (*Medicago truncatula* et apparentées), des Vitacées (*Vitis vinifera* et apparentées) et chez *Arabidopsis thaliana*. Ces familles comportent toutes plusieurs espèces qui sont infectées par différents virus à ARN pour lesquels un rôle de eIF4E dans le cycle infectieux a été démontré.

2. MATÉRIELS ET MÉTHODES

2.1. Collections de génotypes

Les collections de génotypes représentatifs de différents niveaux de diversité préalablement établies sur la base d'information de séquences ou de marqueurs microsatellites suivantes ont été utilisées :

- Solanacées: Les 25 génotypes de *Capsicum annuum* décrits dans [3] représentent l'échantillon intraspécifique. Pour les analyses interspécifiques, 26 accessions de 11 espèces représentatives des

Capsicum domestiqués et des espèces sauvages apparentées ont été utilisées. Pour la tomate, les analyses intraspécifiques ont été réalisées chez 20 génotypes de *Lycopersicon esculentum* et les analyses interspécifiques chez 30 accessions représentatives des 8 espèces sauvages apparentées.

- *Arabidopsis*: Les 48 écotypes de la core collection établie par McKhann *et al.* [10] ont été utilisés. L'espèce *A. lyrata* a été utilisée comme groupe externe afin d'estimer la divergence intraspécifique.
- *Medicago truncatula* et espèces apparentées : L'échantillon étudié comprend (i) la core collection *M. truncatula* de 32 individus, (ii) un échantillon d'une trentaine d'individus représentant un groupe génétique plus homogène identifié à l'aide de marqueurs microsatellites au sein de la collection [12], (iii) une dizaine de lignées représentant les espèces sœurs de *M. truncatula* : *M. littoralis* et *M. tricycla*, et (iv) un ensemble de 12 lignées représentants d'autres espèces du genre.
- Triticées diploïdes: L'échantillon d'étude est constitué de 97 génotypes appartenant à 30 espèces diploïdes de la tribu des triticées (genres *Triticum*, *Aegilops*, *Hordeum*, *Agropyrum*, *Australopyrum*, *Brachypodium*, *Dasyphyrum*, *Eremopyron*, *Henrardia*, *Heteranthelium*...).
- Vitacées : L'échantillon étudié comprend (i) une core collection de 48 individus de *Vitis vinifera* ssp. *sativa* obtenue par l'équipe Génétique de la Vigne de l'UMR DIAPC, (ii) 16 vignes sauvages ou lambrusques (*Vitis vinifera* ssp. *silvestris*) et (iii) 13 individus correspondant à différentes espèces et genres de la famille des Vitacées.

2.2. Analyse des données de séquence

Les séquences sens et antisens ont été alignées et corrigées manuellement avec les logiciels Genalys (<http://software.cng.fr>), BioEdit et/ou Staden. Les sites nucléotidiques polymorphes ont été détectés par Genalys. La séquence peptidique a été prédite par BioEdit. Pour chaque position polymorphe, les fréquences de chaque variant nucléotidique et les estimations du polymorphisme (π , θ_w) ont été obtenues avec le logiciel DnaSP version 4. Le test de neutralité du D de Tajima a été réalisé avec le logiciel DnaSP. Les probabilités (P-values) d'obtenir les D de Tajima observés ont été estimées sur la base de 10 000 simulations de coalescence, sous l'hypothèse d'absence d'évènement de recombinaison et dans une population de taille constante. Les phylogénies ont été réalisées avec le

logiciel MEGA3 (les arbres ont été construits selon la méthode Neighbour-joining). Les profils de divergence de séquence aux locus *eIF4E* ont été analysés en utilisant des modèles de substitution des codons [5] qui contraste la divergence synonyme et non synonyme. Différents modèles sont testés, incluant ou non la possibilité que certains sites soient sous sélection positive. La comparaison des valeurs de vraisemblance permet de déterminer quel modèle est le plus probable. La recherche de sites sous sélection positive a été effectuée avec le logiciel codelm du package PAML version 3.14.

3. PRINCIPAUX RÉSULTATS ET DISCUSSION

3.1. Obtention des séquences

Les gènes *eIF4E* de plantes sont constitués de 5 exons et 4 introns. La conservation de la jonction exon-intron observée chez les eucaryotes et la proportionnalité de la taille de l'intron 1 par rapport à la taille du génome de l'espèce permet de prédire, d'une part, la position des introns et, d'autre part, la taille approximative de l'ADNg [14]. Pour les Solanacées, *Arabidopsis* et *Medicago*, les échantillons ont été séquencés pour d'autres fragments génomiques (10 à 30 fragments anonymes répartis sur le génome) afin de nous renseigner sur les patrons de polymorphisme intraspécifique « neutre » et sur les niveaux de divergence entre espèces.

- *Capsicum* spp. et *Lycopersicon* spp. : Trois gènes *eIF4E* ont été clonés et séquencés à partir de séquences EST de tomate : 2 gènes codant pour une protéine de type *eIF4E* (*eIF4E1* et *eIF4E2*) et 1 codant pour l'isoforme *eIF(iso)4E*. Les ADNc des trois gènes ainsi que les séquences des introns 2, 3 et 4 ont été obtenus pour l'ensemble des génotypes.
- *Arabidopsis thaliana* : Les séquences ADNg des 5 gènes *At-eIF4E* (*eIF4E1*, *eIF4E2*, *eIF4E3*, *eIF(iso)4E* et *nCBP*) ont été obtenues à partir de la séquence du génome d'*A. thaliana* avec des amorces définies dans les régions 5' et 3' UTR.
- *Medicago truncatula* et espèces apparentées : A partir des séquences génomiques complètes des gènes *eIF(iso)4E* (BAC AC174281) et *eIF4E* (BAC AC183923), plusieurs couples d'amorces ont été définis et testés de façon à permettre d'amplifier et de séquencer l'ensemble des cinq exons des deux gènes, accompagnés de fragments d'introns et d'une partie des régions 5' et 3' UTR. Pour *eIF4E*, la séquence de l'ensemble du gène (introns et exons, environ 3500 pb) a été obtenue sur l'ensemble des accessions de *M.*

truncatula. La séquence codante uniquement a été obtenue pour les autres espèces du genre *Medicago* (environ 700 bp). Pour *eIF(iso)4E*, trois fragments ont été obtenus couvrant les 5' et 3' UTR, la totalité des exons ainsi que les introns 2, 3 et 4; seule une séquence partielle de l'intron 1 a pu être obtenue.

- Triticées diploïdes : Deux clones BAC positifs pour *eIF(iso)4E* ont été identifiés et isolés d'une banque de blé dur. Les deux clones ont été assignés aux deux génomes du blé dur en comparant leur séquences à celles obtenues des espèces diploïdes parentales (c'est-à-dire *T. urartu* espèce donatrice du génome A et *Aegilops speltoides* espèce la plus proche au génome B). Les régions contenant le gène cible ont été séquencées. La structure générale du gène est conservée entre les deux génomes : 5 exons et 4 introns de tailles similaires entre les deux clones. Cependant, dans la région en amont des gènes il y a une insertion/délétion de 190 pb. Un seul clone BAC positif (485) pour *eIF4E* a été isolé dans la banque BAC de blé dur et assigné au génome A. Un autre clone BAC contenant la copie du génome B a été isolé d'une banque BAC spécifique du chromosome 3B du blé tendre, ce qui confirme la position de *eIF4E* sur les chromosomes du groupe 3 des Triticeae. Les régions contenant les gènes cibles ont été séquencées. La structure générale du gène *eIF4E* du blé est semblable à celle des autres espèces et à celle de *eIF(iso)4E*, c'est à dire 4 exons et 5 introns.
- Vitacées : La vigne est une espèce hautement hétérozygote, ce qui peut rendre difficile le séquençage de produits PCR obtenus à partir d'ADN génomique. Pour évaluer le taux de polymorphisme dans ces gènes et faciliter leur séquençage, une banque BAC Cabernet Sauvignon a été criblée avec des amorces spécifiques des gènes *eIF4E* et *eIF(iso)4E* [1] puis les BAC portant chacun des allèles aux deux gènes ont été séquencés. L'utilisation des liens entre carte physique et carte génétique, puis des liens entre carte génétique et la séquence du génome (www.genoscope.cns.fr/vitis) ont permis de positionner le gène *eIF(iso)4E* sur le groupe de liaison 5 entre les marqueurs génétiques UDV-053 et VVIV21 (gène n°GSVIVT00020333001). Le gène *eIF4E* (n°GSVIVT00007223001) est en revanche sur un contig de séquence actuellement encore non positionné sur la carte génétique. Ces deux gènes sont bien uniques : le BLATP des deux protéines contre le protéome de *Vitis vinifera* les alignent contre elles-mêmes et l'une contre l'autre. Deux hits supplémentaires, extrêmement partiels sont trouvés pour *eIF(iso)4E* dont l'un dans l'intron d'un autre gène. La famille semble donc peu complexe chez la vigne.

- L'alignement des séquences obtenues pour les deux allèles *eIF4E* du Cabernet Sauvignon met en évidence plusieurs SNPs dont 3 conduisent à un changement d'acides aminés. Pour *eIF(iso)4E*, aucun polymorphisme dans les exons n'est observé. L'importance du polymorphisme observé (1 SNP ou Indel respectivement tous les 99 bp et 69 bp) ainsi que la longueur des introns 1 de *eIF4E* et 3 de *eIF(iso)4E* ont conduit à réaliser le séquençage de ces gènes à partir d'ADNc.

L'ensemble de ces séquences a été déposées dans la base de données EvoluTIF (Evolutionary dynamics of Translation Initiation Factors in plants - <http://www.avignon.inra.fr/evolatif/index.php?page=accueil>) créée dans la cadre du projet.

3.2. Recherche des pressions de sélection

3.2.1. Co-évolution entre le facteur *eIF4E1* du piment et la VPg des potyvirus

L'analyse du polymorphisme des gènes *eIF4E1*, *eIF4E2* et *eIF(iso)4E* chez *C. annuum* montre que seul *eIF4E1*, correspondant au locus *pvr2* de résistance aux potyvirus, présente des traces de sélection positive : parmi les 25 accessions de *C. annuum* représentatives de la diversité de cette espèce, 10 variants alléliques ont été identifiés au locus *pvr2-eIF4E1* et de façon assez remarquable, toutes les substitutions nucléotidiques sont non-synonymes et localisées dans les régions I et II de la protéine, régions préalablement montrées comme étant impliquées dans la résistance aux virus. La plupart de ces acides aminés sont sous pression de sélection positive (ω compris entre 9 et 10 ; $P > 0,95$) et les analyses fonctionnelles ont montré qu'ils correspondaient aux résidus impliqués dans la sensibilité *versus* résistance au PVY, suggérant une pression de sélection positive exercée par les potyvirus [3]. Ces résultats couplés à ceux montrant que les acides aminés de la protéine virale VPg (également sous sélection positive) permettent aux potyvirus de s'adapter aux protéines *eIF4E* de résistance, mettent en évidence une coévolution entre *eIF4E1* du piment et la VPg des potyvirus [3].

Récemment, il a été proposé qu'un facteur important régissant l'évolution des gènes dominants de résistance aux agents pathogènes était le mode de reconnaissance, direct ou indirect, entre la protéine de résistance et la protéine d'avirulence [4]. Dans le cas des reconnaissances indirectes faisant intervenir une protéine de garde, les gènes R présentent une faible diversité allélique maintenue par de la sélection balancée. A l'inverse, dans le cas des reconnaissances directes, une coévolution de type « course aux armements » est retrouvée, avec le maintien d'une forte diversité allélique pour le gène de résistance et le gène d'avirulence. Ainsi, dans le cas d'une

interaction directe, l'agent pathogène peut contourner la résistance en modifiant de façon minimale sa protéine Avr pour échapper à la reconnaissance par la protéine de résistance et *vice versa* pour la protéine R. Ce type de reconnaissance est donc caractérisé par de la sélection diversifiante à la fois chez les protéines R et Avr.

Les cas qui se rapprochent le plus de ce que nous observons pour le gène *eIF4E1* du piment sont ceux des gènes de résistance *Pto* de la tomate et *L* du lin, qui présentent un niveau de substitutions non-synonymes élevé et un grand nombre de variants alléliques résultant d'une pression de sélection pour la diversification [4], [13]. L'évolution de *Pto* semble être influencée à la fois par de la sélection purificatrice et de la sélection diversifiante, comme *eIF4E1* chez le piment. Cependant, une pression de sélection balancée est à l'origine du maintien de la variété allélique au locus *Pto*, ce qui ne semble pas être le cas pour *eIF4E1* de *Capsicum* où aucun polymorphisme synonyme adjacent aux polymorphismes non-synonymes n'a été détecté. Dans le cas du couple lin-*Melampsora lini*, des analyses portant à la fois sur le polymorphisme des gènes de résistance et de virulence de l'agent pathogène qui interagissent directement ont été réalisées. Des traces de sélection positive sont retrouvées à la fois chez les locus *L* et *Avr*, comme dans le cas du pathosystème piment-potyvirus. Ces interactions directes entre protéines de résistance et protéines d'avirulence caractérisent une évolution rapide de type « course aux armements ». Dans le cas de la résistance récessive aux virus à ARN, la résistance est la conséquence de l'incapacité pour le virus d'interagir avec une protéine de l'hôte qui lui est indispensable pour accomplir son cycle infectieux. A l'inverse du mécanisme « actif » des résistances dominantes, le virus doit donc muter pour rétablir l'interaction avec la protéine de l'hôte et non pour échapper à l'interaction. Le parasitisme obligatoire des virus à ARN, comme celui de l'oomycète *M. lini*, implique que les protéines VPg de virulence coévoluent dans un système clos avec les protéines de résistance *eIF4E* de la plante hôte, ce qui impose une course évolutive rapide à la fois chez l'hôte et le pathogène. Cependant, un pré-requis pour une telle issue évolutive est que les protéines de virulence et de résistance puissent accumuler des mutations qui affectent l'interaction sans imposer un coût de fitness trop important pour la plante ou pour le pathogène.

3.2.2. Le profil d'évolution du gène *eIF4E1* impliqué dans la résistance aux potyvirus n'est pas conservé chez les Solanacées

La tomate, Solanacée proche du piment, est également infectée par plusieurs potyvirus et le gène *eIF4E1* a également été démontré comme étant impliqué dans la résistance récessive (gène *pot1-eIF4E1*). Cependant, en comparaison avec le piment, une fréquence beaucoup plus faible de

gènes récessifs de résistance aux potyvirus a été observée dans la variabilité naturelle (~1% contre ~40% chez le piment).

L'analyse du polymorphisme des 3 gènes *eIF4E* chez le genre *Lycopersicon* montre que les trois gènes sont très contraints. Le gène *eIF4E1* ne présente qu'un seul site sous sélection positive qui apparaît spécifique de l'espèce *L. pimpinellifolium*. L'ensemble des résultats obtenus ne permettent pas de détecter de traces de sélection en lien avec l'interaction avec les virus. Cette absence de pression de sélection pourrait s'expliquer par le fait que, contrairement au piment, les potyvirus n'ont qu'une incidence faible sur les cultures de tomate. La majorité des souches de potyvirus n'induisent que des symptômes de mosaïque ou de marbrures discrets sur les feuilles de tomate, même si certains isolats apparus dans les années 1980 sont responsables de symptômes plus sévères (croissance des plantes ralentie, production en fruits diminuée). De plus, chez la tomate, les potyvirus sont quasiment systématiquement retrouvés en infections mixtes avec d'autres genres viraux, comme les cucumovirus.

3.2.3. A. thaliana ne constitue pas un modèle adapté à l'étude de la coévolution plante-potyvirus

L'analyse du polymorphisme des gènes *eIF4E* a été conduite en parallèle chez *Arabidopsis* pour deux raisons : (i) c'est la seule dicotylédone pour laquelle nous disposions, au moment du démarrage du projet, de l'ensemble des membres de la famille *eIF4E* et (ii) curieusement et contrairement aux Solanacées, aucune des résistances récessives (pourtant nombreuses) aux potyvirus identifiées dans la variabilité naturelle de cette espèce ne correspond à un gène *eIF4E*.

Les analyses de polymorphisme montrent que les gènes *At-eIF4E1*, *At-eIF4E3*, *At-eIF(iso)4E* et *nCBP* sont sous pression de sélection purificatrice (polymorphisme très réduit et substitutions non-synonymes très inférieures aux substitutions synonymes). En revanche, le gène *At-eIF4E2*, présentant plus de 95 % d'identité avec *At-eIF4E3* et séparé de 7 kb de ce dernier, présente un polymorphisme beaucoup plus élevé et une valeur de D de Tajima très positive, suggérant qu'une pression de sélection balancée s'exercerait sur ce gène. Chacun des membres de la famille multigénique apparaît donc maintenu par la sélection argumentant en faveur d'une spécialisation fonctionnelle.

Ces résultats ne permettent pas de détecter des traces de sélection en lien avec l'interaction avec les potyvirus et confirment qu'aucun gène récessif de résistance ne semble correspondre à *eIF4E* chez *A. thaliana*. Ces résultats peuvent s'expliquer par le fait qu'*A. thaliana* ne constitue pas un hôte préférentiel pour les potyvirus. Walsh et Jenner [18] insistent sur l'absence de gènes de résistance au TuMV (potyvirus inféodé aux Brassicacées) dans la

variabilité naturelle d'*A. thaliana*, suggérant que ce virus n'infecte que très rarement cette espèce et/ou que le maintien de gènes de résistance représente un coût pour la plante. *A. thaliana*, espèce modèle dans de nombreux domaines d'études en biologie et excellent pathosystème de laboratoire, ne constitue donc pas un modèle pertinent d'étude de la coévolution plante-potyvirus.

3.2.4. Le profil d'évolution des gènes *eIF4E* n'est pas conservé chez les Légumineuses

L'analyse du polymorphisme des gènes *eIF4E* et *eIF(iso)4E* a été conduite chez le genre *Medicago* car de précédentes études mettaient en évidence l'implication de ces gènes dans la résistance aux potyvirus chez d'autres espèces de Légumineuses : (i) chez *Pisum sativum*, plusieurs allèles récessifs de résistance aux potyvirus codent pour des protéines *eIF4E* ou *eIF(iso)4E* ; (ii) chez *Phaseolus vulgaris*, des analyses génétiques indiquent que le locus récessif *desc* de résistance au CIYVV pourrait correspondre à un facteur *eIF4E* [8].

Les analyses de séquence des gènes *eIF4E* menées chez des espèces autogames diploïdes du genre *Medicago* ne permettent pas d'identifier de variants alléliques *eIF4E* ou *eIF(iso)4E* potentiellement impliqués dans la résistance aux potyvirus. De plus, que ce soit au niveau intraspécifique ou interspécifique, aucune pression de sélection positive sur les gènes *eIF4E* et *eIF(iso)4E* n'a été identifiée. Aucun des gènes *eIF4E* chez *Medicago* ne semble donc impliqué dans les processus évolutifs liés aux interactions entre les protéines *eIF4E* et les potyvirus.

De façon assez similaire à ce que l'on observe chez les Solanacées, le profil de polymorphisme des gènes *eIF4E* semble variable selon les espèces de Légumineuses étudiées : toute une série de variants alléliques impliqués dans la résistance aux potyvirus identifiée chez le pois, un seul allèle récessif de résistance détecté chez le haricot à ce jour et *a priori* absence de résistance récessive aux potyvirus correspondant aux gènes *eIF4E* chez *Medicago*. Des analyses de polymorphisme des gènes *eIF4E* chez le pois et le haricot permettraient de vérifier si on retrouve les mêmes traces de sélection positive que celles identifiées pour le gène *eIF4E1* chez le piment.

3.2.5. Le polymorphisme du gène *eIF4E* chez le blé se distingue du polymorphisme détecté chez l'orge

L'orge et le blé, de la tribu des Triticées, sont fréquemment infectés par des virus à ARN du genre *Bymovirus* appartenant aussi à la famille des *Potyviridae*. Comme chez les potyvirus, une protéine VPg est liée de façon covalente à l'extrémité 5' de l'ARN des *bymovirus*. Chez *Hordeum vulgare*, le gène récessif *rym4/5* de résistance à plusieurs souches de *bymovirus* code pour une

protéine eIF4E [8]. Une première analyse sur 56 accessions d'orge avait permis d'identifier 7 variants alléliques au locus *eIF4E-rym4/5* se différenciant par 9 substitutions nucléotidiques exclusivement non-synonymes. Une étude plus récente portant sur un échantillon de 127 accessions d'orge cultivé et 4 accessions sauvages a permis de confirmer l'importante variété allélique présente à ce locus et l'identification d'un nombre important de mutations non-synonymes [16].

A partir de ces résultats et afin de comparer le profil d'évolution des gènes *eIF4E* au sein de la tribu des *Triticeae*, les analyses de polymorphisme de ces gènes ont été conduites chez des espèces diploïdes de blé. Chez *T. monococcum* et *T. urartu*, les gènes *eIF4E* et *eIF(iso)4E* apparaissent soumis à une forte pression purificatrice. Cependant, la recherche de sites sous sélection sur un échantillon interspécifique révèle que 4 sites dans la séquence codante de *eIF(iso)4E* présentent un ω supérieur à 1, avec notamment le codon en position 120 dans l'exon 2 qui présente un ω supérieur à 4 avec une probabilité de 0,994. Ce site ne varie pas uniquement dans un sous-ensemble d'espèces mais dans toute la population, sans cohérence avec la taxonomie. De plus, des analyses de phylogénie des gènes *eIF4E* réalisées sur l'échantillon interspécifique de *Triticées* diploïdes montrent que l'évolution des gènes *eIF4E* du blé se distingue de l'évolution des gènes *eIF4E* de l'orge. D'autres voies pourraient être utilisées chez le blé en réponse aux infections virales. En effet, chez le riz, il a été montré que c'était le facteur eIF4G et non pas eIF4E qui était impliqué dans la résistance [2]. Ainsi, ce scénario pourrait être retrouvé chez d'autres monocotylédones, comme le blé où les gènes *eIF4E* ne présentent pas de polymorphisme caractéristique des gènes récessifs de résistance aux virus à ARN.

3.2.6. Une pression de sélection positive s'exerce sur le gène *eIF4E* chez la vigne

Deux virus du genre *Nepovirus* (Grapevine fanleaf virus, GFLV et ArMV, famille des *Comoviridae*) sont responsables de la maladie du Court-Noué chez la vigne, une des maladies les plus importantes chez cette espèce. Le domaine protéase de la NIa (forme précurseur de la VPg) des *Nepovirus* interagit *in vitro* avec la protéine eIF(iso)4E d'*A. thaliana*, suggérant un rôle potentiel des facteurs eIF4E dans l'interaction plante-nepovirus [7].

Le gène *eIF(iso)4E* apparaît soumis à une intense sélection purificatrice. Par contre, les analyses de polymorphisme de l'ADNc *eIF4E* chez les 48 accessions de *Vitis vinifera* ssp. *sativa* ont permis de mettre en évidence un excès de substitutions non-synonymes par rapport aux substitutions synonymes ($\pi_a \gg \pi_s$) et un signal très fort de sélection positive : 11 % des sites de la protéine eIF4E présentent un ω supérieur à 5, rappelant les traces

de sélection positive détectées au niveau du facteur eIF4E1 chez le piment. De plus, plusieurs des substitutions non-synonymes détectées se retrouvent dans les régions de la protéine eIF4E montrées comme étant impliquées dans la résistance aux potyvirus chez d'autres espèces. L'impact des substitutions en acides aminés identifiées sur la résistance aux nepovirus est en cours d'analyse (individus en cours d'autofécondation pour avoir l'ensemble des variants alléliques à l'état homozygote afin de réaliser des tests de résistance aux nepovirus). Ces résultats pourraient ouvrir la voie à une nouvelle stratégie de lutte contre les nepovirus chez la vigne.

3.3. Vers un modèle d'évolution convergente pour la résistance des gènes *eIF4E*

Parmi les genres étudiés à ce jour, de la sélection positive sur les gènes *eIF4E* n'a été retrouvée que dans le cas du gène *eIF4E1* chez le piment et la vigne. Chez le piment, les analyses fonctionnelles montrent que les acides aminés sous sélection positive correspondent à ceux impliqués dans la résistance [3]. Il apparaît clairement que cette pression de sélection n'est pas conservée entre les genres de la famille des Solanacées, genres pourtant tous infectés par les potyvirus et dont certains possèdent des gènes récessifs de résistance aux potyvirus correspondant à eIF4E, suggérant que les caractéristiques biologiques d'interaction avec les (poty)virus « *in natura* » conduisent à des profils différents d'évolution des gènes *eIF4E*. Plus globalement, ces résultats associés à ceux montrant la conservation de la position des substitutions en acides aminés dans les protéines eIF4E de résistance d'espèces très éloignées - piment, laitue, pois, orge -, suggèrent l'existence d'une évolution convergente des gènes *eIF4E* vers la résistance, qui pourrait être conditionnée par une pression de sélection importante exercée par les virus.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Adam-Blondon A.F., Bernole A., Faes G., Lamoureux D., Pateyron S., Grando M.S., Caboche M., Velasco R., Chalhoub B., 2005 – Construction and characterization of BAC libraries from major grapevine cultivars. *Theor. Appl. Genet.*, 110:1363-1371.
- [2] Albar L., Bangratz-Reyser M., Hebrard E., Ndjiondjop M.N., Jones M., Guesquière A., 2006 – Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to Rice yellow mottle virus. *Plant J.*, 47:417-426.
- [3] *Charron C., Nicolai M., Gallois J.L., Robaglia C., Moury B., Palloix A., Caranta C., 2008. Natural variation and functional analyses provide evidence for coevolution between plant eIF4E and potyviral VPg. *Plant J.*, in press.
- [4] Dodds P.N., Lawrence G.J., Catanzariti A.M., Teh T., Wang C.I., Ayliffe M.A., Kobe B., Ellis J.G., 2006 - Direct protein interactions underlies gene-for-gene

- specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. USA*, 103:8888-8893.
- [5] Goldman N., Yang Z., 1994 – A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.*, 11:725-736.
- [6] Kang B.C., Yeam I., Frantz J.D., Murphy J.F., Jahn M.M., 2005 – The *pvr1* locus in *Capsicum* encodes a translation initiation factor eIF4E that interacts with Tobacco etch virus VPg. *Plant J.*, 42:392-405.
- [7] Leonard S., Chisholm J., Laliberté J.F., Sanfaçon H., 2002 – Interaction *in vitro* between the proteinase of Tomato ringspot virus (genus *Nepovirus*) and the eukaryotic translation initiation factor iso4E from *Arabidopsis thaliana*. *J. Gen. Virol.*, 83:2085-2089.
- [8] Maule A., Caranta C., Boulton M., 2007 – Sources of natural resistance to plant viruses: status and prospects. *Mol. Plant Pathol.*, 5:279-284.
- [9] Mc Dowell J.M., Woffenden B.J., 2003 – Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol.*, 21:178-183.
- [10] McKhann H.I., Camilleri C., Berard A., Bataillon T., David J.L., Reboud X., Le Corre V., Caloustian C., Gut IG., Brunel D., 2004 – Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.*, 38:193-202.
- [11] Robaglia C., Caranta C., 2006 – Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci.*, 11:40-45.
- [12] Ronfort J., Bataillon T., Santoni S., Delalande M., David J.L., Prosperi J.M., 2006 – Microsatellite diversity and broad scale geographic structure in a model legume : building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biol.* 6:28.
- [13] Rose L.E., Michelmore R.W., Langley C.H., 2007 – Natural variation in the Pto disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of Pto. *Genetics*, 175:1307-1319.
- [14] Ruffel S., Caranta C., Palloix A., Lefebvre V., Caboche M., Bendahmane A., 2004 – Structural analysis of the eukaryotic initiation factor 4E gene controlling potyvirus resistance in pepper: exploitation of a BAC library. *Gene*, 38:209-216.
- [15] Ruffel S., Gallois J.L., Moury B., Robaglia C., Palloix A., Caranta C., 2006. Simultaneous mutations in translation initiation factor *eIF4E* and *eIF(iso)4E* are required to prevent Pepper veinal mottle virus infection of pepper. *J. Gen. Virol.*, 87:2089-2098.
- [16] Stracke S., Presterl T., Stein N., Perovic D., Ordon F., Graner A., 2007 – Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding *Bymovirus* resistance in barley. *Genetics*, 175: 805-817.
- [17] Tiffin P., Moeller D.A., 2006 – Molecular evolution of plant immune system genes. *Trends Genet.*, 22:662-670.
- [18] Walsh J., Jenner C., 2002 – Turnip mosaic virus and the quest for durable resistance. *Mol. Plant Pathol.*, 3:289-300.
- [19] Wright S.I., Gaut B.S., 2005 – Molecular population genetics and the search for adaptive evolution on plants. *Mol. Biol. Evol.*, 22:506-519.

Diversité et évolution d'une famille multigénique impliquée dans la biosynthèse de l'amidon chez les graminées, et cas particulier de la domestication du maïs

Jonathan CORBI⁽¹⁾, Marilyn DEBIEU⁽²⁾, Agnès ROUSSELET⁽¹⁾,
Martine LE GUILLOUX⁽¹⁾, Jean-Louis PRIOUL⁽³⁾,
Catherine DAMERVAL⁽¹⁾, Maud TENAILLON⁽¹⁾,
Domenica MANICACCI^{(1)*}

⁽¹⁾INRA, UMR de Génétique Végétale, Ferme du Moulon,
91190 Gif sur Yvette, France

⁽²⁾Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10,
D50829 Koln

⁽³⁾Institut de Biotechnologie des Plantes, Université Paris Sud, F91400 Orsay, France

Abstract: Diversity and evolution of a multigenic family involved in starch biosynthesis in grasses, and the case of maize domestication. ADP glucose pyrophosphorylase (AGPase) is a key enzyme involved in starch biosynthesis, encoded by a multigenic family. In the present study, we identified the large AGPase sub-unit gene expressed in maize leaves, through comparative analysis of BAC and EST sequences. Then, combining both an inter-specific approach in all angiosperms and an intra-specific approach involving wild and cultivated *Zea mays*, we showed that the AGPase multigenic family evolved under various selective pressures. Some duplication events potentially led to phenotypic advances, such as a specific and efficient AGPase in grass kernel. In maize, although all AGPase genes seem to be under larger constraint than average genes in the species, two of them, encoding for the small AGPase subunit in endosperm (*Bt2-a*) and the large sub-unit in leaves (*Sh2-f*), could have been the target of selection during maize domestication.

AGPase/ paralogue gene/ molecular evolution/ expression/ natural selection/ human selection

Résumé : L'AGPase (ADP glucose pyrophosphorylase) est une enzyme clé de la voie de biosynthèse de l'amidon codée par une famille multigénique. Dans un premier temps, nous avons identifié la séquence du gène de grande sous-unité d'AGPase exprimé dans la feuille de maïs grâce à des données récentes de séquençage de BAC et d'EST. Notre étude évolutive combinant une approche inter-spécifique chez l'ensemble des angiospermes et une approche intra-spécifique

* Correspondance et tirés à part : manicacci@moulon.inra.fr

impliquant les compartiments sauvages et cultivés de l'espèce *Zea mays* a permis de montrer que la famille multigénique étudiée a subi, au cours de son évolution, des pressions de sélection variées. Certains événements de duplication ont probablement permis des avancées phénotypiques majeures, comme une AGPase spécifique et plus efficace dans les grains des graminées. Chez le maïs, bien que tous les gènes de cette famille soient relativement contraints, deux d'entre eux semblent avoir donné prise à la sélection au cours de la domestication, *Bt2-a* et *Sh2-f*.

AGPase/ gène paralogue/ évolution moléculaire/ expression/ sélection naturelle/ sélection anthropique

1. INTRODUCTION

La diversité du vivant est la preuve que les espèces ont constamment été capables d'acquérir de nouvelles fonctions. Une hypothèse déjà ancienne [33] et largement soutenue [10], [27], [51], [40] est que la duplication de tout ou partie du génome, en créant des redondances fonctionnelles, a laissé place à la divergence de séquences et de fonctions, permettant une adaptation rapide et efficace à de nouvelles conditions environnementales. Les génomes des végétaux supérieurs ont subi de nombreux événements de duplication partielle et de polyploïdisation [48], dont une conséquence est l'existence de nombreuses familles multigéniques [12]. Bien que le phénomène de duplication semble jouer un rôle majeur dans la spéciation et l'innovation génétique, peu de travaux ont été réalisés sur le devenir à court terme des gènes dupliqués. Nous nous intéressons à l'évolution d'une famille multigénique impliquée dans la voie de biosynthèse de l'amidon (famille de l'AGPase) à deux échelles évolutives : d'une part à l'échelle des angiospermes qui, au cours de leur évolution, ont subi plusieurs événements de duplication conduisant à un nombre de gènes d'AGPase élevé et variable d'une espèce à l'autre, et d'autre part chez le maïs qui, depuis sa domestication, a subi une forte pression de sélection pour l'augmentation des réserves amylacées de ses grains, et dont le génome présente six gènes d'AGPase.

L'ADP Glucose pyrophosphorylase (AGPase) est une enzyme clé de la voie de biosynthèse de l'amidon. Elle est largement répandue et très conservée depuis les bactéries, où elle est composée de quatre sous-unités codées par un même gène (homotétramère) et participe à la synthèse du glycogène [2], jusqu'aux plantes supérieures, où elle se trouve sous forme d'hétérotétramère constitué de deux grandes sous-unités de fonction plutôt régulatrice et de deux petites sous-unités de fonction plutôt catalytique [42]. C'est une duplication très ancienne du gène d'AGPase qui aurait permis la spécialisation fonctionnelle et la complémentarité des deux sous-unités.

Chez le maïs, il a été montré que les AGPases majoritairement exprimées dans l'albumen, l'embryon et la feuille sont codées par trois couples de gènes paralogues différents [16], [34], [15], que nous nommerons ci-après *Sh2-a* et *Bt2-a*, *Sh2-e* et *Bt2-e*, et *Sh2-f* et *Bt2-f*, respectivement. La protéine est plastidiale dans l'embryon et la feuille, alors que la forme majoritaire de l'albumen est cytosolique [5]. Le gène *Sh2-f* n'a pas été identifié à ce jour. A l'origine de notre projet, la phylogénie de l'ensemble de ces gènes était incomplète, mais les divergences entre séquences de chaque sous-unités pour différents organes semblaient plus faibles qu'entre grande et petite sous-unités. Ces données suggèrent que, suite à la duplication initiale ayant donné lieu aux deux sous-unités de type SH2 et BT2, plusieurs duplications indépendantes pour chaque sous-unité ont conduit à des paralogues spécialisés dans différents organes [42], [23].

Chez les céréales, les gènes d'AGPase les plus étudiés sont ceux exprimés dans l'albumen, principal organe de réserves. Chez le maïs, le rôle physiologique crucial de l'AGPase dans l'albumen a été confirmé par l'effet drastique de la mutation des gènes *Sh2-a* (codant la grande sous-unité) ou *Bt2-a* (codant la petite sous-unité) [45], par la corrélation entre la variation d'activité enzymatique de l'AGPase et les variations de teneur en amidon [43], [34], [17], [20], et par la colocalisation de ces gènes avec des QTL de masse du grain ou de teneur en amidon [18], [6], [39]. Depuis la domestication du maïs, il y a environ 10 000 ans, il semble que la teneur en amidon ait été fortement soumise à la sélection humaine. L'amidon, synthétisé dans la feuille et le grain, et spécifiquement accumulé dans l'albumen, permet de répondre à la fois à des exigences en conditions naturelles lors de la germination des grains, mais représente aussi une des principales ressources pour l'Homme lors de la culture du maïs.

Nous avons cherché à identifier les mécanismes évolutifs tels que la sélection, naturelle ou humaine, la dérive génétique, ou les effets démographiques ayant modelé la diversité des gènes d'AGPase. Chez le maïs, espèce dont la domestication a été la plus étudiée du fait de l'importance économique mondiale de cette culture, il est actuellement possible de comparer des formes sauvages actuelles, les téosintes [26], [41], [30], aux formes cultivées ayant subi différentes intensités de sélection : depuis les cultivars traditionnels encore utilisés dans la zone d'origine de l'espèce, l'Amérique centrale, jusqu'au matériel élite aujourd'hui utilisé dans la création variétale. Il est actuellement connu que la domestication du maïs a conduit à une perte globale de diversité résultant à la fois de la sélection sur certains gènes cibles, les gènes du syndrome de domestication [24], [7], [46], [49], mais aussi d'un goulot d'étranglement dû à la réduction d'effectif ayant accompagné les premières étapes de domestication [9], [25], [13].

Au laboratoire, nous avons montré que la variation moléculaire de *Sh2-a* est associée à une variation de la masse du grain et de la teneur en amidon et en protéine de l'albumen ([28], et résultats non publiés). De plus, le patron de diversité du gène *Sh2-a* sur l'ensemble de l'espèce suggère que la grande sous-unité d'AGPase de l'albumen a été soumise à une sélection naturelle ancienne, existant déjà chez les téosintes, conduisant à une forte réduction de la diversité du gène [29]. Par ailleurs, nous avons confirmé les travaux de Whitt *et al.*, [49] montrant que *Bt2-a* a subi une très forte réduction de diversité lors de la domestication, suggérant que ce gène est impliqué dans la différenciation morphologique entre grains de plantes sauvages et cultivées.

Notre projet vise 1) à identifier le gène *Sh2-f* par recherche *in silico* et validation par étude d'expression, 2) à analyser l'évolution de la famille multigénique de l'AGPase chez les angiospermes en recherchant particulièrement les branches de l'arbre phylogénétique au cours desquelles le gène a subi une accélération d'évolution, et enfin 3) à caractériser la variation moléculaire des six gènes d'AGPase de maïs parmi les formes sauvages et cultivées afin de déterminer si certains d'entre eux ont été soumis à des pressions de sélection naturelle ou anthropique.

2. MATÉRIEL ET MÉTHODES

2.1. Identification du gène de la grande sous-unité d'AGPase exprimé dans la feuille chez le maïs, *Sh2-f*, et validation par RT-PCR semi-quantitative.

2.1.1. Identification du gène *Sh2-f*.

La séquence du gène d'orge codant la grande sous-unité d'AGPase exprimée dans la feuille (GenBank U66876; [8]) a été utilisée afin de rechercher *in silico*, par Blastn, une séquence homologue chez le maïs. Nous avons ainsi obtenu deux séquences partielles et chevauchantes, l'une issue du séquençage génomique couvrant 2717 paires de bases de la partie 5' du gène (TIGR data base, AZM4 51892), et l'autre, correspondant à un EST, couvrant 1238 paires de bases de la partie 3' codante du gène (Genbank AY108622). La séquence composite issue de ces deux séquences a été alignée aux gènes d'AGPase de maïs déjà connus.

2.1.2. RT-PCR semi-quantitative

Les ARN totaux de la lignée MBS847 ont été isolés à partir de onze tissus différents: feuilles au stade hétérotrophe (stade 3 feuilles) et autotrophe (4^e feuille), feuille de l'épi et spathes foliacées (stade floraison), ovules non fécondés, grains fécondés au stade 3 JAP (jours après

pollinisation), embryon et albumen de grains en développement au stade 21 JAP, pollen mûr, racines (stade 3 feuilles) et soies non fécondées. Les tissus ont été broyés dans l'azote liquide et les ARN extraits grâce au kit RNeasy mini kit (Qiagen) sans incubation des broyats + tampon RLT à 56°C pour les tissus contenant de l'amidon. Les extraits ont été traités par DNase et rétro-transcrits en cDNA selon la méthode utilisée par Damerval *et al.*, [4]. La concentration des ARN messagers a été ajustée entre organes en utilisant le gène d'ARN ribosomique 18S comme gène constitutif de référence. L'expression des six gènes paralogues dans l'ensemble des tissus a été révélée par PCR à l'aide d'amorces gène-spécifiques (tabl. I), migration sur gel d'agarose et coloration au bromure d'éthidium. Afin d'obtenir des résultats semi-quantitatifs, le nombre de cycles de PCR permettant d'atteindre la phase linéaire d'amplification a été défini, pour chaque gène, en utilisant l'extrait de cDNA de l'organe où l'expression la plus forte était attendue : albumen au stade 21 JAP pour *Sh2-a* et *Bt2-a*, embryon au stade 21 JAP pour *Sh2-e* et *Bt2-e*, et feuille autotrophe pour *Sh2-f* putatif et *Bt2-f* (tabl. I).

Tableau I : Amorces oligonucléotidiques et conditions de RT-PCR pour les gènes d'AGPase.

Gène	Amorce Forward	Amorce Reverse	Ajout mix	TM	NC
<i>Sh2-a</i>	TGCAGCATTCTCAAACACAG	GGATATCAATAAGCCTGTAACAT	1 uL MgCl ₂	55	26
<i>Bt2-a</i>	CTCTGCTTCCTCAACCGTCAC	GTCIGTCACATCAGCATCAAGAACC	2 uL MgCl ₂	56	25
<i>Sh2-e</i>	ATCCCAATGAGCAACTGTIT	GTTCCAATGTCTCCCACT	-	56	22
<i>Bt2-e</i>	TATCCACTGACGAAGAAGAGGGCG	AAGGACATTTTATCATTTCTCA	-	50	30
<i>Sh2-f</i>	GGTGGTGGCGGTATCCTC	CTGCAAGCAGCTCTCTCTT	1x Q, 1 uL MgCl ₂	62	30
<i>Bt2-f</i>	CTCCAGACCTGCCTCGAC	CAGCAACAGTAATATCAGCATCC	1x Q	62	30
<i>18S</i>	CCAATCCCTCCGTAGTTAGCTTCT	CCTGTCCGGCCAAGGCTATATAC	3 uL MgCl ₂	60	10

Ajout mix : complément au tampon de PCR constitué de 1x Tampon de la taq polymérase Qiagen, 1u de taq Qiagen, 0,4 µM de chaque amorce et 0.2 µM de chaque dNTP dans un volume total de 25 µL ; TM : température d'hybridation des amorces ; NC : nombre de cycles de PCR. Les NC cycles de PCR (45" à 95 °C, 30" à la température TM et 2' à 72 °C) sont précédés de 5' à 95 °C et suivis de 5' à 72 °C.

2.2. Évolution de la famille multigénique d'AGPase chez les angiospermes

2.2.1. Recherche *in silico* des séquences de gènes

Nous avons collecté l'ensemble des séquences nucléotidiques codantes correspondant aux gènes de type *Sh2* et *Bt2* par blast (megablast et tblastn) dans les bases de données NCBI (National Center for Biotechnology Information), JGI (Department of Energy Joint Genome Project, pour *Populus trichocarpa*, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, *Ostreococcus tauri* et *Ostreococcus lucimarinus*) et TIGR (pour *Oryza sativa*). Les séquences incomplètes ou trop divergentes n'ont pas été conservées. L'étude concerne ainsi 44 séquences de type *Sh2* (liste complète disponible auprès des

auteurs). Les résultats portant sur les séquences de type *Bt2* ne sont pas reportés ici.

2.2.2. Analyse des séquences

Les séquences ont été alignées à partir de leur traduction protéique en utilisant le logiciel BioEdit [22] avec la matrice BLOSSUM62. L'extrémité N-terminale a été réduite d'environ 82 résidus du fait d'un mauvais alignement des séquences dans cette zone. Une phylogénie du gène *Sh2* a été obtenue par maximum de vraisemblance (500 bootstraps) avec PhyML [21] en utilisant, comme groupe externe, la séquence connue la plus proche de celles des angiospermes : *Sh2* de *Physcomitrella patens* (Fig. 1).

L'alignement et la phylogénie ont ensuite été utilisés pour identifier des sites (codons) présentant une vitesse d'évolution variable au cours de l'histoire de cette sous-unité (modèle dit « branche et site » du programme codeml implémenté dans PAML v4 ; [50]). Le programme utilise le ratio $\omega = dN/dS$ comme estimateur de la vitesse d'évolution, dN étant le taux de substitution non-synonyme et dS le taux de substitution synonyme. Une valeur de 1 indique une évolution neutre, alors qu'une valeur inférieure à 1 indique une contrainte évolutive et une valeur supérieure à 1 une sélection positive. Chaque modèle est évalué par Likelihood Ratio Test (LRT) relativement à un modèle emboîté. Nous avons testé différents scénarios évolutifs permettant d'identifier une variation de la vitesse d'évolution i) suite à la première duplication du gène *Sh2* (fig. 1, branches 1 et 2), ii) suite à la même duplication et jusqu'à la séparation entre monocotylédones et dicotylédones (branches 1 d'une part et 2+3+3' d'autre part), iii) après la séparation entre monocotylédones et dicotylédones pour la sous-unité généralement exprimée dans le grain (branches 4 et 5), et enfin iv) suite à la seconde duplication du gène spécifique des monocotylédones (branches 6 et 7).

Evolution moléculaire de l'AGPase

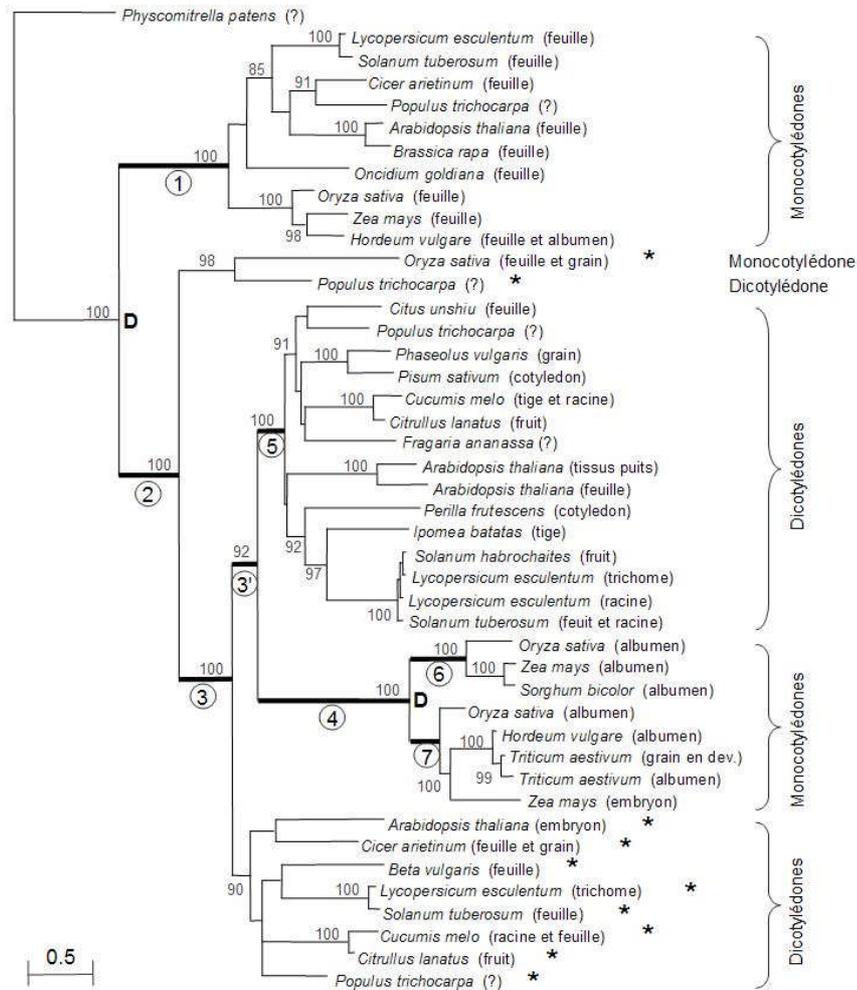


Figure 1: Arbre phylogénique des gènes de la grande sous-unité (SH2) d'AGPase chez les angiospermes. L'arbre a été construit par maximum de vraisemblance (PhyML, 500 bootstraps) avec un modèle G^{TR}+Γ+I dont les paramètres Γ (loi gamma de distribution des taux de substitution) et I (proportion de sites invariants) sont estimés par vraisemblance. La longueur des branches est proportionnelle au taux de substitution synonyme (dS) estimé par PAML. Les organes d'expression sont notés entre parenthèses (? : organe d'expression non décrit). D: événements de duplication. Numéros entourés: branches testées pour une modification du taux d'évolution.*: espèces retirées pour tester le taux d'évolution de la branche 2+3+3'.

2.3. Diversité des gènes d'AGPase dans l'espèce *Zea mays* L.

2.3.1. Matériel végétal

Notre étude a porté sur une collection de maïs cultivés constituée de 13 lignées d'origine américaine et 13 lignées d'origine européenne échantillonnées à partir d'une collection de 372 lignées (décrites par Camus-Kulandaivelu *et al.* [3]) dans le but de maximiser la diversité de marqueurs microsatellites (logiciel Mstrat ; [19]). Par ailleurs, 12 populations traditionnelles (cultivars) originaires d'Amérique centrale et 15 populations de téosintes de la sous-espèce *parviglumis*, forme sauvage actuelle la plus proche des maïs cultivés, ont été choisis pour représenter la diversité répertoriée (accessions CIMMYT). Un individu a été pris au hasard par population de cultivar et de téosinte.

2.3.2. Diversité moléculaire intra-spécifique

Les individus ont été semés en serre et leur ADN extrait à partir de feuilles jeunes (voir Manicacci *et al.* [29]). Une région génomique d'environ 2kb couvrant une partie comparable des six paralogues (allant des exons 4 à 13 pour *Sh2-a* et 3 à 8 pour *Bt2-a*) a été définie à partir de l'alignement des six gènes, et amplifiée chez les lignées, les cultivars et les téosintes (protocoles disponibles auprès des auteurs). Les produits de PCR ont été directement séquencés pour les maïs cultivés (lignées homozygotes) et clonés selon la méthode de TA-cloning (kit Promega ou TopoXL) pour les 27 individus issus des populations de cultivars et de téosintes, potentiellement hétérozygotes. Pour chaque individu, un clone contenant un insert de taille attendue a ensuite été isolé et son insert amplifié par PCR à l'aide d'amorces universelles (PCR à l'aide de taq Qiagen 1 µl pour 20 µl, 30 cycles ; TM ajustée en fonction des amorces), puis séquencé. Pour chaque gène, des amorces internes ont été définies pour séquencer l'intégralité du fragment, puis les séquences ont été alignées et éditées grâce au logiciel Staden, et les séquences correspondant à un même individu ont été contiguës. Les singletons (variants présents une seule fois dans l'alignement de séquences) observés lors du séquençage d'un clone isolé de cultivars ou de téosintes peuvent résulter d'erreurs de la taq-polymérase lors de la PCR. Ils ont été vérifiés soit par séquençage d'un second clone portant le même allèle (gène *Sh2-a*), soit, pour les gènes présentant peu ou pas de polymorphisme d'insertion/délétion, par séquençage direct de produits de PCR sur ADN génomique et identification de polymorphismes nucléotidiques entre allèles par lecture de pics doubles (pour les 5 autres paralogues). Les vérifications de singletons pour le gène *Sh2-f* ont été réalisées pour la majeure partie de la région séquencée. Seule cette partie est analysée, et les dernières vérifications en cours permettront de compléter et de confirmer nos résultats.

Pour chaque gène, l'alignement des séquences a permis d'estimer les taux de diversité π [32] et θ [47] et le nombre d'haplotypes. Le déséquilibre de liaison (LD) a été estimé par r^2 (carré du coefficient de corrélation entre fréquences alléliques) en excluant les sites non informatifs, et nous avons testé sa signification par tests exacts de Fisher et sa corrélation avec la distance physique entre polymorphismes par régression non linéaire [35]. L'hypothèse d'évolution neutre classiquement utilisée en génétique des populations (neutralité, taille de population constante et absence de structure de population) a été testée grâce aux paramètres D [44], D* et F* [11]. Ces analyses ont été appliquées aux lignées, cultivars, maïs (lignées + cultivars) et téosintes, grâce au logiciel DnaSP version 4.00 [38].

3. RÉSULTATS ET DISCUSSION

3.1. Expression des gènes d'AGPase

La séquence composite du gène *Sb2-f* putatif, que nous avons alignée avec les cinq gènes déjà connus chez le maïs, couvre très probablement l'intégralité des exons du gène. Elle est plus proche des gènes de type *Sb2* que des gènes de type *Bt2*, mais suffisamment différente de *Sb2-a* et *Sb2-e* pour pouvoir correspondre au gène *Sb2-f* de maïs.

Le patron d'expression obtenu pour *Sb2-f* putatif (fig. 2) indique 1) qu'il est fonctionnel puisqu'il s'exprime chez le maïs, 2) que son expression est observée dans les organes où elle était attendue puisqu'il s'exprime quasi-exclusivement dans la feuille de maïs, particulièrement au stade jeune, 3) que son patron d'expression est totalement différent de ceux de *Sb2-a* et *Sb2-e*, et enfin 4) qu'il est co-exprimé avec *Bt2-f* puisqu'il présente un patron d'expression similaire. Il semble donc fortement probable que ce gène code bien la grande sous-unité de l'AGPase de feuille de maïs, dont la petite sous-unité est codée par *Bt2-f* (ou *L2* ; [34]).

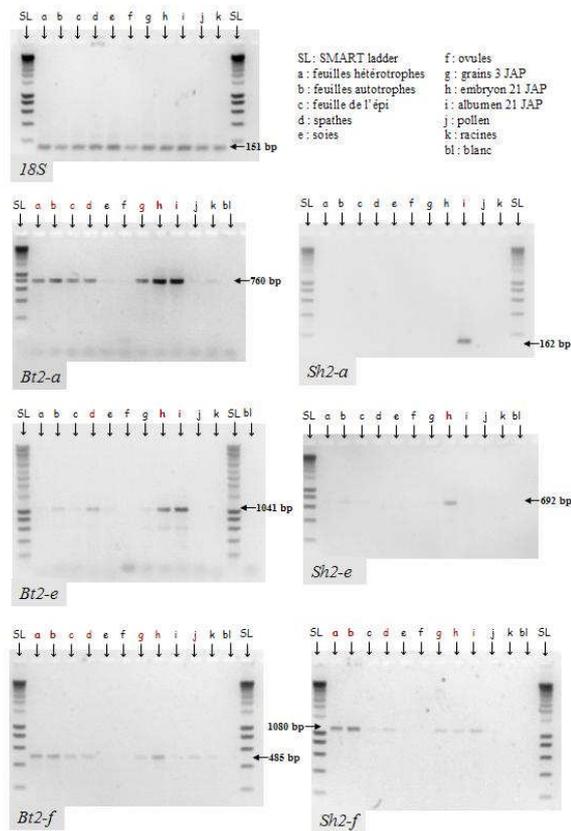


Figure 2 : Expression des gènes d'AGPase. Profils de RT-PCR obtenus pour le gène ribosomique constitutif 18S et les six gènes de la famille multigénique de l'AGPase sur 11 tissus de maïs. SL : marqueur de taille Smart ladder.

Par ailleurs, les patrons d'expression des six gènes d'AGPase sont particulièrement contrastés (fig. 2). Les gènes *Sh2-a* et *Bt2-a* sont décrits pour s'exprimer majoritairement dans l'albumen [16]. L'expression de *Sh2-a* est exclusivement trouvée dans l'albumen à 21 JAP, alors que le gène *Bt2-a* s'exprime aussi dans l'embryon des grains à 21 JAP, et plus faiblement dans le grain jeune (3 JAP) et les feuilles, quel que soit leur stade de développement. Le patron spécifique obtenu pour le gène *Sh2-a* suggère que le patron non spécifique obtenu pour *Bt2-a* est dû à un domaine plus large d'expression plutôt qu'à une contamination des ADNc. Les gènes *Sh2-e* et *Bt2-e* sont décrits pour s'exprimer dans l'embryon [34], [15]. De même que pour les gènes d'albumen, la spécificité de région d'expression est plus grande pour *Sh2-e* que pour *Bt2-e*, qui s'exprime aussi bien dans l'albumen que dans l'embryon des grains en cours de remplissage (21 JAP) ainsi que, plus faiblement, dans les spathes. Ces résultats représentent la première étude d'expression, dans

différents organes de maïs, des gènes de type *Sb2*, et confirment ceux obtenus sur les gènes de type *Bt2* par Rosti et Denyer [37].

3.2. Evolution moléculaire de la famille multigénique d'AGPase chez les angiospermes

Notre approche phylogénétique sur l'ensemble des séquences de type *Sb2* et *Bt2* confirme que le groupe *Sb2* est monophylétique et très éloigné du groupe *Bt2*, et que les séquences de type *Sb2* sont beaucoup plus diverses que celles de type *Bt2* (résultats non présentés ; [14]).

Suite à la plus ancienne duplication du gène *Sb2*, la phylogénie (fig. 1) permet de distinguer deux branches très fortement soutenues, la branche 1 qui conduit à un ensemble de paralogues généralement spécialisés dans la feuille, et la branche 2 (ou 2+3+3' si l'on considère la branche allant de la duplication à la divergence entre monocotylédones et dicotylédones) qui conduit à des paralogues généralement spécialisés dans le grain. Ces deux branches, et plus particulièrement la branche 2 (ou 2+3+3'), montrent une accélération de la vitesse d'évolution de certains codons (tabl. II). Parmi eux, le résidu en position 261 est fortement susceptible d'avoir subi une sélection positive dans les branches 1 ($PP > 80\%$) et 2 ($PP > 95\%$). Ce codon, localisé dans le site actif de la sous-unité SH2 [1], présente un acide aminé S (ou A) de très faible encombrement parmi les gènes exprimés dans la feuille, et un acide aminé Q, H, L ou D, d'encombrement élevé, parmi les gènes exprimés dans le grain. On peut donc penser que, suite à la première duplication du gène *Sb2* chez les angiospermes, une sélection divergente des deux paralogues, portant sur le site actif, a permis leur spécialisation dans la feuille ou dans le grain.

Tableau II : Résultats du modèle « branches et site » (PAML) montrant une variation significative du taux d'évolution dans la phylogénie des gènes de type *Sb2*.

Branche	<i>P</i>	% sites	ω	0,80 < <i>PP</i> < 0,90	0,90 < <i>PP</i> < 0,95	0,95 < <i>PP</i>
1	3,0 10 ⁻²	3,0	43,1	261	225	228 412 423
2	9,2 10 ⁻⁴	2,3	45,6	227		114 261 368
2+3+3'	3,9 10 ⁻⁶	5,7	5,7	261 263 291	64	368
4	2,8 10 ⁻⁸	6,2	6,4	149 159 424	445	
				502		

P: degré de signification du LRT (Likelihood ratio test) dont l'hypothèse nulle est un modèle à deux catégories de sites (contraints et neutres), et le modèle testé possède une catégorie supplémentaire de sites soumis à sélection positive dans la branche considérée, et contraints ou neutres dans le reste de la topologie. La statistique de test (ratio de vraisemblances) suit un Chi-2 à deux degrés de liberté. Le tableau indique la proportion de sites sous sélection positive dans la branche considérée (% sites) et leur valeur de ω , ainsi que la position, relativement à la séquence de *Sb2-a* de *Z. mays* (GenBank AAB52952), des sites ayant une

probabilité postérieure (*PP*) >80% d'être sous sélection positive (algorithme Bayes Empirical Bayes de PAML).

Par ailleurs, le paralogue exprimé dans le grain montre une très forte accélération d'évolution dans la branche conduisant aux monocotylédones (branche 4, fig. 1, tabl. II), impliquant notamment le résidu 502 localisé dans le motif d'activation de la sous-unité. Ce résultat suggère qu'une modification adaptative de la régulation de la sous-unité SH2 ait pu conduire à une protéine plus efficace, permettant la forte accumulation d'amidon caractéristique de l'albumen des graminées.

3.3. Évolution moléculaire des gènes d'AGPase au cours de la domestication du maïs

Aucun des six gènes d'AGPase étudiés ne présente de diminution significative du taux de diversité entre cultivars et lignées. Ces deux origines ont donc été regroupées sous le terme maïs et leur diversité a été comparée à celle des téosintes (fig. 3). Ces six gènes, à l'exception du gène *Sh2-f* pour les téosintes, montrent un faible niveau de diversité nucléotidique en comparaison à la diversité moyenne obtenue pour des gènes neutres chez cette espèce, suggérant de fortes contraintes évolutives sur l'ensemble de la famille multigénique. La structure haplotypique des polymorphismes révèle un déséquilibre de liaison très faible chez les téosintes pour les six gènes (résultats non présentés), alors qu'il est élevé chez les formes cultivées et peu affecté par la distance physique (la figure 4 donne l'exemple du gène *Bt2-d*). Ceci indique qu'un nombre réduit d'haplotypes, issu de la large diversité des téosintes, est présent parmi les formes cultivées.

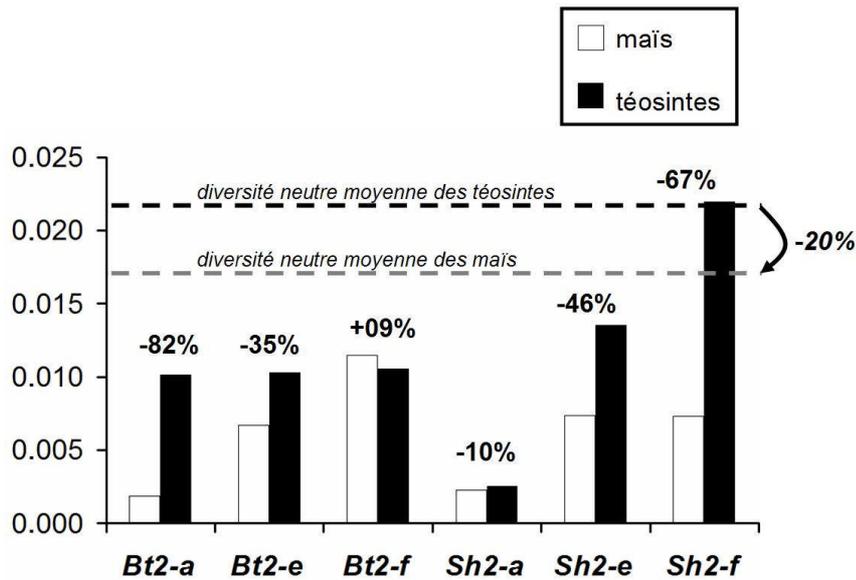


Figure 3 : Diversité nucléotidique des gènes d'AGPase, estimé par π sur sites silencieux (non codants et synonymes) parmi les maïs (lignées élites et cultivars) et les téosintes. Les lignes pointillées indiquent la diversité nucléotidique moyenne des sites silencieux recensée dans des gènes neutres chez les maïs et les téosintes. Les pourcentages en gras indiquent la perte de diversité observée entre téosintes et maïs pour chaque gène, ainsi qu'en moyenne sur les gènes neutres.

La domestication du maïs s'est accompagnée d'un goulot d'étranglement qui a entraîné une forte chute de diversité sur l'ensemble du génome (environ 20 %, [36]). Deux des gènes que nous avons étudiés montrent une perte de diversité nettement supérieure, puisqu'elle est de 67% pour *Sh2f* et de 82% pour *Bt2-a*. Par ailleurs, plusieurs tests de neutralité révèlent une trace de sélection purifiante parmi les maïs et les lignées pour *Sh2f* ($D_{Tajima} = -1,68$, $P < 0,10$ et $D_{Tajima} = -1,86$, $P < 0,05$ pour les maïs et les lignées, respectivement), et parmi les cultivars pour *Bt2-a* ($D_{Fu&Li} = -1,69$, $P < 0,05$). Ces deux gènes ne montrent aucune trace de sélection chez les téosintes. Ces résultats suggèrent que ces deux gènes ont pu jouer un rôle majeur au cours de la domestication et/ou pendant les phases de sélection humaine qui ont suivi. Cette sélection aurait quasiment éliminé la diversité du gène d'albumen *Bt2-a* chez les maïs. Il serait intéressant de mettre en place une approche de génétique d'association afin de déterminer si la diversité du gène *Sh2f*, encore importante parmi les maïs actuels, est liée à une variation phénotypique de caractères tels que la croissance de la plante ou l'accumulation en amidon des grains.

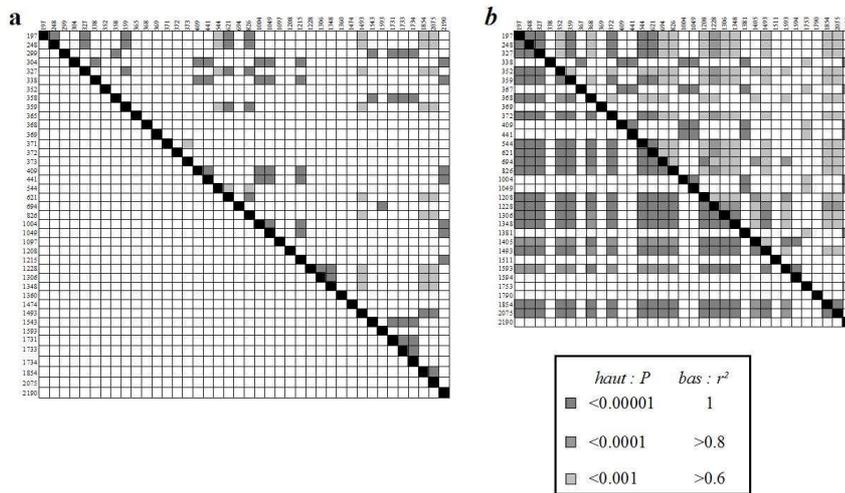


Figure 4 : Déséquilibre de liaison entre sites polymorphes du gène *Bt2-e* parmi a) les téosintes et b) les maïs cultivés (lignées et cultivars). Seuls les polymorphismes informatifs ont été retenus.

L'ensemble de ces résultats est préliminaire et sera complété par des analyses en cours qui devraient permettre de préciser les mécanismes ayant affecté l'évolution de la petite sous-unité d'AGPase, BT2, chez les angiospermes, et de mettre à jour des événements de coévolution entre sous-unités présentes dans un même organe. Enfin la comparaison des études inter et intra-spécifiques nous indiquera si les mêmes résidus ou régions protéiques ont été la cible de la sélection lors de l'histoire des angiospermes et à l'intérieur de l'espèce *Zea mays*, en particulier lors de la domestication du maïs.

REMERCIEMENTS

Nous remercions Milouda Sehabi et Pierre Garrabos pour leur contribution au séquençage des gènes d'AGPase, Guillaume Achaz et Julien Dutheil pour leurs conseils dans l'utilisation de PAML, Gabriella Aquileta pour son aide dans l'interprétation des résultats de l'étude inter-spécifique, Steven Ball pour les discussions enrichissantes à l'occasion du comité de thèse de JC. Cette étude a été financée par le Bureau des Ressources Génétiques (projet n°8 de l'appel à projets 2005-2006), par l'ANR Jeunes Chercheurs (ANR-05-JCJC-0067-01), et par une bourse BDI attribuée à JC.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Akihiro T., Mizuno K., Fujimura T., 2005. Gene expression of ADP-glucose pyrophosphorylase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA. *Plant and Cell Physiology*, 46(6): 937-46.
- [2] Ballicora M.A., Iglesias A.A., Preiss J., 2003. ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiol Mol Biol Rev* 67(2): 213-25, table of contents.
- [3] Camus-Kulandaivelu L., Veyrieras J.B., Madur D., Combes V., Fourmann M., Barraud S., Dubreuil P., Gouesnard B., Manicacci D., Charcosset A., 2006. Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. *Genetics* 172(4): 2449-63.
- [4] Damerval C., Le Guilloux M., Jager M., Charon C., 2007. Diversity and evolution of CYCLOIDEA-like TCP genes in relation to flower development in Papaveraceae. *Plant Physiol* 143(2): 759-72.
- [5] Denyer K., Dunlap F., Thorbjornsen T., Keeling P., Smith A.M., 1996. The major form of ADP-glucose pyrophosphorylase in maize endosperm is extraplasmidial. *Plant Physiol* 112(2): 779-85.
- [6] Doebley J., Bacigalupo A., Stec A., 1994. Inheritance of kernel weight in 2 maize-teosinte hybrid populations - Implications for crop evolution. *Journal of Heredity* 85(3): 191-195.
- [7] Doebley J., Stec A., Hubbard L., 1997. The evolution of apical dominance in maize. *Nature* 386(6624): 485-8.
- [8] Eimert K., Luo C., Dejardin A., Villand P., Thorbjornsen T., Kleczkowski L. A., 1997. Molecular cloning and expression of the large subunit of ADP-glucose pyrophosphorylase from barley (*Hordeum vulgare*) leaves. *Gene* 189(1): 79-82.
- [9] Eyre-Walker A., Gaut R.L., Hilton H., Feldman D.L., Gaut B.S., 1998. Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences of the United States of America* 95(8): 4441-4446.
- [10] Force A., Lynch M., Pickett F.B., Amores A., Yan Y.L., Postlethwait J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151(4): 1531-45.
- [11] Fu Y.X., Li W.H., 1993. Statistical tests of neutrality of mutations. *Genetics* 133(3): 693-709.
- [12] Gaut B.S., Doebley J.F., 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci*, 94(13): 6809-14.
- [13] Gaut B.S., Le Thierry d'Ennequin M., Peek A. S., Sawkins M. C. 2000. Maize as a model for the evolution of plant nuclear genomes. *Proc Natl Acad Sci U S A* 97(13): 7008-15.
- [14] Georgelis N., Braun E.L., Shaw J.R., Hannah L.C., 2007. The two AGPase subunits evolve at different rates in angiosperms, yet they are equally sensitive to activity-altering amino acid changes when expressed in bacteria. *Plant Cell* 19(5): 1458-72.
- [15] Giroux M., Smith-White B., Gilmore V., Hannah L.C., Preiss J., 1995. The large subunit of the embryo isoform of ADP glucose pyrophosphorylase from maize. *Plant Physiol* 108(3): 1333-4.

- [16] Giroux M.J., Hannah L.C., 1994. ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize. *Mol Gen Genet* 243(4): 400-8.
- [17] Giroux M.J., Shaw J., Barry G., Cobb B.G., Greene T., Okita T., Hannah L.C., 1996. A single gene mutation that increases maize seed weight. *Proceedings of the National Academy of Sciences of the United States of America* 93(12): 5824-5829.
- [18] Goldman I.L., Rocheford T.R., Dudley J.W., 1993. Quantitative trait loci influencing protein and starch concentration in the Illinois long-term selection maize strains. *Theoretical and Applied Genetics* 87(1-2): 217-224.
- [19] Gouesnard B., Bataillon T.M., Decoux G., Rozale C., Schoen D.J., David J.L., 2001. MSTRAT: an algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. *Journal of Heredity* 92(1): 93-4.
- [20] Greene T.W., Kavakli I.H., Kahn M.L., Okita T.W., 1998. Generation of up-regulated allosteric variants of potato ADP-glucose pyrophosphorylase by reversion genetics. *Proceedings of the National Academy of Sciences of the United States of America* 95(17): 10322-10327.
- [21] Guindon S., Gascuel O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5): 696-704.
- [22] Hall T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium series* 41: 95-98.
- [23] Hannah L.C., Shaw J.R., Giroux M.J., Reyss A., Prioul J.L., Bae J.M., Lee J.Y., 2001. Maize genes encoding the small subunit of ADP-glucose pyrophosphorylase. *Plant Physiol* 127(1): 173-83.
- [24] Hanson M.A., Gaut B.S., Stec A.O., Fuerstenberg S.I., Goodman M.M., Coe E.H., Doebley J.F., 1996. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* 143(3): 1395-1407.
- [25] Hilton H., Gaut B.S., 1998. Speciation and domestication in maize and its wild relatives: Evidence from the globulin-1 gene. *Genetics* 150(2): 863-872.
- [26] Iltis H. H. 1983. From teosinte to maize: the catastrophic sexual transmutation. *Science* 222: 886-894.
- [27] Lynch M., Force A., 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1): 459-73.
- [28] Manicacci D., De Vienne D., Causse M., Prioul J.L., Touzet P., Falque M., 2001. Utilisation d'associations entre au moins un polymorphisme de séquence nucléique du gène *Sb2* et au moins une caractéristique de qualité de la graine, dans des procédés de sélection des plantes. *patent EP1421219*.
- [29] Manicacci D., Falque M., Le Guillou S., Piegü B., Henry A.M., Le Guilloux M., Damerval C., De Vienne D., 2007. Maize *Sb2* gene is constrained by natural selection but escaped domestication. *J Evol Biol* 20(2): 503-16.
- [30] Matsuoka Y., Vigouroux Y., Goodman M.M., Sanchez G.J., Buckler E.S., Doebley J., 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *PNAS* 99(9): 6080-6084.
- [31] McDonald J.H., Kreitman M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351(6328): 652-4.
- [32] Nei M., 1987. *Molecular evolutionary genetics*, Colombie University Press, New York.

- [33] Ohno S., 1970. Evolution by gene duplication, Springer-Verlag, New York.
- [34] Prioul J. L., Jeannette E., Reyss A., Gregory N., Giroux M., Hannah L. C., Causse M., 1994. Expression of ADP-glucose pyrophosphorylase in maize (*Zea mays* L.) grain and source leaf during grain filling. *Plant Physiol* 104(1): 179-87.
- [35] Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S., 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America* 98(20): 11479-11484.
- [36] Ross-Ibarra J., Morrell P.L., Gaut B.S., 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci U S A* 104 Suppl 1: 8641-8.
- [37] Rosti S., Denyer K., 2007. Two paralogous genes encoding small subunits of ADP-glucose pyrophosphorylase in maize, Bt2 and L2, replace the single alternatively spliced gene found in other cereal species. *J Mol Evol* 65(3): 316-27.
- [38] Rozas J., Sanchez-DelBarrio J.C., Messeguer X., Rozas R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19(18): 2496-7.
- [39] Sene M., Causse M., Damerval C., Thevenot C., Prioul J.L., 2000. Quantitative trait loci affecting amylose, amylopectin and starch content in maize recombinant inbred lines. *Plant Physiology and Biochemistry* 38(6): 459-472.
- [40] Simillion C., Vandepoele K., Van Montagu M.C., Zabeau M., Van de Peer Y., 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 99(21): 13627-32.
- [41] Smith B.D., 1995. The emergence of agriculture. Freeman, New York.
- [42] Smith-White B. J., Preiss J. 1992. Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources. *J Mol Evol* 34(5): 449-64.
- [43] Stark D.M., Timmerman K.P., Barry G.F., Preiss J., Kishore G.M., 1992. Regulation of the amount of starch in plant-tissues by ADP glucose pyrophosphorylase. *Science* 258(5080): 287-292.
- [44] Tajima F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2): 437-60.
- [45] Tsai C. Y., Nelson O. E. 1966. Starch-deficiency maize mutant lacking adenosine diphosphate glucose pyrophosphorylase activity. *Science* 151: 341-3.
- [46] Wang R.L., Stec A., Hey J., Lukens L., Doebley J., 1999. The limits of selection during maize domestication. *Nature* 398(6724): 236-9.
- [47] Watterson G.A., 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2): 256-76.
- [48] Wendel J. F. 2000. Genome evolution in polyploids. *Plant Mol Biol* 42(1): 225-249.
- [49] Whitt S.R., Wilson L.M., Tenaillon M.I., Gaut B.S., Buckler E.S., t. 2002. Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci U S A* 99(20): 12959-62.
- [50] Yang Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 13(5): 555-6.

- [51] Zhang L., Pond S.K., Gaut B.S., 2001. A survey of the molecular evolutionary dynamics of twenty-five multigene families from four grass taxa. *J Mol Evol* 52(2): 144-56.

Diversité et évolution des déshalogénases bactériennes : détection bioinformatique et perspectives de recherche

Maude DAVID⁽¹⁾, Françoise BRINGEL⁽²⁾, Marco PAGNI⁽³⁾,
Niamh GILMARTIN⁽²⁾, Hasna BOUBAKRI⁽¹⁾, Thierry NADALIG⁽²⁾,
Pascal SIMONET⁽¹⁾, Timothy VOGEL⁽¹⁾, Stéphane VUILLEUMIER^{(2)*}

⁽¹⁾ Environmental Microbial Genomics, Laboratoire AMPERE, UMR CNRS 5005,
Ecole Centrale de Lyon, Université de Lyon, 36 avenue Guy de Collongue,
69134 Ecully, France

⁽²⁾ Equipe Adaptations et Interactions Microbiennes dans l'Environnement,
UMR 7156 ULP – CNRS, Génétique Moléculaire, Génomique, Microbiologie,
Université Louis Pasteur, 28 rue Goethe, 67083 Strasbourg cedex, France

⁽³⁾ Swiss Institute of Bioinformatics, Vital-IT group, Quartier Sorge -
Bâtiment Génopode, CH-1015 Lausanne, Suisse

Abstract: Diversity and evolution of bacterial dehalogenases: Detection by bioinformatic tools and research perspectives. In general, our environment has only recently been exposed to significant quantities of the halogenated compounds due to their synthesis for industrial applications. This may have led in the bacterial world to the rapid evolution of genes, such as dehalogenases genes, for the degradation of these often toxic and xenobiotic compounds. Different modes of evolution have been proposed for the emergence of these catabolic genes, but the sequential accumulation of point mutations appears unsuitable for this purpose. Rather, such genes may be more rapidly arrived at by frame-shift mutations in pre-existing genes, a process which in many cases seems to involve repetitive sequence elements. Alternatively, genes coding for functional enzymes may be assembled by reshuffling of DNA fragments of different origins. We are experimentally investigating these different mechanisms in the laboratory, focussing on dehalogenase enzymes which perform the key steps in the degradation of halogenated pollutants. Here, we briefly discuss how the combination of different bioinformatic tools may also help us to detect and analyse the global diversity of dehalogenase genes, and thereby to gain a better understanding of evolutionary processes. In the context of the growing availability of DNA sequences from the environment in sequence databases, the development of streamlined bioinformatic protocols to address questions of gene evolution towards a given function is a challenging perspective for the near future.

**Bioremediation/ halogenated pollutants/ dehalogenation/ bioinformatics/
gene evolution**

* Correspondance et tirés à part : vuilleumier@gem.u-strasbg.fr

Résumé : Les composés halogénés utilisés industriellement, souvent des xénobiotiques toxiques introduits récemment en quantités significatives dans l'environnement, ont vraisemblablement conduit à l'évolution rapide de gènes cataboliques comme ceux des déshalogénases dans le monde bactérien. Différents modes d'évolution ont été proposés pour expliquer l'émergence de ces gènes, dont il semble peu probable qu'ils puissent résulter de l'accumulation successive de mutations ponctuelles. Comme alternative, une évolution de ces gènes par des mutations modifiant le cadre de lecture de gènes préexistants a été proposée, selon des processus qui impliqueraient souvent des éléments de séquence répétés. L'assemblage de gènes codant pour des enzymes fonctionnelles par brassage de fragments de gènes préexistants d'origines diverses représente une autre possibilité. Nous étudions expérimentalement au laboratoire le potentiel de ces différents mécanismes pour l'évolution de gènes, en nous consacrant en particulier aux déshalogénases, enzymes-clé de la dégradation des polluants chlorés. Ici, nous discutons brièvement comment l'utilisation combinée de différents outils bioinformatiques aujourd'hui disponibles peut aussi nous aider à détecter et à analyser la diversité globale des gènes déshalogénases, mais pourrait aussi à terme nous aider à gagner une meilleure connaissance des processus évolutifs à l'origine de l'existence de ces gènes. Dans le contexte actuel de l'augmentation importante de séquences d'ADN environnemental dans les bases de données de séquence, le développement de protocoles bioinformatiques permettant d'aborder les questions d'évolution de gènes pour le codage de fonctions spécifiques émerge comme un défi important pour l'avenir.

Bioremédiation/ polluants halogénés/ déshalogénation/ bioinformatique/ évolution moléculaire

1. INTRODUCTION

Les procaryotes ont été les premiers organismes vivants : ils évoluent sur notre planète depuis 3,8 milliards d'années, durant lesquelles ils ont acquis une diversité physiologique qui leur a permis de coloniser les écosystèmes les plus divers [11]. Les développements récents en génomique ont permis de révéler que cet exceptionnel potentiel adaptatif bactérien est directement lié aux caractéristiques de leurs génomes.

Schématiquement parlant, les génomes ont la capacité d'évoluer par deux types de mécanismes [2]. Les processus verticaux, tout d'abord, entraînent à chaque génération des changements locaux, ponctuels, de l'ADN cellulaire. Ils permettent aussi des réarrangements génomiques par recombinaison du matériel génétique de la cellule, par le biais de séquences d'insertion et de transposons. Par ailleurs, les bactéries disposent de systèmes faisant varier la fréquence des mutations et des réarrangements, et permettant d'accroître leur diversité génétique de façon transitoire (stade hypermutateur) [20].

La transmission horizontale de l'information génétique, en revanche, conduit à des échanges intercellulaires de fragments d'ADN (par transduction, conjugaison ou transformation active ou passive). Ces processus permettent l'intégration par recombinaison d'ADN étranger au matériel génétique d'une cellule [6].

L'ensemble de ces processus, en diversifiant le matériel génétique disponible, représente ainsi un moyen important permettant aux bactéries de s'adapter aux perturbations de l'environnement ou à de nouvelles niches écologiques. Cette évolution des génomes bactériens a fait l'objet de nombreux travaux récents d'analyse bioinformatique (*in silico*). Par exemple, Nakamura et al. [15] ont détecté des événements d'échange de gènes sur de nombreux génomes et, dans certains cas, ont pu identifier les taxons d'origine de ces gènes. Ces études présentent cependant plusieurs limites : elles ne permettent ni de dater précisément ces événements, ni de déterminer les conditions environnementales qui ont conduit à l'acquisition de certains gènes puis à leur fixation dans le matériel génétique des populations étudiées. Ces limitations soulignent la nécessité d'utiliser des modèles d'étude de l'évolution sur une échelle de temps connue. L'utilisation à grande échelle des pesticides produits industriellement par l'homme, parfois qualifiés de xénobiotiques (littéralement "étrangers au vivant"), fournit l'opportunité d'étudier l'adaptation des bactéries à des composés auxquels elles n'ont pas été confrontées au préalable.

Les composés halogénés constituent une classe de xénobiotiques intéressante parce que beaucoup d'entre eux n'ont été introduits que très récemment dans l'environnement, par exemple à partir des années 1920 pour le dichlorométhane et des années 40 pour le lindane. Relevons qu'un nombre significatif de composés toxiques de ce type sont maintenant connus pour être produits naturellement, que ce soit de manière abiotique ou biotique [12], [16]. Il semble toutefois raisonnable d'admettre que la concentration de la plupart de ces composés dans l'environnement, facteur majeur de leur toxicité et des préoccupations sociétales actuelles, a été négligeable jusqu'à l'époque industrielle.

On constate aujourd'hui qu'une part significative des bactéries détectées dans des environnements exposés à des composés chlorés, expriment un métabolisme de dégradation spécifique de ces composés, qui leur permet de les exploiter comme nutriments, parfois même exclusifs, pour leur croissance. Ces voies de dégradation revêtent ainsi un grand intérêt pour l'étude des processus d'évolution microbienne [9]. Les nombreux travaux de recherche dans ce domaine, et ce dès les débuts du génie génétique (voir p. ex. [5]), ont déjà permis d'élucider plusieurs mécanismes responsables de la mobilisation de gènes fonctionnels dans le contexte de l'assemblage de voies de dégradation de polluants. Cependant, si de nombreuses voies

cataboliques sont aujourd'hui élucidées, on ne possède le plus souvent pas d'informations suffisantes sur l'évolution des gènes codant pour les enzymes impliquées, ni sur le degré de divergence de ces séquences par rapport aux séquences d'origine [9].

Dans ce contexte, nos travaux se sont focalisés en particulier sur les déshalogénases, les enzymes qui réalisent les étapes-clé de la dégradation des composés chlorés. Deux hypothèses ont été émises quant à l'origine des gènes correspondants. D'après Janssene et al. [9], ces gènes pourraient provenir d'un gène ancêtre précurseur, fonctionnel ou cryptique, à partir duquel une courte évolution a permis d'obtenir le gène catabolique. On peut envisager par exemple qu'un décalage de la phase de lecture au niveau d'une séquence qui présente des répétitions internes conduise à son évolution en un gène fonctionnel catabolique. Ce gène pourrait alors être dispersé par transfert horizontal au sein de la communauté bactérienne, la sélection exercée par la présence du polluant aidant à sa fixation au sein d'une ou de plusieurs populations. Des indications pour l'existence de ce mécanisme ont été décrites [18]. Une hypothèse alternative pour l'évolution des gènes de déshalogénases, propose que le transfert d'informations entre bactéries s'opère non par la mise à disposition pour la communauté d'un gène fonctionnel, mais par la génération de gènes de novo par brassage moléculaire de plusieurs fragments d'ADN d'origines distinctes [4]. De ce fait même, un nouveau gène fonctionnel construit de la sorte, par transferts horizontaux successifs, garderait la possibilité d'être dispersé au sein de la communauté par recombinaison, puis y être fixé dans les bactéries les mieux adaptées à son action [4].

Nous nous intéressons à dresser un inventaire des déshalogénases microbiennes et de leurs caractéristiques moléculaires, dans le but de renseigner nos hypothèses sur le chemin évolutif de la classe fonctionnelle d'enzymes déshalogénases et d'approfondir ainsi nos connaissances des mécanismes qui contribuent à l'évolution de cette ressource. Notre projet s'articule en deux volets. D'une part, nous travaillons à l'étude bioinformatique exhaustive de l'ensemble des déshalogénases, caractérisées ou putatives, en comparant les séquences nucléiques et protéiques de ces enzymes avec celles disponibles en bases de données, mais aussi en comparant ces séquences entre elles. D'autre part, cette approche bioinformatique nourrit les approches expérimentales que nous poursuivons, et qui visent à l'évolution de déshalogénases au laboratoire, par l'application de différentes techniques de mutagenèse aléatoire, et par brassage et remodelage de fragments d'ADN *in vitro* [4].

2. MATÉRIELS ET MÉTHODES

Les recherches de similarité de séquence sont d'abord réalisées par BLAST [1]. La base de données Uniprot ([22], <http://beta.uniprot.org>) permet de rechercher et d'organiser les familles de séquences en groupes de différents niveaux d'identité ("clusters") ainsi que leurs identifiants bioinformatiques (par exemple ceux des bases de données Hamap, Prosite, Tigrfam, Interpro, Pfam et Blocks utilisées dans le tableau I). Les alignements protéiques et nucléiques sont réalisés avec les différents outils disponibles sur la plateforme d'analyse bioinformatique MyHits ([17], <http://myhits.isb-sib.ch>) et par l'intermédiaire de ClustalX [10]. Le logiciel MEME a aussi été utilisé afin d'identifier des motifs protéiques et nucléiques ([3], <http://bioweb.pasteur.fr/seqanal/motif/meme>). Des outils expérimentaux d'analyse de profils au niveau protéique ([8], cf. <http://www.expasy.org/tools/scanprosite>, et [25] pour un exemple d'application), mais aussi nucléique en cours de développement, ont été utilisés sur la plateforme Vital-IT du Swiss Institute of Bioinformatics (<http://www.vital-it.ch>).

3. RÉSULTATS

Les résultats présentés ici ont à leur origine la recherche de séquences homologues aux déshalogénases identifiées dans la base de données NCBI (National Center for Biotechnology Information) en juin 2006. A cette date, celle-ci contenait 317 génomes bactériens totalement séquencés (en février 2008, ce chiffre était monté à 586). Sur cette base, 80 gènes codant pour des protéines présentant un niveau de similarité significatif avec une déshalogénase connue ont été identifiés. L'élimination des séquences par trop similaires nous a permis de définir une trentaine de gènes différents, représentant l'ensemble des déshalogénases des bases de données de séquence, tous types confondus. A souligner qu'une activité enzymatique correspondante n'avait été démontrée que pour une quinzaine d'entre elles. Le tableau I présente un extrait succinct et actualisé (février 2008) des différents types de déshalogénases bactériennes connus dans la base de données Uniprot, organisés en groupes (clusters) de séquences présentant au moins 50% d'identité entre elles au niveau protéique. Cette préstructuration automatisée des séquences disponibles en base de données est utile car elle permet de définir aisément des représentants-type de chaque sous-famille de séquences dans l'analyse des caractéristiques communes de familles de protéines d'intérêt. Elle permet ainsi de définir 677 clusters "déshalogénase", représentant 2702 séquences. Les 557 clusters d'origine procaryote dans lesquels le mot "dehalogenase" figure dans le titre, comprennent 2459 séquences. A noter que de très nombreux clusters portent le même nom, en

particulier les clusters intitulés sur des variations de la désignation "haloacid dehalogenase" (58 clusters comprenant un total de 287 séquences), sans qu'il soit a priori possible de les interclasser, notamment de par le manque de données expérimentales disponibles. De fait, la grande majorité de ces clusters ne contient pas de protéine pour laquelle une activité a été vérifiée expérimentalement, et la plus grande partie de ces clusters n'ont pas été revus manuellement. Ce constat souligne la nécessité d'une nouvelle analyse experte et sans a priori (c'est-à-dire sans tenir compte de l'annotation existante), par exemple par génération de profils dédiés (cf. p. ex. [25]), de l'ensemble des séquences disponibles en base de données, afin de recouvrer la diversité globale des familles de protéine d'intérêt et en affiner l'analyse.

Pour ce qui concerne les questions d'évolution moléculaire de gènes de déshalogénase qui nous intéressent ici, rappelons que la simple détection de séquences similaires (par Blast, par exemple) dans les séquences bactériennes disponibles en base de données pour chacune des séquences de déshalogénases choisies définit simplement un ensemble de séquences présentant un taux de similarité suffisamment élevé entre elles. Un chemin évolutif qui structure leur distribution ou mette en évidence un gène potentiellement plus "ancestral" n'est pas défini.

Il est toutefois frappant de noter dans ce contexte que pour tous les types de déshalogénases abordés, les analyses ont systématiquement révélé un profil de similarité discontinu, s'étageant entre de forts niveaux de similarité (supérieurs à 50 %), et de faibles niveaux de similarité (25 %), sans valeurs intermédiaires. S'il se confirmait à l'avenir, ce phénomène suggère fortement que l'évolution de gènes codant pour une fonction sous forte pression de sélection dans l'environnement, telle la dégradation de composés toxiques, ne procède que peu souvent par des mécanismes de mutagenèse ponctuelle. Un exemple est présenté dans la figure 1 avec l'analyse de la protéine DcmA (dichlorométhane déshalogénase) de *Methylobacterium dichloromethanicum* DM4, que nous avons étudié en détail [24]. Très peu de séquences présentent une forte similarité avec cette enzyme, jusqu'à une valeur-seuil inférieure à 60 % d'identité environ pour la dichlorométhane déshalogénase codée par le gène *dcmA* de *Methylophilus leisingeri* également caractérisée au laboratoire [21]. Les gènes avec des valeurs de similarité inférieures à 25 % (fig. 1), issus d'organismes phylogénétiquement éloignés et ne partageant pas la même niche écologique, ne peuvent être qualifiés d'homologues au sens strict. Il est donc important de rappeler que la superfamille enzymatique des glutathion S-transférases dont font partie les dichlorométhane déshalogénases comporte d'autres représentants capables de dégrader le dichlorométhane (GST theta 1 des mammifères, 24 % d'identité, [21]) ou d'autres composés chlorés,

comme la tetrachlorohydroquinone déshalogénase LinD [14] intervenant dans la voie de dégradation du lindane (moins de 12 % d'identité).

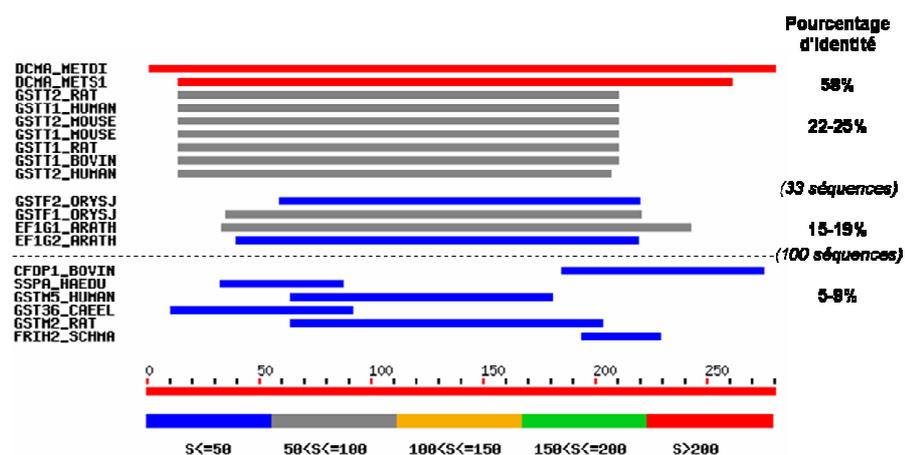


Figure 1: Résultat du Blast de la séquence de la dichlorométhane déshalogénase/glutathion S-transférase (GST) DcmA de *Methylobacterium dichloromethanicum* DM4 avec la base de données Swissprot, montrant les décrochements typiques observés pour les déshalogénases entre les séquences de déshalogénases proprement dites et les séquences de la même superfamille des GSTs. Les premières séquences en grisé sont des GSTs de mammifères dégradant (GSTT1) ou non (GSTT2, 54-58 % d'identité avec GST T1, et comme GST T1 24 % d'identité avec DcmA [7]). Les dernières séquences détectées ("E-value" < 10) sont toujours des GSTs, mais sans activité de déchloration connue. En revanche, d'autres GSTs bactériennes connues pour leur activité de déchloration, comme LinD (cf. tabl. I) ne sont pas détectées dans cette analyse.

Tableau 1 : Les types majeurs de déshalogénases microbiennes et leurs descripteurs bioinformatiques

disername	Gene No.	Acc. No.	divers /total seps	EC	HSSP	Prosite	HAMAP	Tigrfam	Pfam	Interpro	Pins
(S)-haloacid dehalogenase	<i>hddA</i>	Q51645	4 / 51	3.81.2	1QQ5					IPR004493:HAD_S1	
(R)-haloacid dehalogenase	<i>hddB</i>	Q8KLS9	2 / 7	3.81.9			TIG:R01549:HAD_SF1A-V1			IPR004493:HAD_S1	
2,5-dichloroethanol reductase	<i>hddC</i>						TIG:R01493:HAD_SF1A-V2			IPR004493:HAD_S1	
3-chloro-4-hydroxyphenylacetate dehalogenase	<i>hddD</i>	P98806	1 / 6	2.53.38			TIG:R01509:HAD_SF1A-V3			IPR004493:HAD_S1	
(6-chloro)phenol reductive dehalogenase precursor	<i>gpcA</i>	P81594	1 / 44	3.8.1-	PSS0405:GST_CTER		TIG:R01428:HAD_gpcA-II	PF00702:Hydrda.se		IPR004046:GST_C	PR_00413:HADHALOGENASE
4-chlorobenzoyl-CoA dehalogenase	<i>hdbB</i>	O68600	2 / 15		PSS0404:GST_NTER			PF00043:GST_C		IPR010987:GST_C_like	
CmuA chloromethanetera hydrofolate methyltransferase	<i>gpcA</i>	P81594	1 / 44	3.8.1-	P80098			PF00037:Her4		IPR001450:4FeS_FeS_b1	
chloromethane dehalogenase	<i>cmuA</i>	Q9X7F9	1 / 99		4FEAS_HERREDOXIN		TIG:R02486:RDH	PF08425:NiFe_dehyd_N		IPR01534:NiFe_byl_ssa_N	
chloromethane dehalogenase	<i>dmuA</i>	P21161	1 / 16	4.51.3	P80166			PF00378:EC1		IPR012832:RDH	
					1MJ3	ENCOYL_OA_HYDRATASE				IPR001753:Creonase_core	
										IPR015839:Cmbd	
										IPR0015839:Cmbd	
										IPR006759:Met_synth_Cb1-b1	
										IPR006360:Met_sse_McuA_CmuA	
										IPR000257:Uro_diambyls	
										IPR004046:GST_C	
										IPR010987:GST_C_like	
										IPR004045:GST_N	
										IPR012335:Threosokm_fold	

Diversité et évolution des déshalogénases bactériennes

cluster name	Gene Acc.No.	EC	HSSP	Prostc	HAMAP	Tigfam	Pfam	Interpro	Prints
haloacetate dehalogenase	<i>dhl1</i> Q01598	3.8.1.3	IEHY				PF0561_Abhydrolase_1	IPR03089_Ab_hydrolase IPR00073_Ab_hydrolase_1 IPR00639_Ipox_hydrolase	PR0111_AbHYDROLASE PR0412_EPOXYHYDROLASE PR0111_AbHYDROLASE PR0412_EPOXYHYDROLASE PR0412_EPOXYHYDROLASE
haloallene dehalogenase	<i>dhl4</i> P22643	3.8.1.5	1HDE		ME_01230		PF0561_Abhydrolase_1	IPR03089_Ab_hydrolase IPR00073_Ab_hydrolase_1 IPR00639_Ipox_hydrolase	PR0111_AbHYDROLASE PR0412_EPOXYHYDROLASE PR0412_EPOXYHYDROLASE
haloallene dehalogenase	<i>hah</i> P51698	3.8.1.5	1CV2		ME_01231		PF0561_Abhydrolase_1	IPR03089_Ab_hydrolase IPR00073_Ab_hydrolase_1 IPR00639_Ipox_hydrolase IPR02547_AbH_startC2 IPR02198_SDR	PR0111_AbHYDROLASE PR0412_EPOXYHYDROLASE PR0412_EPOXYHYDROLASE
Halohydrin dehalogenase	<i>hka4</i> Q93MS3	3 / 9	1ZMO						
hexachlorocyclohexane dehalogenase	<i>hca4</i> P51697	1 / 50	4.51.-						
trichloroethene reductive dehalogenase	<i>prtA</i> Q81G31	8 / 37	PS0198 4E48_FERREDOXIN			TIGR02486_RDH1	PF0057_Fer4	IPR01450_4E48_Fe_S_bd IPR01283_RDH1 IPR04046_GST_C IPR01097_GST_C_lik IPR04046_GST_N IPR01235_Thioredoxin_fold IPR01450_4E48_Fe_S_bd IPR01283_RDH1	PR0353_4E48FERREDOXIN
trichloro-p-hydroquinone reductive dehalogenase	<i>prgC</i> Q03520	1 / 2	25.1.18 IEGB	PS0465_GST_CTER PS0404_GST_NTER			PF0043_GST_C PF02798_GST_N		
trichloroethene reductive dehalogenase	<i>trcA</i> Q9KIP5	1 / 22	1FCA	PS0198 4E48_FERREDOXIN		TIGR02486_RD11	PF0057_Fer4		
vinyl chloride reductive dehalogenase	<i>trcF1</i> Q6XCM4	1 / 5				TIGR02486_RD11			

4. DISCUSSION

Ce que nous avons évoqué succinctement ici (tabl. I, fig. 1) nous semble caractéristique d'un état des lieux de l'analyse bioinformatique en microbiologie aujourd'hui. Des efforts spécifiques doivent être mis en œuvre non seulement pour une annotation précise et utile et une caractérisation expérimentale des séquences disponibles, mais aussi pour tenter de mieux comprendre les mécanismes évolutifs qui ont conduit à leur émergence. Dans le contexte thématique des déshalogénases qui nous intéresse plus particulièrement, quelques questions concrètes alimentent notre travail aujourd'hui :

- Peut-on estimer aujourd'hui combien de fois l'activité déshalogénase a été acquise au sein des GSTs ou des A, B-hydrolases ?
- Les GSTs impliquées dans des activités moins "exotiques" que la déchloration ont-elles d'autres modes ou tempi d'évolution ?
- Est-il possible ou souhaitable d'établir des relations d'orthologie entre les gènes de déshalogénase ?
- Les profils phylogénétiques (impliquant le voisinage de gènes ([23], <http://string.embl.de>), ou l'interaction avec des composés chimiques spécifiques ([13], <http://stitch.embl.de>) apportent-ils une valeur ajoutée à notre analyse ?
- Y a-t-il systématiquement des éléments génétiques mobiles au voisinage des gènes d'intérêt ?
- Est-il possible d'étendre l'analyse aux protéines associées à l'action des déshalogénases (transporteurs d'acide ou de chlorure, enzymes de réparation de l'ADN ?)

Nous avons pour l'heure constaté que différents outils (comme par exemple ici Blast (et en particulier Psi-Blast [19]), Profilescan [8], [17] et Meme [3]) peuvent apporter des éléments de réponse complémentaires aux questions d'évolution de gènes que nous souhaitons traiter. En tout état de cause, le développement d'un ensemble de protocoles d'analyse de séquences adapté à la fois à l'analyse de séquences sous forte pression de sélection, et à la masse de plus en plus importante et diverse de séquences disponibles émerge comme une perspective de recherche importante, et la classe d'enzymes des déshalogénases s'offre comme un sous-ensemble de séquences d'intérêt fonctionnel bien défini pour la mettre en œuvre.

REMERCIEMENTS

Les auteurs remercient, pour leur soutien aux travaux effectués dans le contexte de ce projet, le Bureau des Ressources Génétiques (projet "Diversité, transmission et évolution des déshalogénases microbiennes dans l'environnement" 2005-2006, le CNRS (ATIP Vuilleumier 2002-2005), le réseau alsacien des laboratoires en ingénierie et sciences de l'environnement Realise (financements CPER 2000-2006, PPF 2005-2008), et le Vital-IT Transnational Access Programme du Swiss Bioinformatics Institute (Vuilleumier 2007) et la Région Rhône-Alpes (M. DAVID, 2006-2008).

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J.H., Zhang Z., Miller W., Lipman D.J., 1997. - Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.*, 25: 3389-3402.
- [2] Arber W., 2000. - Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol. Lett.*, 24: 1-7.
- [3] Bailey T.L., Williams N., Misleh C., Li W.W., 2006. - MEME: discovering and analyzing DNA and protein sequence motifs. *Nucl. Acids Res.*, 34: W369-W373.
- [4] Boubakri H., Beuf M., Simonet P., Vogel T.M., 2006. - Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. *Gene*, 375: 87-94.
- [5] Chakrabarty A.M., Friello D.A., Bopp L.H., 1978. - Transposition of plasmid DNA segments specifying hydrocarbon degradation and their expression in various microorganisms. *Proc. Natl. Acad. Sci. USA*, 75: 3109-3112.
- [6] Frost L.S., Leplae R., Summers A.O., Toussaint A., 2005. - Mobile genetic elements: the agents of open source evolution. *Nature Rev. Microbiol.*, 3: 722-732.
- [7] Gisi D., Maillard J., Flanagan J.U., Rossjohn J., Chelvanayagam G., Board P.G., Parker M.W., Leisinger T., Vuilleumier S., 2001. - Dichloromethane mediated in vivo selection and functional characterization of rat glutathione S-transferase theta 1-1 variants. *Eur. J. Biochem.*, 268: 4001-4010.
- [8] Hulo N., Bairoch A., Bulliard V., Cerutti L., De Castro E., Langendijk-Genevaux P.S., Pagni M., Sigrist C.J.A., 2006. - The PROSITE database. *Nucl. Acids Res.*, 34: D227-D230.
- [9] Janssen D.B., Dinkla I.J.T., Poelarends G.J., Terpstra P., 2005. - Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ. Microbiol.*, 7: 1868-1882.
- [10] Jeanmougin F., Thompson J.D., Gouy M., Higgins D.G., Gibson T.J., 1998. - Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, 23: 403-405.
- [11] Kassen R., Rainey P.B., 2004. - The ecology and genetics of microbial diversity. *Annu. Rev. Microbiol.*, 58: 207-231.
- [12] Keppler F., Eiden R., Niedan V., Pracht J., Scholer H.F., 2000. - Halocarbons produced by natural oxidation processes during degradation of organic matter. *Nature*, 403: 298-301.

- [13] Kuhn M., von Mering C., Campillos M., Jensen L.J., Bork P., 2008. - STITCH: interaction networks of chemicals and proteins. *Nucl. Acids Res.*, 36: D684-D688.
- [14] Miyachi K., Suh S.K., Nagata Y., Takagi M., 1998. - Cloning and sequencing of a 2,5-dichlorohydroquinone reductive dehalogenase gene whose product is involved in degradation of gamma-hexachlorocyclohexane by *Sphingomonas paucimobilis*. *J. Bacteriol.*, 180: 1354-1359.
- [15] Nakamura Y., Itoh T., Matsuda H., Gojobori T., 2004. - Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genet.*, 36: 760-766.
- [16] Öberg G., 2002. - The natural chlorine cycle - fitting the scattered pieces. *Appl. Microbiol. Biotechnol.*, 58: 565-581.
- [17] Pagni M., Ioannidis V., Cerutti L., Zahn-Zabal M., Jongeneel C.V., Falquet L., 2004. - MyHits: a new interactive resource for protein annotation and domain identification. *Nucl. Acids Res.*, 32: W332-W335.
- [18] Poelarends G.J., Kulakov L.A., Larkin M.J., Vlieg J., Janssen D.B., 2000. - Roles of horizontal gene transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways. *J. Bacteriol.*, 182: 2191-2199.
- [19] Przybylski D., Rost B., 2007. - Consensus sequences improve PSI-BLAST through mimicking profile-profile alignments. *Nucl. Acids Res.*, 35: 2238-2246.
- [20] Saint-Ruf C., Matic I., 2006. - Environmental tuning of mutation rates. *Environ. Microbiol.*, 8: 193-199.
- [21] Stourman N.V., Rose J.H., Vuilleumier S., Armstrong R.N., 2003. - Catalytic mechanism of dichloromethane dehalogenase from *Methylophilus* sp. strain DM11. *Biochemistry*, 42: 11048-11056.
- [22] Suzek B.E., Huang H.Z., McGarvey P., Mazumder R., Wu C.H., 2007. - UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23: 1282-1288.
- [23] von Mering C., Jensen L.J., Kuhn M., Chaffron S., Doerks T., Kruger B., Snel B., Bork P., 2007. - STRING 7 - recent developments in the integration and prediction of protein interactions. *Nucl. Acids Res.*, 35: D358-D362.
- [24] Vuilleumier S., Ivos N., Dean M., Leisinger T., 2001. - Sequence variation in dichloromethane dehalogenases/glutathione S-transferases. *Microbiology*, 147: 611-619.
- [25] Vuilleumier S., Pagni M., 2002. - Bacterial glutathione S-transferases: new lessons from bacterial genomes. *Applied Microbiol. Biotechnol.*, 58: 138-146.

Étude de QTL liés à la trypanotolérance bovine dans une population métisse ouest-africaine par analyse d'association

Guiguibaza Kossigan DAYO^(1,5), Sophie THEVENON^{(2)*},
Jean-Paul POIVEY^(3,4), Issa SIDIBE⁽⁵⁾, Zakaria BENGALY⁽⁵⁾,
André EGGEN⁽⁶⁾, Didier BOICHARD⁽⁷⁾, Mathieu GAUTHIER⁽⁶⁾.

⁽¹⁾IRD, UMR Trypanosomes, TA A-17/A Campus international de Baillarguet,
34398 Montpellier cedex 5, France

⁽²⁾CIRAD, UMR Trypanosomes, TA A-17/A Campus international de Baillarguet,
34398 Montpellier cedex 5, France

⁽³⁾CIRAD, UPR Systèmes d'élevage, TA C-18/A,
Campus international de Baillarguet, 34398 Montpellier cedex 5, France

⁽⁴⁾INRA, UR631 Station d'Amélioration Génétique des Animaux,
31326 Castanet-Tolosan, France

⁽⁵⁾CIRDES, BP454 Bobo Dioulasso, Burkina Faso

⁽⁶⁾INRA, UR339 Laboratoire de Génétique Biochimique et de Cytogénétique,
78352 Jouy-En-Josas, France

⁽⁷⁾INRA, UR337 Station de Génétique Quantitative et Appliquée,
78352 Jouy-En-Josas, France

Abstract: QTL linked to bovine trypanotolerance in West African crossbred population. African animal trypanosomiasis is a blood parasitic disease transmitted by tse-tse flies, and present in sub-Saharan Africa. This disease is highly deleterious for indicine cattle population, but is less deleterious for West African taurine breeds, that have the ability to limit parasitemia and anemia and remain productive in enzootic areas. This ability is called trypanotolerance. Several QTLs (Quantitative Traits Loci) have been identified as associated to trypanotolerance traits in an experimental F2 population, crossbred between taurine and zebu. However, these QTLs regions are currently too broad for efficient marker-assisted selection or to look for candidate genes. The next step consists in reducing the confidence interval of some QTLs regions involved in trypanotolerance and trying to confirm these QTL in natural conditions of infection. For that, 363 West African cattles, naturally crossbred between taurine and zebu breeds, have been genotyped and phenotyped for traits related to trypanotolerance. Genotyping concerned four QTLs regions, on BTA02, BTA04, BTA07 and BTA13 respectively, and sixty two microsatellite markers were used. First, linkage disequilibrium between markers was estimated in the studied cattle population in order to estimate the resolution of subsequent association test. Then, the phenotype linked to trypanotolerance was assessed by analyzing data from the previous

* Correspondance et tirés à part : sophie.thevenon@cirad.fr

monitoring of the cattle during two years in an enzootic area. Finally, association studies were performed between the estimated phenotype and the markers genotypes.

A high level of expected heterozygosity (0.70), mean number of alleles per loci (9.8) and a slight shift from Hardy-Weinberg equilibrium were found. Linkage disequilibrium extended over shorter distances than what has been observed in cattle from developed countries. Linkage disequilibrium extended over small distances, which suggests that whole genome scan would require a large number of markers. However, association studies using this population are expected to be effective, and using markers genotyped instead of haplotypes is preferable.

Anemia being one the main symptoms produced by trypanosomiasis, anemia control is one trait associated to trypanotolerance. Anemia level was assessed by measuring Packed Cell Volume (PCV). Trypanosomiasis diagnosis was performed using three methods: direct observation of the parasites in the buffy coat, parasites DNA amplification by PCR (Polymerase Chain Reaction) and antibodies detection by serology. Diagnosis allows determining the period when an animal was infected. Average level of PCV of the monitored animals was modeled using a mixed linear model. This model consists of age, sex, year of survey, season, water points, and the fact to be infected or not by trypanosomes as fixed effects, and animals and groups of breeders as random effects. The animal random effect, added by the mean residuals per individual, was used as the individual phenotype.

Finally, association analyses using ANOVA and linear model were performed between the alleles of each marker and the phenotype associated to mean individual PCV. A marker, MNB-42 on BTA04, was very significantly associated to this phenotype. In perspective, fine mapping will be performed by genotyping new microsatellite markers 500kb around MNB-42. Moreover, phenotype linked to parasitemia control will be estimated.

cattle / trypanosomiasis / linkage disequilibrium / microsatellite / association study

Résumé : La trypanosomose animale africaine est une maladie parasitaire du système sanguin, transmise principalement par les glossines, et présente en Afrique subsaharienne. Cette maladie est très délétère pour les populations bovines de type zébu, mais certaines races taurines d'Afrique de l'Ouest sont tolérantes à cette maladie et restent productives en zones d'enzootie. Plusieurs QTLs, liés à la trypanotolérance, ont été identifiés, dans une population F2 taurin x zébu après infection expérimentale. Cependant, les régions QTLs sont larges et il est nécessaire de réduire leur intervalle de localisation et de les valider en condition naturelle. Pour cela, des bovins d'Afrique de l'Ouest, métissés entre taurins et zébus, ont été phénotypés vis-à-vis de la trypanotolérance et génotypés pour 4 chromosomes. Le déséquilibre de liaison (DL) a été estimé au sein de la population étudiée afin d'estimer la puissance attendue des analyses d'association. Le DL décroît très rapidement avec la distance entre marqueurs. Les principaux facteurs de variation de l'hématocrite moyen des individus ont été analysés, débouchant sur la construction d'un phénotype propre à l'individu et

indépendant des autres facteurs de variation. Ce phénotype a fait l'objet d'analyses d'association avec les marqueurs génotypés, mettant en évidence un marqueur fortement associé sur le chromosome 4.

bovin / trypanosomose / déséquilibre de liaison / microsatellite / test d'association

1. INTRODUCTION

Les trypanosomoses sont des maladies parasitaires du tissu sanguin provoquées par des protozoaires (genre *Trypanosoma*), transmis essentiellement par les glossines ou mouches tsé-tsé (genre *Glossina*). Elles ont des conséquences sanitaires dramatiques sur les populations de ruminants domestiques en Afrique sub-saharienne : les zébus (*Bos indicus*) et les races taurines exotiques souffrent particulièrement de la maladie. Cependant, les taurins ouest africains (*Bos taurus*), dont la race à longue corne (N'Dama) et les races à courtes cornes (Baoulé, Somba, Lagunaire), présentent la particularité de mieux résister naturellement aux trypanosomoses que les zébus. Les races taurines autochtones présentent en effet une meilleure capacité de contrôle de la parasitémie et de l'anémie qui s'accompagne d'une meilleure aptitude à croître et à produire, dans une zone infestée de trypanosomes, que celle présentée par les zébus ou les races taurines exotiques. Ces races sont qualifiées de « trypanotolérantes », et constituent une ressource extrêmement précieuse pour le développement de l'élevage en Afrique sub-saharienne.

Depuis longtemps, des études de génétique quantitative ont montré que le maintien de l'hématocrite et de la productivité sous pression trypanosomienne présente une bonne héritabilité, le contrôle de l'hématocrite étant fortement corrélé à la productivité [28], [29], [30]. Par exemple, chez la N'Dama, l'héritabilité de l'hématocrite moyen est de $0,64 \pm 0,33$ et la corrélation génétique entre l'hématocrite moyen et la croissance est de $0,70 \pm 0,42$ [29]. Le niveau de trypanotolérance est variable selon les animaux au sein même des races trypanotolérantes et certains individus zébus tendent également à montrer une meilleure tolérance que d'autres à la maladie.

Cependant, la sélection massale de ce caractère se heurte aux difficultés classiques de mesure des phénotypes pour les caractères liés à des résistances aux maladies, auquel s'ajoute dans ce cas l'absence de critère spécifique permettant de classer les animaux de façon fiable en tolérants ou sensibles. Des croisements sont d'ailleurs réalisés par les éleveurs dans les zones de contact entre les aires d'élevage traditionnel des races Baoulé et Zébu Peul, mais le caractère empirique de ces croisements fait qu'ils ont peu

d'impact sur le développement de la trypanotolérance des animaux métis ; en outre, l'utilisation massive et systématique des médicaments trypanocides limite la sélection naturelle des métis tolérants.

Aussi apparaît-il essentiel de pouvoir caractériser les gènes impliqués dans la trypanotolérance afin de concevoir et réaliser directement une sélection ou des croisements sur les marqueurs d'intérêt. Des études préliminaires, menées dans des populations de N'Dama de race pure, ont permis de démontrer que la trypanotolérance est un caractère inné sous contrôle multigénique. A partir d'un dispositif expérimental correspondant à un croisement F2 entre taurins N'Dama et zébus Boran, dont les individus ont été soumis à une infection expérimentale contrôlée, les travaux menés par l'ILRI (International Livestock Research Institute) [10], ont révélé quinze régions chromosomiques intervenant significativement dans la variabilité génétique (QTL) de caractères liés à la trypanotolérance (anémie, parasitémie...). Sept des allèles responsables de l'effet QTL observé ont une origine taurine, deux ont une origine zébu et six ont une origine mixte. Ces résultats signifient qu'il serait possible d'obtenir des animaux métis plus tolérants que leurs parents grâce à des croisements appropriés.

L'étape qui suit l'identification de QTLs est la confirmation de ces QTLs en condition naturelle et la réduction de leur intervalle de localisation pour permettre l'identification de « gènes candidats positionnels ». Sachant qu'il est difficile de disposer de pedigrees fiables dans les populations bovines africaines, une première approche de validation des QTLs consiste à réaliser des analyses d'association entre marqueurs et phénotypes liés à la trypanotolérance. Cette approche, bien que limitée par certaines contraintes liées à l'obtention de phénotypes en conditions naturelles, et à des risques de résultats erronés en raison d'une structuration non contrôlée au sein des populations considérées, permet néanmoins d'affiner la localisation des QTLs puisqu'elle bénéficie d'un nombre plus élevé de générations de recombinaison entre le marqueur et la(es) mutation(s) causale(s) recherchée(s) que dans les dispositifs expérimentaux.

Dans le cadre de notre étude, 4 régions QTLs, parmi les plus significatives, identifiées dans le dispositif F2 de l'ILRI, ont été choisies comme candidates pour une étude d'association sur une population de bovins d'Afrique de l'Ouest métis entre taurins à courtes cornes (Baoulé) et zébus (Zébu Peul). Ces mêmes individus avaient fait l'objet d'un suivi sur une période de deux ans en étroite collaboration avec les éleveurs pour recueillir diverses données permettant l'élaboration de phénotypes liés à la trypanotolérance.

L'analyse des données (génotypiques et phénotypiques) s'est déroulée selon 4 étapes principales : 1) étude de la variabilité et de la structure

génétique de la population ; 2) estimation de l'étendue du déséquilibre de liaison entre marqueurs afin d'évaluer le niveau de résolution attendue par les méthodes d'analyse d'association ; 3) estimation des phénotypes en considérant l'hématocrite moyen comme premier critère ; 4) analyses d'association entre le phénotype et les marqueurs.

2. MATÉRIEL ET MÉTHODES

2.1. Site d'étude et matériel animal

Le site d'étude (fig. 1) est situé au sud-ouest du Burkina Faso, dans la province de la Comoé, une zone sub-humide à climat soudano-guinéen. Il est caractérisé par une forte pression trypanosomienne à *Trypanosoma congolense* et *T. vivax* (résultats fournis par le service d'élevage de la région et les sondages parasitologiques et entomologiques réalisés par le Centre International de Recherche-Développement sur l'Élevage en zone Sub-humide). Cinq cent seize (516) bovins, femelles et mâles, âgés de 1 à 2 ans au démarrage du suivi, ont été identifiés par des boucles auriculaires. Ces animaux étaient des métis zébus x taurins appartenant à 76 éleveurs, majoritairement d'ethnie peul, regroupés en 14 campements et vivant sur le site d'étude. Les animaux étaient élevés selon un même système d'élevage extensif, transhumant.

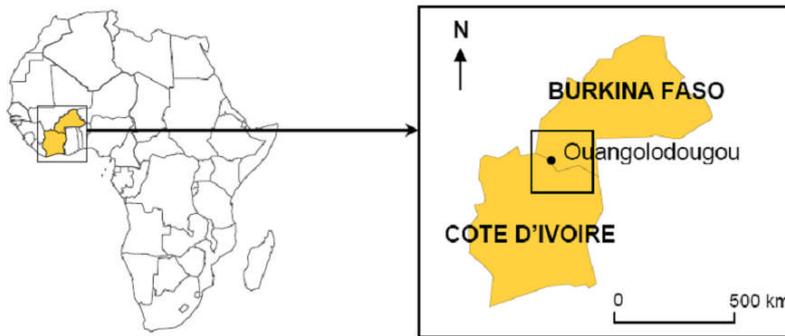


Figure 1 : Zone géographique de l'étude du phénotype lié à la trypanotolérance.

Les animaux ont été suivis tous les mois pendant deux ans, de juin 2003 à juin 2005. Un ensemble de variables a été récolté chaque mois pour chaque animal, à savoir :

- l'état d'embonpoint de l'animal, selon une grille d'observation,
- les informations concernant les événements sanitaires et zootechniques recueillies auprès des éleveurs,
- l'hématocrite,
- la parasitémie, évaluée après centrifugation du sang et observation de l'interface leucocytes/érythrocytes,
- le diagnostic moléculaire des trypanosomoses par PCR (Polymerase Chain Reaction),
- le diagnostic sérologique par test ELISA (Enzyme Linked ImmunoSorbent Assay) indirect détectant les anticorps dirigés contre les antigènes totaux de trypanosomes,
- les traitements administrés par l'éleveur entre chaque visite mensuelle,
- les traitements administrés lorsque l'animal était déclaré malade, selon des critères fonction de la parasitémie, de l'hématocrite, et de l'état général. Les animaux malades ont été traités à l'acéturate de diminazène à la dose de 7 mg/kg.

Les principaux points d'eau, pâturages, et parcs utilisés par les éleveurs, ont été enregistrés par GPS (Global Positionning System).

2.2. Typage des marqueurs moléculaires et analyses des génotypes

L'ADN de chaque animal a été extrait à partir du sang total avec le kit Promega, en plusieurs exemplaires. Quatre zones QTLs, préalablement identifiées par Hanotte *et al.*, [10], sur les chromosomes bovins BTA02, BTA04, BTA07 et BTA13, ont été choisies. Soixante deux (62) marqueurs microsatellites (voir tabl. I) ont été sélectionnés pour leur polymorphisme et leur répartition sur les zones d'intérêt (d'après Ihara *et al.*, [14]) dont :

- 16 marqueurs situés sur le BTA02 dans le QTL associé au contrôle de l'hématocrite;
- 19 marqueurs situés sur le BTA04 dans le QTL associé au contrôle de la parasitémie;
- 17 marqueurs situés sur le BTA07 dans le QTL associé au contrôle de la parasitémie, de l'hématocrite et du poids;
- 10 marqueurs situés sur le BTA13 dans le QTL associé au contrôle de l'hématocrite et de la durée de la parasitémie.

Tableau I : Résumé des informations sur la position des marqueurs (Ihara *et al.*, 2004) et la variabilité dans la population des bovins phénotypés (A : nombre d'allèles ; He : hétérozygotie attendue, $Hobs$: hétérozygotie observée ; P -value du F_{is} : * si P -value < 5 %, ** si P -value < 1 %, *** si P -value < 0,1 %).

Groupe de liaison	Position relative (cM) par multi-pt	Nom des marqueurs	Remarques	A	He	$Hobs$	F_{is}	
2	23,1	TGLA61		15	0,81	0,73	0,10	***
2	27,3	DIK1081		9	0,41	0,39	0,04	
2	34,7	DIK4334		7	0,74	0,67	0,09	**
2	38,9	BM3010		6	0,39	0,38	0,04	
2	42,3	DIK5022		4	0,61	0,55	0,09	**
2	45,3	DIK2853		7	0,77	0,73	0,05	*
2	46,3	DIK1140		11	0,82	0,78	0,05	*
2	49,6	DIK2496		4	0,65	0,60	0,07	*
2	50,9	DIK2729		11	0,73	0,70	0,05	*
2	52,9	DIK2705		15	0,82	0,79	0,04	*
2	56,9	DIK4025		11	0,71	0,66	0,08	**
2	59,3	DIK4673		11	0,86	0,88	-0,02	
2	60,3	BM4440		11	0,79	0,79	0,01	
2	62,0	DIK2719		9	0,79	0,76	0,03	
2	69,8	DIK4972		11	0,84	0,84	0,01	
2	79,6	DIK4208		6	0,40	0,32	0,19	***
4	74,0	DIK4682	Problèmes de lecture	11	0,82	0,80	0,02	
4	77,6	MNB-42		13	0,84	0,82	0,03	
4	79,9	DIK617		10	0,73	0,70	0,04	
4	82,6	DIK123		6	0,68	0,62	0,09	**
4	86,2	DIK026	Pas d'amplification					
4	87,3	LEP-2	Problème de lecture	20	0,91	0,91	0,01	
4	88,7	DIK2740	Pas d'amplification					
4	90,5	DIK4042		6	0,73	0,69	0,05	*
4	93,1	MS2079		9	0,73	0,59	0,20	***
4	94,1	DIK4259		6	0,43	0,38	0,12	***
4	95,2	DIK4236		6	0,59	0,58	0,01	
4	96,4	DIK2646		17	0,81	0,79	0,02	
4	100,9	DIK2579		3	0,15	0,14	0,03	
4	101,3	DIK4290	Allèles nuls	5	0,61	0,20	0,68	***

4	104,5	DIK089		12	0,87	0,83	0,04	*
4	108,9	DIK4404		20	0,90	0,84	0,06	***
4	112,8	MGTG4B		10	0,79	0,82	-0,03	
4	114,6	DIK4521		5	0,43	0,40	0,06	
4	119,9	DIK4542	Allèles nuls	7	0,75	0,35	0,54	***
7	47,9	DIK2819		8	0,77	0,74	0,04	
7	55,3	DIK4606		23	0,90	0,88	0,02	
7	58,6	DIK4739		9	0,81	0,78	0,03	
7	59,3	DIK2689		11	0,82	0,83	-0,01	
7	62,3	DIK4145		9	0,66	0,63	0,04	
7	63,5	DIK4563		16	0,85	0,80	0,06	**
7	66,9	CSSM057		3	0,45	0,44	0,03	
7	67,7	DIK2256		6	0,29	0,28	0,03	
7	70,3	BMS904		3	0,61	0,58	0,05	
7	71,7	DIK4135		23	0,84	0,78	0,07	***
7	72,1	DIK050		13	0,85	0,83	0,02	
7	72,9	MNB-15		15	0,88	0,76	0,13	***
7	73,7	DIK2407		3	0,38	0,39	-0,03	
7	75,5	DIK4386		7	0,64	0,61	0,05	
7	76,2	DIK2915		9	0,81	0,72	0,10	***
7	82,5	INRABERN192	Allèles nuls	9	0,80	0,72	0,10	***
7	90,7	DIK630		9	0,82	0,78	0,05	*
13	74,4	DIK2309		10	0,77	0,76	0,02	
13	75,1	DIK2890		6	0,73	0,72	0,01	
13	76,5	DIK5305		6	0,65	0,65	-0,01	
13	77,1	DIK4350		12	0,68	0,66	0,03	
13	81,0	BL1071		15	0,84	0,82	0,02	
13	82,1	DIK2867	Problèmes de lecture	6	0,69	0,54	0,22	***
13	85,9	DIK537		11	0,76	0,79	-0,04	
13	87,0	DIK4871		10	0,77	0,73	0,05	*
13	91,4	DIK5250		6	0,51	0,52	-0,01	
13	94,5	DIK2117		11	0,82	0,78	0,05	*
Moyenne sur 54 marqueurs				9,8	0,70	0,67	0,045	***

Le typage des marqueurs microsatellites a été réalisé de manière standard par amplification PCR puis migration sur ABI3730x1 Genetic Analyser sequencer d'Applied Biosystems pour les marqueurs des BTA04, BTA07 et

BTA13 et sur MegaBace 1000, GE HealthCare pour les marqueurs de BTA02.

La localisation physique des marqueurs sur les chromosomes bovins a été vérifiée par le criblage d'un panel d'hybrides d'irradiation (RH) et par le criblage de la banque génomique de BAC (Bacterial Artificial Chromosome) INRA. Les distances physiques entre marqueurs, établies par cartographie RH, ont été analysées à l'aide du logiciel CARTHAGENE. D'autre part, des PCR *in-silico* et le blast des séquences attenantes ont été réalisées sur la séquence bovine en cours d'assemblage Btau3.1 (<http://www.hgsc.bcm.tmc.edu/projects/bovine/>) à l'aide du logiciel BLAST [2].

A l'aide des logiciels GENETIX 4.03 (<http://www.genetix.univ-montp2.fr>) et FSTAT [8], nous avons estimé les fréquences alléliques, la diversité génétique aux marqueurs (hétérozygotie attendue, hétérozygotie observée et nombre d'allèles par locus) et l'écart à l'équilibre Hardy-Weinberg (F_i). En outre, nous avons vérifié s'il existait une possible structuration en sous-populations à l'aide du logiciel STRUCTURE 2.2 [20]. Le programme STRUCTURE 2.2 estime la vraisemblance d'une sous-structuration en populations ou la vraisemblance que les individus soient originaires de plusieurs populations ancestrales mélangées, en inférant la proportion du génome de chaque animal issue de ces possibles populations ancestrales. Le nombre (K) de populations est fixé *a priori*. Ici, nous avons testé les valeurs K=1 à 5 et nous avons estimé la vraisemblance de l'observation des génotypes selon K. Garnier *et al.* [7] qui suggèrent de choisir la valeur de K qui maximise le gain d'information, c'est-à-dire de choisir K au-delà duquel l'écart [$\ln P(D)_K - \ln P(D)_{K-1}$] chute, avec $P(D)_K$ la probabilité *a posteriori* des données pour K.

Les phases des génotypes (haplotypes) aux marqueurs microsatellites ont été inférées à l'aide du logiciel PHASE [24]; [25]. La méthode faisant appel à une approche bayésienne, la reconstruction a été effectuée 15 fois de manière indépendante soit 5 répétitions (comme suggéré par Stephens *et al.*, [24]) pour trois valeurs du paramètre de génétique de population ρ ($\rho=4N_e c$, N_e =taille efficace et c = probabilité de recombinaison entre paires de base). Les fréquences haplotypiques ont été utilisées pour vérifier la cohérence de la reconstruction pour les différentes chaînes implémentées. Le déséquilibre de liaison (DL) a ensuite été estimé au niveau gamétique, selon les haplotypes reconstruits, par les estimateurs [12], [15], r^2 [13] χ^2 , [31] (logiciel POWERMARKER). La signification statistique du DL a été testée au niveau génotypique (logiciel GENEPOP) et gamétique (logiciel POWERMARKER).

2.3. Analyses des phénotypes

Le phénotype trypanotolérant dans son ensemble peut être décomposé en phénotype « contrôle de l'hématocrite » et en phénotype « contrôle de la parasitémie ». Dans cette étude, nous nous sommes limités à l'étude de l'hématocrite, car ce dernier est bien corrélé au maintien du poids corporel et à la productivité [29], [30]. Pour l'évaluation du phénotype « hématocrite moyen pendant le suivi longitudinal », des modèles linéaires mixtes, qui intégraient plusieurs facteurs de variation considérés en effets fixes (effet de l'infection différenciant l'espèce de trypanosomes, effet saison, effet de l'année, effet du lieu d'abreuvement, effet de l'âge, effet du sexe...) et les facteurs liés à l'animal et au troupeau modélisés en effets aléatoires, ont été testés. Le meilleur modèle a été choisi en faisant le test du rapport des vraisemblances entre des modèles emboîtés (modèles ajustés par la méthode du maximum de vraisemblance) et en utilisant le critère d'AIC (Akaike's Information Criterion), qui permet d'optimiser la vraisemblance du modèle tout en limitant le nombre de paramètres à estimer. Le bon ajustement des données par les modèles comparés a également été vérifié en examinant la normalité et l'homoscédasticité des résidus.

Le modèle, choisi pour l'estimation de l'hématocrite moyen selon les critères du matériel et méthodes, est le suivant :

$$y_{ijklmnop} = \mu + Se_k + \alpha Ag_{ijklmnop} + \beta Ag_{ijklmnop}^2 + T_{v_l} * T_{c_m} + Sa_n * An_o + Pe_p + An_i + Gr_j + \varepsilon_{ijklmnop}$$

Avec :

- $y_{ijklmnop}$ = hématocrite moyen
- An_i = l'effet aléatoire animal ($i = 1, \dots, 363$)
- Gr_j = l'effet aléatoire du groupe d'éleveur ($j = 1, \dots, 7$)
- Se_k = l'effet sexe ($k = 1, 2$)
- $Ag_{ijklmnop}$ = covariable âge, mesuré en mois et compris entre 5 et 51 mois
- α et β = coefficients de régression linéaire et quadratique sur l'âge de l'animal
- T_{v_l} = l'effet de la positivité à *T. vivax* (toute méthode diagnostique confondue, $l = 0$ ou 1 ; avec $0 =$ « non infecté » ; $1 =$ « infecté »)
- T_{c_m} = l'effet de la positivité à *T. congolense* (toute méthode diagnostique confondue, $m = 0$ ou 1 , avec $0 =$ « non infecté » ; $1 =$ « infecté »)

- Sa_n = l'effet saison ($n = 1, \dots, 4$)
 An_o = l'effet de l'année du suivi ($o = 1, 2$)
 Pe_p = l'effet du point d'eau ($p = 1, \dots, 8$)
 $\varepsilon_{ijklmnop}$ = l'erreur résiduelle, supposée suivre une loi normale $N(0, \sigma^2)$.

L'estimation de l'effet aléatoire animal du modèle mixte, additionné de la moyenne des résidus par animal, ceci contenant encore une part de variabilité due à l'animal, a été considérée comme le phénotype de l'animal. Cette performance est appelée « hémocrite moyen ajusté » ou HMA dans la suite du texte. L'effet du parasite a été ajusté, car l'hémocrite durant la totalité du suivi a été considéré, comprenant les périodes où un animal est vu infecté ou non. D'autre part, il est considéré que *T. congolense* est plus pathogène que *T. vivax*, et que l'interaction des deux espèces de parasites est extrêmement délétère [17].

2.4. Analyses d'association

Pour chacun des allèles présentant une fréquence supérieure à 15 %, une ANOVA à un facteur (avec 3 niveaux : zéro, une ou deux copies de l'allèle) a été réalisée, la variable dépendante étant le phénotype (HMA). Une correction pour les tests multiples a été réalisée en calculant le FDR (False Discovery Rate) sur l'ensemble de l'expérience. Par ailleurs, le modèle a été pondéré par le nombre d'observations par animal. Les analyses statistiques ont été effectuées à l'aide du logiciel R (R Development Core Team, [22]).

3. RÉSULTATS

Sur les 516 bovins initialement sélectionnés pour le suivi longitudinal, 363 ont finalement été génotypés et utilisés pour les analyses des phénotypes. En effet, certains animaux sont rapidement sortis du suivi pour divers motifs : éleveur ayant décidé de ne plus participer à l'étude, animaux vendus, troupeaux à très faible incidence trypanosomienne éliminés du suivi car sans intérêt pour l'étude (non soumis au risque trypanosomien).

3.1. Analyse de la variabilité génétique et la structure de la population étudiée

Deux marqueurs n'ont pu être amplifiés, les profils de six autres marqueurs étaient difficiles à lire et/ou présentaient des allèles nuls. Sur les 54 locus restants, les résultats montrent une très forte variabilité génétique avec une hétérozygotie moyenne attendue de 0,70 (écart-type de 0,17 entre

marqueurs), un nombre moyen d'allèles par locus de 9.8 (écart-type de 4,7). Le F_{is} moyen est de 0,045 et significatif ($P < 0,001$) (tabl. I). Ni l'analyse factorielle des correspondances (AFC) implémentée dans GENETIX ni les analyses réalisées par STRUCTURE n'ont montré de structuration au sein de la population (non montrées).

3.2. Analyse du déséquilibre de liaison

La construction des cartes RH a permis de confirmer l'ordre de la carte génétique [14], excepté pour deux inversions sur BTA07 qui concernent les marqueurs CSSM057 et DIK2256 d'une part, et MNB-15 et DIK2407 d'autre part. Le doute sur la position relative des marqueurs a été levé par le criblage des banques de BAC. D'autre part, l'ancrage des marqueurs sur la séquence bovine en cours d'assemblage a également confirmé une inversion sur la carte génétique. La seconde n'a pas été confirmée car la position du MNB-15 n'a pas pu être retrouvée sur l'assemblage (version Btau3.1) de la séquence du génome bovin. L'ordre des marqueurs selon la carte RH a donc finalement été utilisé pour la reconstruction des haplotypes des individus.

Les résultats de la reconstruction des haplotypes par PHASE sont cohérents, chaque reconstruction pour un groupe de liaison donnant les mêmes haplotypes à fréquence élevée. Pour l'ensemble des individus génotypés, le même haplotype est obtenu dans 54 %, 34 % et 18 % pour les groupes de liaison sur BTA13, BTA07 et BTA04, ce qui signifie inversement que des haplotypes différents sont obtenus dans 46, 66 et 82 % des cas, la différence pouvant concerner un ou plusieurs locus.

La signification statistique du DL, calculé au niveau génotypique, a été estimée par GENEPOP. La figure 2 montre la proportion de paires de locus dont la P -value du DL est inférieure à 1 % en fonction du taux de recombinaison. La proportion chute rapidement au-delà de 5cM. Pour des locus indépendants (50cM), 5 % des marqueurs ont une P -value < 1 %, ce qui dépasse la proportion attendue sous l'effet du hasard.

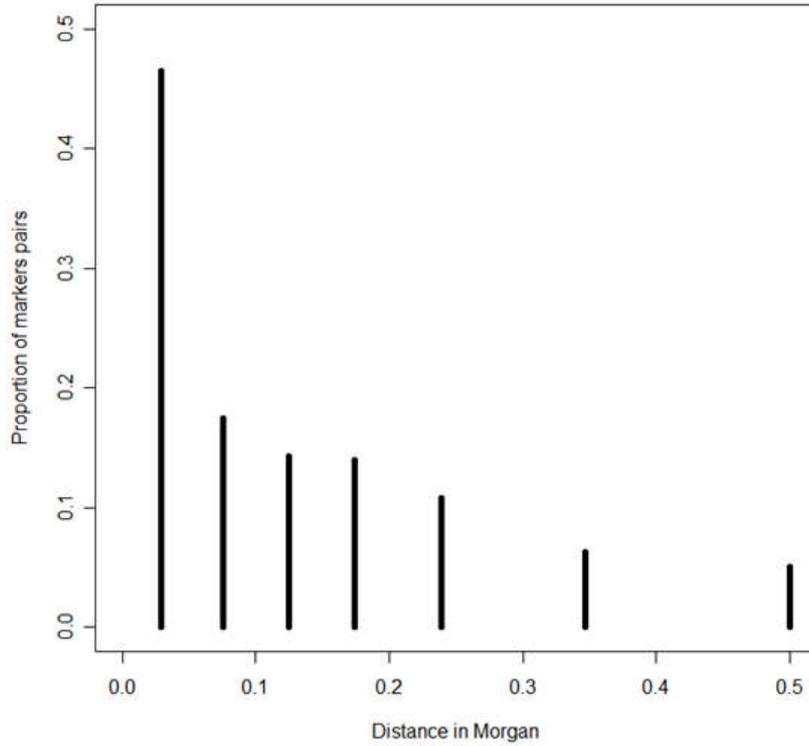


Figure 2: Proportion de paires de marqueurs dont le déséquilibre de liaison (mesuré au niveau des génotypes) a une P -value inférieure à 0,01, en fonction du taux de recombinaison entre les marqueurs (classes de 5 cM, puis de 10 cM pour des distances supérieures à 20 cM).

Les valeurs du DL ont été estimées pour l'ensemble des haplotypes des 15 reconstructions de PHASE par groupe de liaison. Les valeurs moyennes des 3 mesures D' , r^2 , χ^2 , et leurs écarts-types, sont donnés en figure 3. Pour un intervalle entre marqueurs inférieur à 3cM, la valeur moyenne du D' est 0.255 (écart-type = 0.095 entre paires de marqueurs) et chute à 0.158 au dessus de 25cM (écart-type = 0.053).

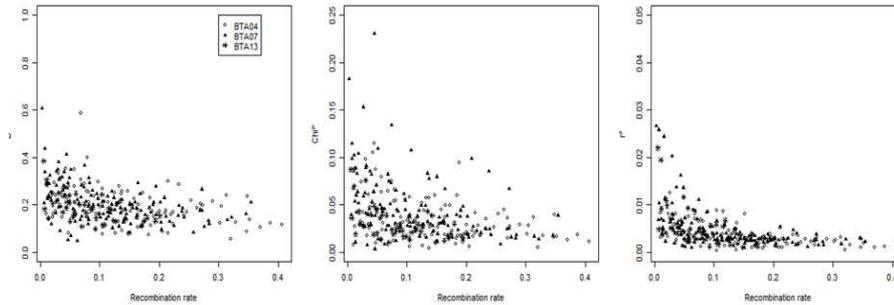


Figure 3 : Trois mesures du déséquilibre de liaison, D' , r^2 , χ^2 (estimées sur les haplotypes reconstruits), en fonction du taux de recombinaison.

3.3. Analyse des phénotypes

Une première étude a été conduite sur l'hématocrite. La saison sèche chaude a un effet très négatif sur l'hématocrite par rapport aux autres saisons ($P < 10^{-4}$), ce qui s'explique par le fait que l'eau et le fourrage sont peu disponibles pour les animaux à cette période. L'année intervient en interaction avec la saison ($P < 10^{-4}$). Le sexe mâle a un effet légèrement négatif sur l'hématocrite ($P = 10^{-4}$), ce qui avait déjà été noté par (Rowlands *et al.*, [23]). L'accroissement de l'âge des individus a un effet positif ($P < 10^{-4}$), ce qui avait été mis en évidence par d'Ieteren [3]. Les points d'eau fréquentés par les animaux ont des effets différents sur l'hématocrite

Comme attendu, le fait d'être infecté par *T. vivax* ou *T. congolense* influe négativement, et de manière très significative ($P < 10^{-4}$) l'hématocrite, et l'infection mixte a un effet encore plus délétère sur la variable d'intérêt ($P < 10^{-3}$).

3.4. Tests d'association

L'analyse de variance effectuée sur HMA a montré que l'allèle 171 du marqueur MNB-42 est associé de manière très significative à HMA ($P < 10^{-5}$ et FDR $< 1\%$). L'allèle 245 du marqueur DIK123, qui se trouve à 5cM de MNB-42 sur BTA04, a également un effet significatif sur HMA (P -value = 0.01 et FDR $< 20\%$). L'effet du premier est détériorateur de l'hématocrite moyen tandis que celui du second est améliorateur.

Les allèles d'autres marqueurs ont une P -value significative mais un FDR supérieur à 20% (voir tabl. II).

Tableau II : Informations sur les marqueurs microsattellites ayant une P -value $< 0,1$ en analyse d'association avec le phénotype HMA : **Intercept**: effet moyen (absence de l'allèle) ; **Effet1** : effet de la présence de l'allèle en une copie ; **Effet2** : effet de la présence de l'allèle en 2 copies.

BTA										
	Marqueurs	Allèles	Intercept	Effet1	Effet2	F-value	P-value	Q-value (FDR)		
4	MNB-42	171	0,25110654	-0,67188512	-1,23783351	9,83573711	6,95E-05	0,00834484		
4	DIK123	245	-0,22678766	0,57841966	0,29592746	5,82763131	0,00323492	0,1940953		
7	DIK2915	196	-0,15643637	0,20756033	0,85657366	5,23195839	0,00576279	0,23051178		
7	DIK2407	199	0,17886323	-0,44621556	-1,08125426	4,84852588	0,00836082	0,25082449		
2	DIK4972	253	0,1924243	-0,34899727	-0,96921748	4,59048843	0,01076297	0,25831118		
7	INRABERN192	188	0,18211655	-0,20248417	-0,92677754	4,33028817	0,01388452	0,2776903		
2	DIK4972	224	-0,1021703	0,45118485	0,5184245	3,24323343	0,04020704	0,61219977		
7	DIK2407	203	-0,41493253	0,24424915	0,59530845	2,85107209	0,05909159	0,61219977		
13	DIK2867	195	-0,16938805	0,23484942	0,50041279	2,73247704	0,06644612	0,61219977		
4	DIK4404	185	-0,07940268	0,21372343	0,72622821	2,56893307	0,07807805	0,61219977		
13	DIK2117	213	-0,10011453	0,24608728	0,61836105	2,38129999	0,09389709	0,61219977		
7	MNB,15	210	-0,08793434	0,41033132	0,40776247	2,54199588	0,08015336	0,61219977		
7	DIK2646	190	-0,1050241	0,14280763	0,61840345	2,36521883	0,0953869	0,61219977		

4. DISCUSSION

4.1. Analyse de la variabilité génétique des bovins dans la population étudiée

Un léger écart à la panmixie (F_{is} moyen significatif) existe dans la population de bovins étudiée. Cependant, les analyses, effectuées à l'aide du programme STRUCTURE et par des méthodes multivariées, ne mettent pas en évidence l'existence de plusieurs sous-populations. On peut donc considérer notre population comme un ensemble homogène, le F_{is} significatif pouvant être dû à la présence d'allèles nuls sur certains marqueurs (voir tabl. I) et/ou à la présence de quelques individus apparentés.

Les analyses des génotypes, décrites dans le rapport intermédiaire du projet, montraient une variabilité génétique très forte dans la population, supérieure à celle observée dans les populations de bovins soumises à la sélection et à l'insémination artificielle en Europe, et également supérieure à celle estimée dans des races pures d'Afrique de l'Ouest, hétérozygotie observée de 0.46 à 0.54 chez le Baoulé et autour de 0.61 chez le zébu Peul Fulani, [5], [6]. Cette forte hétérozygotie était associée à une taille efficace estimée assez large [27]. Cependant, l'estimation de la taille efficace en fonction du DL au cours des générations passées montre une décroissance du N_e [27]. Celle-ci peut être effective, avec une réelle diminution du N_e suite à une sélection par l'homme et/ou naturelle (trypanosomose) ou être due au métissage entre populations zébus et taurines, car le métissage peut biaiser l'estimation du N_e en fonction du DL puisqu'il accroît le DL estimé. La forte hétérozygotie observée conforte l'hypothèse d'un métissage entre races différentes, et celle-ci est confirmée par des génotypages réalisés sur d'autres races avec les mêmes marqueurs (résultats non montrés).

4.2. Analyse du déséquilibre de liaison dans la population des bovins étudiés

Le DL observé dans notre population est plus faible que celui déjà estimé dans des populations bovines sélectionnées [4], [11], [18], [19], [26]. Ces dernières présentent un DL élevé qui persiste sur une longue distance et reflète la diminution rapide de l'effectif efficace sur les dernières générations. La différence du niveau de DL montre de nouveau la forte variabilité génétique de la population étudiée dans notre projet.

D'un point de vue pratique, la caractérisation du DL permet d'estimer le pouvoir statistique des études d'association ayant pour objectif de mettre en évidence des gènes associés à des traits d'intérêt zootechnique. Abecasis *et al.*, [1] et Pritchard, Przeworski, [21] ont proposé un critère pour une

utilisation du DL correspondant à la mesure $r^2 \geq 0.1$: ceci signifie que la taille nécessaire pour détecter une association au marqueur n'excéderait pas 10 fois la taille d'échantillon nécessaire si la mutation causale elle-même était typée. La valeur du DL dans notre étude est relativement faible, et la mesure χ^2 , considérée par Zhao *et al.*, [32] comme étant plus intéressante que le r^2 pour identifier des QTLs avec des marqueurs très polymorphes, atteint rarement ce seuil de 10 %. Ce résultat signifie que notre population est peu intéressante pour une primo-localisation d'un QTL, mais peut en revanche permettre une forte réduction de la zone d'intérêt à condition de densifier suffisamment celle-ci en marqueurs. Dans notre étude, il faudrait donc densifier les zones étudiées en marqueurs. En outre, le faible DL observé explique notre préférence pour une simple analyse d'association entre chaque allèle de chaque marqueur et le phénotype, plutôt que l'utilisation d'haplotypes. De plus certains des QTLs identifiés par infection expérimentale peuvent ne pas avoir été confirmés dans notre étude du fait d'une densification insuffisante en marqueurs.

4.3. Analyse des phénotypes et analyse d'association

Les analyses statistiques d'association révèlent que l'allèle 171 du marqueur MNB-42 est très significativement associé à l'hématocrite moyen des bovins durant le suivi. D'autre part, l'allèle 245 du marqueur DIK123 est également lié à l'hématocrite moyen. Les deux marqueurs sont distants de 5cM sur BTA04, mais le DL entre ces 2 marqueurs est faible ($r^2 = 0.004$). DIK123 n'a pas été localisé dans la séquence du génome bovin et la distance en paires de base est donc inconnue. La forte association de MNB-42 avec l'hématocrite moyen, et sa proximité avec un autre marqueur intéressant, suggèrent que la région contient probablement un gène dont le polymorphisme est lié au phénotype. Notons qu'aucun QTL associé à l'hématocrite n'avait été préalablement détecté sur le chromosome bovin 4.

D'autres marqueurs suggèrent une association, mais le risque, qu'ils s'agissent de faux positifs, est supérieur à celui associé à MNB-42.

4.4. Implications des résultats sur la gestion de la diversité génétique

Notre étude montre une très forte variabilité génétique dans la population étudiée, qui est issue du croisement entre une race zébu et une race taurine d'Afrique de l'Ouest. Cette variabilité génétique est largement supérieure à celle présente dans les races européennes, qui sont fortement sélectionnées pour répondre à des besoins spécifiques de production. L'Afrique de l'Ouest comprend de nombreuses races, de type taurin, zébu et métis. Grâce à la forte diversité génétique des races bovines et à l'adaptation

de ces races à leur milieu, les ressources génétiques ouest-africaines représentent une ressource suffisante pour l'amélioration génétique locale, par de la sélection intra-race ou des croisements raisonnés entre races. L'importation de races exotiques, non adaptées au milieu et dont la diversité génétique est inférieure, devrait rester très limitée.

En outre, les résultats sur l'évolution de la taille efficace estimée par le DL confortent l'hypothèse de métissage entre races de type zébu et taurine. L'introgression progressive des races taurines ouest-africaines, par les races zébus arrivées par l'est [9], représente une menace pour l'intégrité génétique des races taurines [16] et entraîne le risque de disparition de caractères adaptatifs spécifiques, tel celui de la trypanotolérance. La préservation des races pures doit être une priorité des Systèmes Nationaux de Recherche Agronomique et des centres de recherche régionaux d'Afrique de l'Ouest. L'identification de marqueurs génétiques, puis des gènes, associés à des caractères d'adaptation au milieu, comme la trypanotolérance, doit être réalisée afin de préserver ces traits et de les valoriser dans des programmes d'amélioration génétique. La valorisation des ressources génétiques locale doit être un des moyens d'amélioration de la productivité du bétail.

5. CONCLUSIONS

Les premiers tests d'association, entre les allèles et le phénotype correspondant à l'hématocrite moyen des bovins, ont notamment montré un effet significatif de l'allèle 171 du marqueur MNB-42 et de l'allèle 245 du marqueur DIK123 sur l'hématocrite moyen. L'association est très forte pour MNB-42. Des marqueurs microsatellites, distants de 500kb et encadrant MNB-42 sur 1cM ou positionnés entre MNB-42 et DIK123, seront typés sur les animaux phénotypés. De plus, une recherche de gènes candidats dans la zone va être réalisée, grâce au nouvel assemblage de la séquence du génome bovin (Btau4.1) disponible depuis mi-octobre 2007, et à l'utilisation de données issues de travaux d'immunologie et de génomique fonctionnelle.

Pour l'instant, seuls des tests d'association entre les marqueurs et l'hématocrite moyen pendant le suivi ont été entrepris. Or, la trypanotolérance correspond à la capacité de contrôler l'anémie et la parasitémie. L'étude du phénotype liée à la parasitémie est en cours. La variable « parasitémie », en raison de sa distribution extrêmement particulière (variable quantitative discontinue, avec beaucoup de valeurs nulles et quelques valeurs très élevées et dispersées), est plus difficile à modéliser que l'hématocrite (variable quantitative continue, quasi-gaussienne). Des modèles linéaires mixtes généralisés, utilisant une distribution de Poisson avec un excès de valeurs nulles, sont actuellement

testés. Leur validité statistique et biologique doit être vérifiée avant d'entreprendre des tests d'association.

Enfin, nous disposons, sur les 4 groupes de liaison, de génotypes d'autres races bovines, taurines d'Afrique et d'Europe, de zébus métissés d'Afrique, et de zébus purs de Madagascar (non montrés). Une recherche d'éventuelles signatures de sélection (se caractérisant en particulier par une baisse de variabilité génétique chez les populations soumises à une pression trypanosomienne) pourra être entreprise.

En conclusion, une association très significative a été observée entre un marqueur, MNB-42, et le phénotype lié à l'hématocrite moyen des animaux. Les analyses se poursuivent pour affiner la zone et chercher des gènes candidats positionnels. Le travail, sur de nouveaux phénotypes et de nouvelles méthodes pour rechercher des marqueurs associés aux gènes sélectionnés, est en cours mais devra nécessiter des génotypes additionnels. L'approche, qui consiste à valider des QTLs identifiés sur des populations expérimentales dans des populations naturelles, semble donc extrêmement intéressante pour préciser la caractérisation du déterminisme moléculaire de la trypanotolérance.

REMERCIEMENTS

Nous remercions les éleveurs burkinabés, qui ont permis la réalisation du travail, ainsi que les techniciens et chauffeurs du CIRDES qui ont participé au suivi. Les génotypages des marqueurs sur BTA04, 07 et 13 ont été effectués par Labogena. La bourse de thèse de G.K. Dayo est financée par l'IRD. Enfin, nous remercions le BRG et le programme CORUS (Coopération pour la Recherche Universitaire et Scientifique du Ministère des Affaires Étrangères), qui ont financé les travaux présentés dans ce papier.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Abecasis G.R., Noguchi E., Heinzmann A., Traherne J.A., Zhang Y., Lench N.J., Carey A., Cardon L.R., Moffatt M.F., Cookson W.O., 2001. Extent and distribution of linkage disequilibrium in three genomic regions. *American Journal of Human Genetics*, 68: 191-197.
- [2] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25: 3389-3402.
- [3] d'Ieteren G.D., 2001. Trypanotolerance and its exploitation for genetic improvement. Situation report - 30th June 2001. Nairobi, Kenya. ILRI.

- [4] Farnir F., Coppieters W., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mni M., Nezer C., Simon P., Vanmashoven P., Wagenaar D., Georges M., 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Research*, 10: 220-227.
- [5] Freeman A.R., Meghen C.M., Machugh D.E., Loftus R.T., Achukwi M.D., Bado A., Sauveroche B., Bradley D.G., 2004. Admixture and diversity in West African cattle populations. *Molecular Ecology*, 13: 3477-3487.
- [6] Freeman A.R., Bradley D.G., Nagda S., Gibson J.P., Hanotte O., 2006. Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Animal Genetics*, 37: 1-9.
- [7] Garnier S., Alibert P., Audiot P., Prieur B., Rasplus J.Y., 2004. Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, 13: 1883-1897.
- [8] Goudet J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- [9] Hanotte O., Bradley D. G., Ochieng J. W., Verjee Y., Hill E. W., Rege J. O., 2002. African pastoralism: Genetic imprints of origins and migrations. *Science*, 296: 336-339.
- [10] Hanotte O., Ronin Y., Agaba M., Nilsson P., Gelhaus A., Horstmann R., Sugimoto Y., Kemp S., Gibson J., Korol A., Soller M., Teale A., 2003. Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proceedings of National Academy of Sciences of United States of America*, 100: 7443-7448.
- [11] Harmegnies N., Farnir F., Davin F., Buys N., Georges M., Coppieters W., 2006. Measuring the extent of linkage disequilibrium in commercial pig populations. *Animal Genetics*, 37: 225-231.
- [12] Hedrick P.W., 1987. Gametic disequilibrium measures: proceed with caution. *Genetics*, 117: 331-341.
- [13] Hill W.G., Robertson A., 1968. The effects of inbreeding at loci with heterozygote advantage. *Genetics*, 60: 615-628.
- [14] Ihara N., Takasuga A., Mizoshita K., Takeda H., Sugimoto M., Mizoguchi Y., Hirano T., Itoh T., Watanabe T., Reed K.M., Snelling W.M., Kappes S.M., Beattie C.W., Bennett G.L., Sugimoto Y., 2004. A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research*, 14: 1987-1998.
- [15] Lewontin R., 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49: 49-67.
- [16] MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P., Bradley D.G., 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics Society of America*, 146:1071-1086.
- [17] Mattioli R.C., Faye J.A., Büscher P., 1999. Susceptibility of N'Dama cattle to experimental challenge and cross-species superchallenges with bloodstream forms of *Trypanosoma congolense* and *T. Vivax*. *Veterinary Parasitology*, 86: 83-94.
- [18] McRae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M., Slate J., 2002. Linkage disequilibrium in domestic sheep. *Genetics*, 160: 1113-1122.

- [19] Nsengimana J., Baret P.V., 2004. Linkage disequilibrium and the genetic distance in livestock populations: the impact of inbreeding. *Genetic Selection Evolution*, 36: 281-296.
- [20] Pritchard J.K., Stephens M., Donnelly P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.
- [21] Pritchard J.K., Przeworski M., 2001. Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics*, 69: 1-14.
- [22] R Development Core Team, 2007. R: A language and environment for Statistical Computing. Vienne, Austria: R fondation for Statistical Computing.
- [23] Rowlands G.J., Mulatu W., Nagda S.M., d'Ieteren G.D., 1995. Variations in packed red cell volume and trypanosome prevalence and relationships with reproductive traits in east African Zebu cows exposed to drug-resistant trypanosomes. *Acta Tropica*, 59: 105-116.
- [24] Stephens M., Smith N.J., Donnelly P., 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68: 978-989.
- [25] Stephens M., Donnelly P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, 73: 1162-1169.
- [26] Tenesa A., Knott S.A., Ward D., Smith D., Williams J.L., Visscher P.M., 2003. Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science*, 81: 617-623.
- [27] *Thevenon S., Dayo G.K., Sylla S., Sidibe I., Berthier D., Legros H., Boichard D., Eggen A., Gautier M., 2007. The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. *Animal Genetics*, 38: 277-286.
- [28] Trail J.C.M., d'Ieteren G.D.M., Feron A., Kakiese O., Mulungo M., Pelo M., 1991. Effect of trypanosome infection, control of parasitaemia and control of anaemia development on productivity of N'Dama cattle. *Acta Tropica*, 48: 37-45.
- [29] Trail J.C.M., d'Ieteren G.D.M., Maille J.C., Yangari G., 1991b. Genetic aspects of control of anaemia development in trypanotolerant N'Dama cattle. *Acta Tropica*, 48: 285-291.
- [30] Trail J.C.M., d'Ieteren G.D.M., Murray M., Ordner G., Yangari G., Colardelle C., Sauveroché B., Maille J.C., Viviani P., 1993. Measurement of trypanotolerance criteria and their effect on reproductive performance of N'Dama cattle. *Veterinary Parasitology*, 45: 241-255.
- [31] Yamazaki T., 1977. The effects of overdominance of linkage in a multilocus system. *Genetics*, 86: 227-236.
- [32] Zhao H., Nettleton D., Soller M., Dekkers J.C., 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetical Research*, 86: 77-87.

Étude des complémentarités entre gestion dynamique à la ferme et gestion statique en collection : cas de la variété de blé Rouge de Bordeaux

Elise DEMEULENAERE^{(1)*}, Christophe BONNEUIL⁽¹⁾,
François BALFOURIER⁽²⁾, Alain BASSON⁽³⁾,
Jean-François BERTHELLOT⁽³⁾, Vincent CHESNEAU⁽³⁾, Henri FERTÉ⁽³⁾,
Nathalie GALIC⁽⁴⁾, Guy KASTLER⁽³⁾, Jean KOENIG⁽²⁾,
Florent MERCIER⁽³⁾, Joël PAYEMENT⁽³⁾, Alain POMMART⁽³⁾,
Bernard RONOT⁽³⁾, Yves ROUSSELLE⁽⁴⁾, Nicolas SUPIOT⁽³⁾,
Hélène ZAHARIA⁽³⁾, Isabelle GOLDRINGER⁽⁴⁾

¹ INRA-SenS, Institut Francilien Recherche Innovation Société,
77455 Marne-la-Vallée, France

² UMR Génétique, Diversité et Ecophysiologie des Céréales, 234 av. du Brézat,
63100 Clermont-Ferrand, France

³ Réseau Semences Paysannes, Cazalens, 81600 Brens, France

⁴ UMR de Génétique Végétale INRA-Univ. Paris-Sud-CNRS-AgroParisTech,
Ferme du Moulon, 91190 Gif-sur-Yvette, France

Abstract: Complementary aspects of dynamic management on farm and static conservation in genebank based on a case study: the wheat variety 'Rouge de Bordeaux'. While on-farm conservation was considered minor in the development of the National Charter on Genetic Resources in 1998, there is increasing recognition of its important role in the conservation of genetic diversity. In addition to amateur gardening associations that save and exchange seeds, farmers in France have formed networks around systems of shared seed conservation and exchanges, with the goal of protecting the diversity of cultivated species. Parallel to this, the contribution of farmers to the dynamic management of agricultural biodiversity has been recognized by many scientific studies and in international treaties. These developments led us to examine the complementary nature of on-farm and gene-bank conservation efforts in terms of the management of genetic resources. This study, which combines ethnobotanical and genetic approaches, was conducted on bread wheat, for which France has a National Collection of 10000 accessions. There is also an active network of farmers who cultivate historic varieties and landraces, including the Rouge de Bordeaux, which was chosen for a detailed study. By conducting individual interviews with farmers, we were able to understand better their management practices, seed exchanges with other farmers and with the na-

* Correspondance et tiré à part : elise.demeulenaere@ens.fr

tional collection, and strategies for maintaining and selecting this variety on their own farms. We characterized the genetic diversity conserved by several farmers and in the samples preserved in the national collection for Rouge de Bordeaux by using a comparative genetic analysis of samples obtained from farmers and from the collection, in light of the management practices and exchange networks that exist among farmers and between farmers and the national collection. Our results demonstrate that the diversity of these populations is far from being redundant. While some samples are quite homogeneous and similar to samples from the National Collection, others are highly heterogeneous. There is also a high degree of genetic differentiation among populations, with clear groupings of populations identified. The structure of the diversity may be explained by the structure of exchanges and the development of local adaptation within the populations to environmental conditions and management practices, and by differing strategies of selection and conservation. Further analysis will help us understand more precisely what defines a variety and what type of genetic diversity or phenotypic traits are conserved with different management strategies. Our results also support a greater degree of seed circulation between farmers' fields and the National Collection, to conserve the adaptive potential and a broader range of genetic diversity for each variety.

**Dynamic management/ on-farm conservation/ gene-bank conservation/
genetic resources/ seed exchange networks/ adaptation**

Résumé – Considérée comme mineure lors de la rédaction de la Charte nationale des ressources génétiques en 1998, la gestion à la ferme a depuis gagné en importance et en reconnaissance. En effet, après les associations de jardiniers amateurs, des réseaux d'agriculteurs se sont fédérés en France autour de systèmes mutualistes de sélection/conservation de la diversité cultivée. Parallèlement, la contribution des paysans à la gestion dynamique de l'agrobiodiversité a été scientifiquement et institutionnellement reconnue. Ces éléments nous ont conduit à examiner les complémentarités dans la gestion des ressources génétiques à la ferme et en collection. Cette étude, qui croise des approches ethnobotanique et génétique, a été menée sur le blé tendre pour lequel il existe en France une Collection Nationale de 10 000 accessions et un réseau actif d'agriculteurs-collectionneurs de variétés. Pour la variété Rouge de Bordeaux, nous avons tenté de caractériser la diversité dans les champs et en collection, en procédant à une analyse génétique d'échantillons obtenus auprès d'agriculteurs et auprès de la collection. Nos résultats montrent que les ressources génétiques conservées dans les deux compartiments sont loin d'être redondantes. La structuration de la diversité s'explique par les réseaux d'échanges de semences (entre paysans, et entre les paysans et la collection), par l'adaptation locale des populations aux conditions du milieu et aux pratiques de culture, et par les pratiques de sélection/conservation. Ces résultats appellent des analyses complémentaires pour comprendre précisément ce qui est conservé/cultivé sous un nom variétal donné, et plaident en faveur d'une circulation accrue des semences entre champs et collection.

**Gestion dynamique/ gestion à la ferme/ gestion statique/ ressources
génétiques/ réseaux d'échange/ adaptation**

1. INTRODUCTION

La gestion des ressources génétiques se fait selon deux modes, statique ou dynamique, et peut s'appuyer sur des dispositifs *ex situ* ou *in situ*. Selon la Charte Nationale de gestion des ressources génétiques, pour des espèces annuelles à graines orthodoxes comme les céréales et dans le contexte agricole de la France, deux types de gestion *ex situ* sont décrites : (i) la conservation des graines sèches à basse température (dans la Collection Nationale) (ii) la gestion dynamique de populations génétiquement hétérogènes cultivées en stations expérimentales et ressemées d'une année sur l'autre à partir d'échantillons de la récolte précédente†. Pour le blé, il existe effectivement une riche banque de ressources génétiques de céréales à paille gérée par l'INRA à Clermont-Ferrand. Il existe également un programme expérimental de gestion dynamique de la diversité génétique en populations sur lequel on dispose de 20 années de recul et qui constitue une base de réflexion pour la compréhension des mécanismes évolutifs en jeu et pour l'optimisation d'un tel système [17], [20], [12]. Mais il reste un compartiment non évoqué dans la Charte nationale, qui pourrait également jouer un rôle dans la gestion des ressources génétiques : celui de la ferme. Considérée comme mineure lors de la rédaction de la Charte en 1998, la gestion à la ferme a gagné depuis en importance et en reconnaissance et mérite désormais d'être reconsidérée.

En effet, depuis les années 80, des jardiniers amateurs, des agriculteurs, des semenciers artisanaux, des botanistes, produisant ou réhabilitant savoirs et savoir-faire relatifs aux variétés délaissées par le marché, se sont organisés en collectifs et affirmés en tant qu'acteurs à part entière de la gestion du vivant [6], [5]. Des agriculteurs en particulier, dans une démarche d'application des principes de l'agriculture biologique et paysanne (conduites culturales avec faible apport d'intrants, petites exploitations dont la viabilité est maintenue grâce à une pluriactivité et une valorisation des produits en circuits courts), se sont fédérés autour de réseaux mutualistes de sélection/conservation de la diversité cultivée, parfois en collaboration avec des chercheurs institutionnels. Il est à noter que ce n'est pas un hasard si de tels systèmes voient le jour dans le contexte d'une agriculture biologique de proximité. En effet, ce modèle d'agriculture se différencie fortement du système de culture conventionnel par un grand nombre d'items, dont l'hétérogénéité des conditions culturales et des itinéraires techniques, la diversité des besoins des agriculteurs en termes de génotypes/phénotypes végétaux, l'absence actuelle dans le commerce de

† BRG, <http://www.brg.prd.fr/brg/ecrans/charte.htm>. La Charte estime que la gestion dynamique et la conservation à la ferme relèvent de la conservation *ex situ* au sens large, même si l'on peut objecter que le champ, en station expérimentale ou à la ferme, représente le milieu « naturel » des espèces cultivées.

variétés adaptées (du fait de faibles investissements de R&D publique et privée), les demandes spécifiques (organoleptiques, nutritionnelles, sanitaires...) des consommateurs. Nombre de ces caractéristiques sont à mettre en parallèle avec celles des agricultures vivrières et familiales des pays du Sud, pour lesquelles la reconnaissance des limites du modèle de la révolution verte a favorisé les démarches de gestion/sélection participative décentralisée [30].

Par ailleurs, de nombreuses études réalisées sur les systèmes d'agriculture traditionnels des pays du Sud ont montré que les pratiques de gestion des semences par les agriculteurs maintiennent l'ensemble des processus évolutifs (notamment la sélection à travers une diversité de pressions de sélection, et la migration, à travers les réseaux d'échanges de semences), et conduisent de ce fait à une adaptation locale des semences qui maintient dans le même temps la diversité génétique au niveau global [35], [1], [21], [4], [9]. La gestion à la ferme a été rapprochée d'une gestion dynamique en métapopulations, telle que décrite et caractérisée dans le programme expérimental mené sur l'arabette [19]. La contribution des paysans à la gestion dynamique de l'agrobiodiversité a notamment été reconnue dans le Traité International sur les Ressources Génétiques des Plantes pour l'Alimentation et l'Agriculture signé en 2001 entré en vigueur en 2004 sous l'égide de la FAO, qui engage les États à « *encourager (...) les efforts des agriculteurs et des communautés locales pour gérer et conserver à la ferme leurs ressources phytogénétiques* » (art. 5.1.c).

L'intérêt de la gestion dynamique pour la conservation de ressources génétiques adaptées et adaptables a été montré [25], [26], [11], [12], [13], [14], [31], [19], [27], mais il semble qu'un réseau constitué de stations expérimentales, tel celui utilisé dans le programme pilote sur le blé, ne puisse constituer à lui seul un système durable et opérationnel. En effet, (i) le nombre de stations prêtes à s'engager à long terme dans ce dispositif est limité ; (ii) les conditions de culture très standardisées dans les stations couvrent une gamme de variation nettement plus faible que celles rencontrées dans les fermes ; (iii) les pressions de sélection naturelles à l'œuvre dans la gestion dynamique ne contribuent pas efficacement au maintien de la variabilité pour les caractères dits « d'intérêt » au niveau agronomique ou économique (en lien avec la valorisation des produits).

La prise en compte de l'ensemble de ces éléments nous a conduit, dans un projet financé par le Bureau des Ressources Génétiques, à **reconsidérer le rôle de la gestion à la ferme dans le paysage de la conservation des ressources génétiques en France**, et à examiner ses **complémentarités avec la conservation en collection**. Nous avons mené cette étude sur le blé tendre (*Triticum aestivum*), pour lequel il existe à la fois une collection de plus de 10 000 accessions (1/3 d'origine française), décrites pour des

caractères agronomiques, des spectres NIRS et dont 4 000 ont été caractérisées par des marqueurs moléculaires [33], et un réseau actif d'agriculteurs, dont certains s'impliquent dans la recherche, la culture, la conservation ou la sélection à la ferme de plus d'une centaine de variétés anciennes ou locales, depuis plus de 10 ans. Nous avons tenté pour une variété de blé donnée, de caractériser la diversité conservée à la ferme et en collection, en procédant à une analyse génétique comparative d'échantillons obtenus auprès d'agriculteurs et auprès de la collection. Nous avons mis cette diversité conservée à la lumière des pratiques de gestion dans les champs et dans la collection, et à la lumière des échanges de graines existant entre paysans et acteurs institutionnels de la gestion des ressources génétiques.

2. MATÉRIEL ET MÉTHODES

2.1. Enquêtes sur les échanges de semences et sur les pratiques et les usages

Une trentaine d'entretiens individuels avec des producteurs de blé engagés dans la gestion/sélection de semences paysannes ont été réalisés au printemps 2005. Nous nous sommes appuyés pour identifier ces producteurs sur des indications données par le gestionnaire du CRG INRA de Clermont-Ferrand, ainsi que et sur le Réseau Semences Paysannes (RSP), association créée en 2003 pour promouvoir et défendre les semences paysannes, et qui a mené lors de sa création le premier travail d'inventaire des initiatives paysannes autour des variétés locales et anciennes [34], [8]. Ces entretiens ont été complétés par une participation régulière aux réunions du Réseau Semences PaysannesRSP, au cours desquelles les stratégies de conservation et d'échanges étaient discutées entre adhérents.

Les entretiens individuels, conduits autour des motivations des agriculteurs à s'engager dans la culture de semences paysannes, autour des pratiques de conservation à la ferme, autour des origines des variétés et des savoirs et savoir-faire associés et de leurs modalités de transmission, ont permis de réaliser un inventaire des variétés cultivées, et d'établir leur circulation dans un réseau comportant plus de 200 acteurs (comprenant agriculteurs et acteurs institutionnels de la conservation, dont le CRG INRA Clermont). Selon les acteurs, la notion de variété peut prendre plusieurs sens et recouvrir des degrés d'hétérogénéité génétique différents [10]. Dans ces enquêtes, nous avons considéré les catégories variétales telles que définies et nommées par les agriculteurs, sans préjuger de leur recoupement avec les variétés des sélectionneurs. Certaines d'entre elles sont très hétérogènes et sont plus volontiers qualifiées de « variétés-populations » par les

agronomes‡. Notre inventaire des variétés et des échanges s'est accompagné d'un recueil d'informations sur les discours et les pratiques de gestion de la diversité cultivée.

Nous disposons par ailleurs de la liste des lots de semences diffusés à l'extérieur par le CRG INRA Clermont depuis 1994 (depuis l'informatisation de la base de données).

2.2. Analyses phénotypique et génétique des populations de blé

2.2.1. Variété-population étudiée : la variété Rouge de Bordeaux

À partir du réseau de circulation des variétés, nous avons identifié la variété « Rouge de Bordeaux » comme un bon modèle pour notre étude. Comme illustré sur la figure 1, cette variété ancienne (1884), appréciée pour ses qualités en panification artisanale, est en culture chez de nombreux « paysans-boulangers ». Cette variété possédait à l'origine une certaine hétérogénéité génétique, puisqu'elle est issue de sélection massale (à partir de Noé, elle-même sélection massale dans des variétés de pays d'origine russe), et a probablement acquis à travers les cycles de culture dans les champs encore plus de variabilité. C'est pourquoi nous la qualifions de variété-population. La figure 1 montre en outre que des lots de semences ont circulé (dans un sens ou dans l'autre) entre le CRG INRA Clermont et des agriculteurs. Nous avons choisi d'étudier les populations en fonction de leur position dans le schéma d'échanges, mais également de la facilité à obtenir des échantillons. Il s'agit de 19 populations dites de Rouge de Bordeaux (RB) cultivées par 10 agriculteurs différents, certains cultivant plusieurs origines de RB ou bien une même origine selon plusieurs modalités. Quatre échantillons, correspondant chacun aux 4 accessions répertoriées dans la collection de Clermont-Ferrand sous l'étiquette RB ou « type » RB, ont été intégrés dans l'étude, dont l'un (code erge 6310) sert de référence pour la variété Rouge de Bordeaux.

Quelques variétés-populations supplémentaires également très cultivées par les paysans ont été ajoutées à cette étude : deux échantillons de Concorde, le « mélange de James » et la « Touselle de Mayan », ainsi que les échantillons de deux Touselles de la collection faisant partie du mélange (Touselle blanche barbue, blé tétraploïde, et Touselle annone). Elles ont été utilisées dans certains cas comme références extérieures à l'ensemble des RB.

‡ L'agronome Jean Bustarret propose le terme en 1944, pour l'opposer aux variétés lignées pures et aux clones [7].

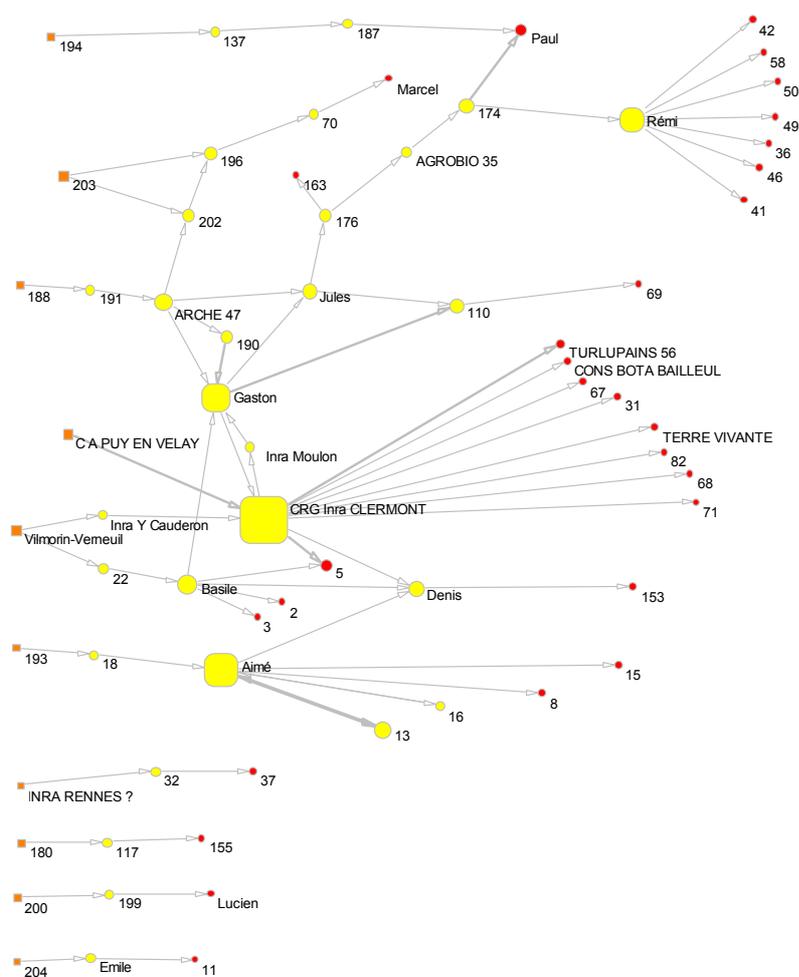


Figure 1 : Circulation de la variété-population Rouge de Bordeaux dans un réseau hybride comprenant agriculteurs (rendus anonymes par des numéros ou des prénoms fictifs) et acteurs institutionnels de la sélection-conservation.

2.2.2. Caractérisation phénotypique des populations

Les 23 échantillons ont été évalués en station au Moulon (région parisienne) dans un dispositif à deux répétitions en blocs complets dans des micro-parcelles de deux rangs, comportant chacune 35 à 40 plantes. Le semis a été réalisé le 8 novembre 2006. Les mesures de précocité d'épiaison ont été réalisées en pourcentage de plantes épiées au cours du temps afin d'estimer la date d'épiaison moyenne de la parcelle et son étalement. Des mesures sur les épis récoltés plante à plante sont en cours. Dans un autre

dispositif, ces mêmes échantillons ont été semés au printemps (le 26 avril 2007) sans vernalisation préalable afin d'évaluer la proportion de plantes de type hiver (nécessitant une période de froid pour épier) et celle de type printemps (qui épient indépendamment des températures froides). Le nombre de plantes épiées sur un total de 26 à 34 plantes dans chaque échantillon a été compté chaque semaine du 22 juin au 13 août.

Les échantillons ont également été mis en culture dans six fermes parmi leurs sites d'origine, suivant un protocole établi en étroite collaboration avec les agriculteurs concernés : 10 échantillons/site avec un jeu commun, une évaluation prenant en compte les critères d'observation des paysans. Ces données ne sont pas encore disponibles.

2.2.3. Marquage microsatellite

Un fragment de feuille a été prélevé sur 20 (cas d'un échantillon) à 42 individus par échantillon le 19 mars dans le dispositif au champ. Les ADN ont été extraits avec purification sur membranes de silice selon un protocole adapté d'après la procédure Qiagen au laboratoire de l'UMR de Génétique Végétale du Moulon. Le génotypage microsatellite de ces individus (938) pour 21 locus microsatellites a été réalisé à la plateforme de génotypage INRA de Clermont-Ferrand. Dix-neuf marqueurs (*Xgwm*) développés par Röder *et al.* [32] et le marqueur Cfd71 développé par Guyomarc'h *et al.* [16] qui amplifie deux locus sur deux chromosomes ont été choisis pour couvrir les 21 chromosomes du blé. Suite à des difficultés techniques, Les résultats ne sont disponibles que pour 13 marqueurs, qui se sont tous avérés polymorphes dans les échantillons étudiés. Les protocoles PCR ont été adaptés de Röder *et al.* [32] et Guyomarc'h *et al.* [16]. Les bandes amplifiées ont été séparées sur un séquenceur ABI 3100 semi-automatique, et analysées avec GeneMapper software (v 3.7, Applied Biosystems).

2.2.4. Analyses de la diversité génétique

Pour chaque population, les fréquences alléliques, le nombre moyen d'allèles, la diversité de Nei [24] ou hétérozygotie attendue H_e , l'hétérozygotie observée H_o et le coefficient de consanguinité ($F_{IS} = 1 - H_o/H_e$) ont été estimés à l'aide du logiciel GENETIX v4.05.2 [3]. La différenciation génétique entre populations (proportion de diversité totale expliquée par l'écart des fréquences alléliques entre populations) a été évaluée à travers l'estimateur θ proposé par Weir et Cockerham [36], estimé ici avec le logiciel FSTAT version 2.9.3.2 [15]. La signification des valeurs de θ estimées entre populations a été testée par des permutations de génotypes entre les populations échantillonnées dans FSTAT. La proximité des populations a été visualisée à l'aide d'un arbre construit selon la méthode UPGMA à partir des θ , dont la robustesse des nœuds a été évaluée avec

1000 bootstraps sur les locus. Trois variétés sans rapport de parenté avec le Rouge de Bordeaux ont été utilisées comme « out groups ». Nous avons également réalisé une analyse factorielle des correspondances (AFC) à l'aide du logiciel GENETIX v4.05.2 [3] afin de décrire la distribution des individus entre et à l'intérieur des populations.

3. RÉSULTATS

3.1. Organisation de la diversité inter-populations

Les résultats qui suivent sont à prendre avec précaution, étant donné le faible nombre de marqueurs (13) et le taux de données manquantes relativement élevé (respectivement 254, 40, 173, 5, 6, 62, 289, 412, 45, 35, 15, 37 et 130 sur 938 individus génotypés, soit des taux allant de 0,5% à 43,9%)§.

Les indices de différenciation par paires de populations estimés par le paramètre θ sont présentés dans le tableau I, avec leurs *P-values* associées. Les populations sont en majorité significativement différenciées et présentent des θ élevés. En revanche, certaines apparaissent très proches génétiquement, comme l'accession « référence » de la collection, Col6310, et les populations de Denis (2003 et 2006) et Basile (2003 surtout), celles-ci étant également proches entre elles. Les populations de trois autres agriculteurs (Gaston, Lucien et Jules) apparaissent également très proches génétiquement. Les variétés conservées ou cultivées par un même agriculteur mais avec des pratiques différentes (RB-BasileCons06 et RB-BasileSol06, RB-Basile-2ep03 et RB-Basile-Mel03), ou bien prélevés à des générations différentes (RB-Denis03 et RB-Denis06, RB-GastonSel03 et RB-Gaston06) sont également très proches génétiquement. Parfois, s'il s'agit d'une sélection de quelques épis au sein d'une variété (Rouge-Roc05 et RB-GastonSel03), on peut observer une différenciation significative.

§ Des analyses sont en cours pour compléter ces données.

Tableau 1: The θ (estimateur du F_{st} de Weir et Cockerham, 1984) par paire de populations (symétrique au-dessus de la diagonale) et P-values obtenues avec 5000 permutations (symétrique sous la diagonale)

Théta	RB- Aimé	RB- Basile Cons06	RB- Basile Isd06	RB- Denis 03	RB- Denis 06	Rouge- Charml les	RB- Gasbn 06	Rouge- Roc05	RB- Jules	RB- Lucien	RB- Marcel	RB- PaulIN2	RB- PaulIN3	Col- C d	Col- C d	RB- Basile 2ep03	
RB-Aimé	0.7942	0.8204	0.8080	0.8710	0.8465	0.3361	0.0649	0.1375	0.0557	0.1691	0.8480	0.0160	0.3763	0.8699	0.8207	0.8376	0.2605
RB-Basile Cons06	0.0002	0.0000	0.1647	0.1571	0.9540	0.9344	0.5123	0.7841	0.5988	0.4445	0.7452	0.9046	0.3750	0.3066	0.7214	0.9488	0.8573
RB-Basile Isd06	0.0002	0.6553	0.2893	0.2893	0.9228	0.9245	0.5546	0.8062	0.6400	0.4929	0.8403	0.9046	0.9566	0.3575	0.7088	0.9350	0.8634
RB-Denis03	0.0002	0.1850	0.1443	0.0257	0.9135	0.9041	0.5846	0.8578	0.6496	0.5247	0.5737	0.8937	0.9319	0.0000	0.8575	0.9297	0.8380
RB-Denis06	0.0003	0.0016	0.0002	1.0000	0.9476	0.9287	0.6240	0.8642	0.6962	0.5638	0.8130	0.9334	0.9596	0.0170	0.9096	0.9666	0.8669
Rouge- Charmlies	0.0003	0.0002	0.0002	0.0022	0.9263	0.9662	0.6662	0.8592	0.7482	0.6564	0.9531	0.9268	0.9608	0.9453	0.9106	0.9541	0.8719
RB-Emile	0.0003	0.0002	0.0002	0.0002	0.0002	0.2807	0.2807	0.3953	0.3648	0.4020	0.9243	0.4052	0.0701	0.9262	0.9153	0.0509	0.9244
RB-Gaston06	0.3792	0.0002	0.0002	0.0710	0.0235	0.0063	0.0000	0.1131	0.0000	0.0024	0.5649	0.1744	0.3187	0.5939	0.6727	0.6850	0.5867
Rouge- Roc05	0.0127	0.0111	0.0130	0.0002	0.0002	0.0004	0.0002	0.0000	0.1069	0.2246	0.8488	0.1902	0.4753	0.8325	0.8546	0.8248	0.2713
RB-Jules	0.2158	0.0004	0.0002	0.0027	0.0046	0.0002	1.0000	0.0009	0.0025	0.6342	0.1983	0.4139	0.6666	0.7285	0.7500	0.2636	0.6577
RB-Lucien	0.1614	0.0002	0.0002	0.0069	0.0012	0.0002	0.5257	0.0004	0.3810	0.4861	0.3005	0.4380	0.5360	0.6564	0.6989	0.3117	0.5220
RB-Marcel	0.0003	0.0002	0.0002	0.0694	0.0213	0.0030	0.1652	0.0002	0.2951	0.1433	0.9276	0.9654	0.6937	0.8759	0.9649	0.6411	0.7321
RB-PaulIN2	0.2629	0.0002	0.0002	0.0006	0.0002	0.0002	1.0000	0.0005	0.0202	0.0093	0.0111	0.5114	0.9181	0.9112	0.9168	0.3037	0.9206
RB-PaulIN3	0.0003	0.0002	0.0002	0.0002	0.0002	0.2638	0.0044	0.0002	0.0002	0.0002	0.0030	0.0002	0.9582	0.9395	0.9487	0.0772	0.9562
Col6310	0.0003	0.0042	0.0002	NA	0.6435	0.0002	0.0061	0.0002	0.0008	0.0002	0.0002	0.0002	0.0002	0.8847	0.9531	0.8573	0.0000
Col23760	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0024	0.0002	0.0002	0.0002	0.0044	0.0002	0.0002	0.9368	0.8629	0.8912	0.0000
Col23756	0.0003	0.0002	0.0002	0.0010	0.0002	0.0002	0.0129	0.0004	0.0030	0.0002	0.0121	0.0002	0.0002	0.0002	0.0002	0.8516	0.9536
Col23761	0.0003	0.0002	0.0002	0.0002	0.0002	0.0024	0.0053	0.0002	0.0006	0.0002	0.0040	0.0002	0.0014	0.0002	0.0002	0.8514	0.9536
RB-Basile-2ep03	0.0003	0.0063	0.0002	1.0000	1.0000	0.0002	0.0204	0.0002	0.0057	0.0006	0.0200	0.0002	0.0002	0.4142	0.0002	0.0002	0.0002
RB-Basile-Miel03	0.0003	0.0032	0.0002	NA	1.0000	0.0002	0.0040	0.0002	0.0006	0.0002	0.0036	0.0002	0.0002	0.0601	0.0002	0.0002	0.2830
RB-Rem03	0.0003	0.0002	0.0002	0.0014	0.0002	0.5838	0.0188	0.0002	0.0042	0.0004	0.0134	0.0006	0.0632	0.0002	0.0002	0.0002	0.0032
RB-Gaston-Set03	0.0029	0.0004	0.0002	0.0006	0.0046	0.0002	0.0125	0.0002	0.0123	0.0123	0.0741	0.0002	0.0004	0.0002	0.0002	0.0002	0.0079

Les valeurs de théta en gras sont significatives au seuil 0.0002, soulignées, elles sont significatives à 1% et en italique, elles sont proches de 0 (inférieures à 1%).

L'arbre représentant l'organisation globale des différentes populations à partir des indices de différenciation θ (fig. 2) montre les deux groupes déjà décelés dans la matrice : le groupe (1) avec les populations des agriculteurs Basile et Denis associés au numéro 6310 de la collection d'un côté, et le groupe (2) avec les populations des agriculteurs Gaston, Lucien et Jules qui agglomèrent ensuite ceux de Paul, Rémi, Émile et Aimé ainsi que le numéro

23761 de la collection. Les trois variétés non apparentées utilisées comme « out-groups » s'enracinent bien à la base de l'arbre, mais on retrouve également là, à l'extérieur du groupe des Rouges de Bordeaux, deux accessions de la collection et le Rouge des Charmilles. La variété RB de Marcel se rattache faiblement au premier groupe.

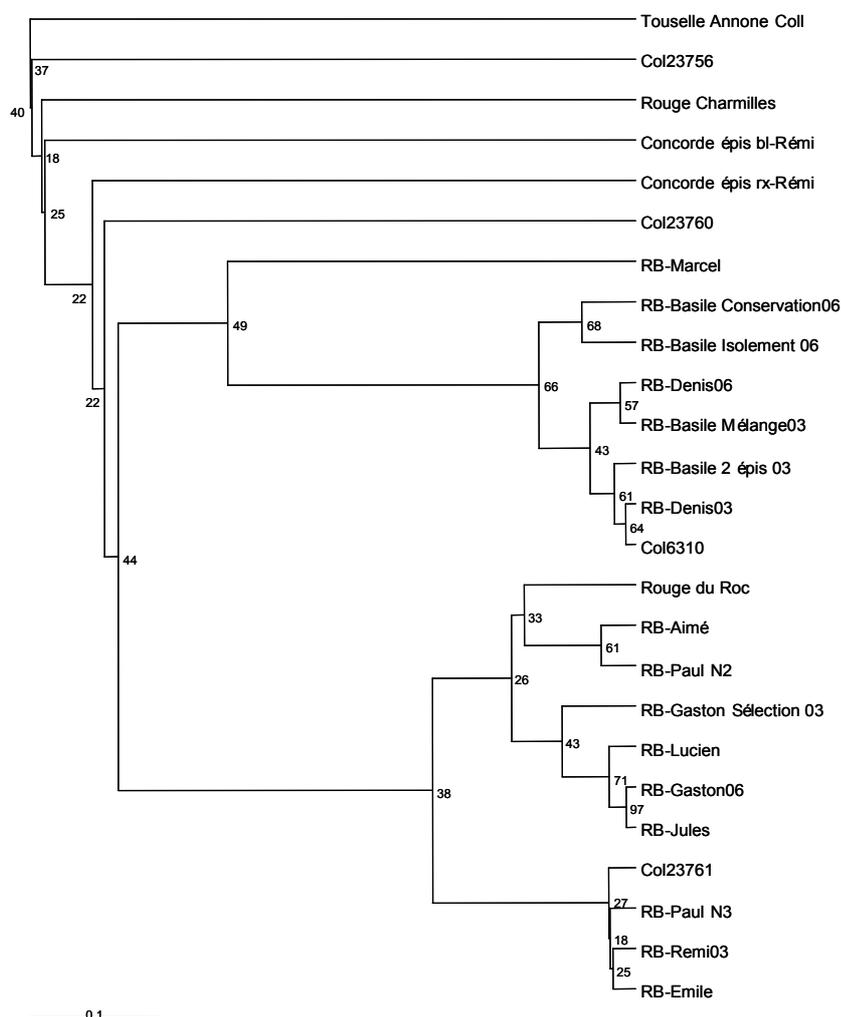


Figure 2 : Arbre reconstruit par la méthode UPGMA à partir des indices de différenciation entre populations estimés par le paramètre θ sur les données de génotypage microsatellite (13 locus). Les valeurs de bootstrap des nœuds ont été obtenues avec 1000 tirages des locus. Les variétés «Touselle Annone Coll», «Concorde épis bl-Rémi», «Concorde épis rx-Rémi» ont été utilisées comme «out-groups» car elles ne sont pas apparentées à la variété Rouge de Bordeaux.

La projection des individus des groupes de RB identifiés ci-dessus dans le premier plan de l'AFC (fig.3) permet de voir que tous les individus de l'accession 6310 de la collection sont bien regroupés (en haut à droite), associés aux individus appartenant aux populations de Basile et Denis (groupe (1)). Aucun d'entre eux ne se trouve projeté dans le quartier en bas à droite où se trouvent la plupart des individus du groupe (2). Au contraire, de nombreux individus des populations de Gaston et Lucien se retrouvent dans le quartier en haut à droite, indiquant une plus forte hétérogénéité au sein de ces populations et/ou un mélange de génotypes.

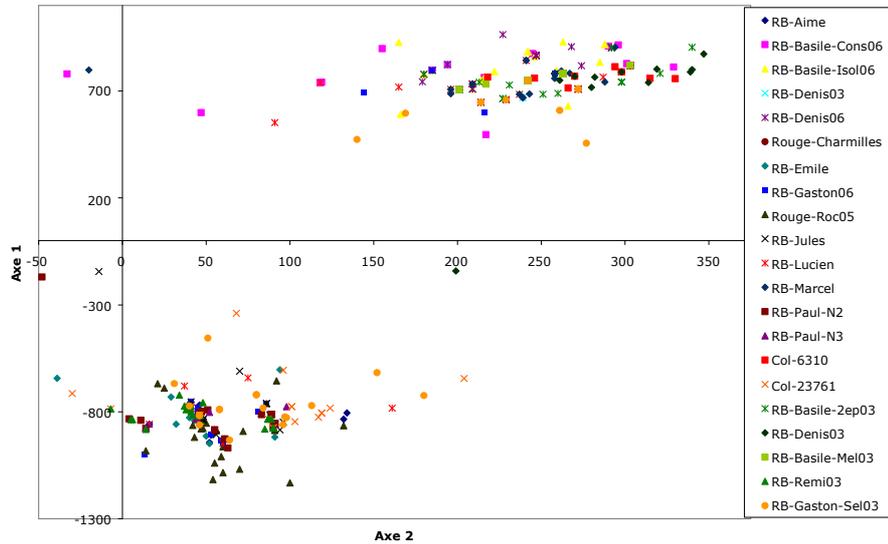


Figure 3 : Projection des individus dans le plan défini par les deux premiers axes de l'AFC réalisée sur les données de génotypage microsatellite (13 locus). La figure se focalise sur un secteur du plan qui contient les deux groupes de RB étudiés identifiés sur l'arbre. L'axe 1 explique 11,32 % de la variation et l'axe 2, 6,45 %.

3.2. Diversité intra-populations

Seul un échantillon est complètement homogène génétiquement (RB-Basile-Mel03), mais un bon nombre est assez homogène avec une diversité génétique intra-population inférieure à 10% (tabl. II). C'est le cas pour trois des quatre accessions de la collection et pour toutes les populations de Denis et de Basile, c'est-à-dire pour tous ceux du groupe (1). En plus, les populations d'Émile, Marcel et Paul ainsi que le Rouge des Charmilles ont une diversité du même ordre.

Tableau 2: Paramètres de diversité intra-échantillon.

	RB- Aimé	RB- Basile Cons06	RB- Basile Iso06	RB- Denis06	RB- Rouges Charmilles	RB- Emile06	RB- Gaston Rouge05	RB- Jules Lucien Marcel	RB- Paul N3	RB- Col 6310	RB- Col 23760	RB- Col 23756	RB- Col 23761	RB- Basile Meil03	RB- Basile Meil03	RB- Gaston Remi03								
He (écart-type)	0.151	0.066	0.044	0.059	0.016	0.026	0.067	0.367	0.181	0.313	0.365	0.016	0.070	0.026	0.028	0.069	0.037	0.152	0.027	0.000	0.110	0.310		
Ho (écart-type)	0.131	0.135	0.139	0.116	0.025	0.050	0.084	0.199	0.181	0.210	0.236	0.046	0.098	0.066	0.089	0.163	0.106	0.130	0.059	0.000	0.084	0.248		
He (écart-type)	0.009	0.151	0.013	0.032	0.007	0.009	0.027	0.032	0.013	0.046	0.012	0.015	0.019	0.022	0.044	0.039	0.056	0.035	0.000	0.019	0.055			
Moyen d'ailles/lo	1.846	1.25	1.231	2	1.385	1.3077	2	3.4615	23.846	22.31	2.4615	1.1538	1.923	1.308	1.231	1.539	2.308	1.308	1	2.462	2.3077			
Fis	0.984	0.036	0.922	0.692	0.887	0.904	0.822	0.95	0.817	0.9284	0.968	0.799	0.892	0.717	0.701	0.757	0.528	0.603	0.492	NA	0.933	0.914		

He : diversité de Nei ou hétérozygotie attendue (estimateur non biaisé selon Nei, 1978)

Ho : hétérozygotie observée

Au contraire, les populations de Gaston, Jules et Lucien présentent des diversités extrêmement fortes, supérieures à 30%. Des valeurs intermédiaires sont obtenues pour les RB de Aimé et Rémi, ainsi que pour le Rouge du Roc et le numéro 23761 de la collection. Similairement, le nombre moyen d'allèles détectés dans chaque population varie de 1 (RB-Basile-Mel03) à 3,46 (RB-Gaston06) (tabl. II). L'hétérozygotie observée donne en théorie une indication sur les croisements et permet l'estimation du taux d'allogamie. Nous ne l'avons pas fait ici car une partie des marqueurs microsatellites présentait des profils ambigus pour les hétérozygotes dus à la présence de stutters. Il faudra l'estimer sur le jeu de marqueurs le plus approprié lorsque nous aurons complété les données de génotypage. Ici, les taux d'hétérozygotie sont le plus souvent inférieurs à 1 % (populations d'Aimé, Jules, Marcel, Paul, Rémi, accession de la collection 6310 et populations RB-Basile-Isol06, RB-Basile-Mel03, RB-Denis06). Les autres sont compris entre 1 et 7%. Ces valeurs faibles sont cohérentes avec le régime de reproduction majoritairement autogame du blé, mais indiquent malgré tout l'existence quasi-systématique de quelques croisements au moins.

3.3. Variabilité phénotypique pour les caractères de précocité

L'évaluation phénotypique en station montre que des populations d'origines différentes sont significativement différentes pour leur développement précoce, leur date d'épiaison en conditions vernalisées (fig. 4) et leur besoin en vernalisation. L'héritabilité pour la date d'épiaison en conditions vernalisées varie de 0.71 à 0.87 indiquant qu'une part prédominante de la variation entre échantillons est due à des différences génétiques. De plus, les populations de RB montrent des niveaux de variabilité différents. Par exemple, l'étalement au sein d'une population de la date d'épiaison en conditions vernalisées varie de 2 (collection 23761) à 10 jours (RB-Gaston06, RB-Lucien, RB-Jules) selon l'origine de la variété-population (fig. 4). On trouve ainsi une relation assez nette entre la diversité génétique moléculaire intra-population et la variabilité pour le caractère de précocité d'épiaison, caractère majeur dans l'adaptation climatique. La distribution des types printemps et hiver, déterminés à partir de l'observation en conditions non vernalisées avec semis au printemps, est également très variable allant de 0 à 100% de plantes hiver. Si l'on excepte le Rouge des Charmilles et les deux accessions de la collection 23756 et 23760 qui sont entièrement constituées de plantes de type « hiver », des plantes de type « printemps » sont systématiquement présentes, la population RB-Denis03 étant la plus « hiver » alors que les RB provenant des agriculteurs Émile, Paul, Rémi et Aimé sont très majoritairement « printemps ». L'accession 6310 de la collection présente 65% de plantes « printemps ».

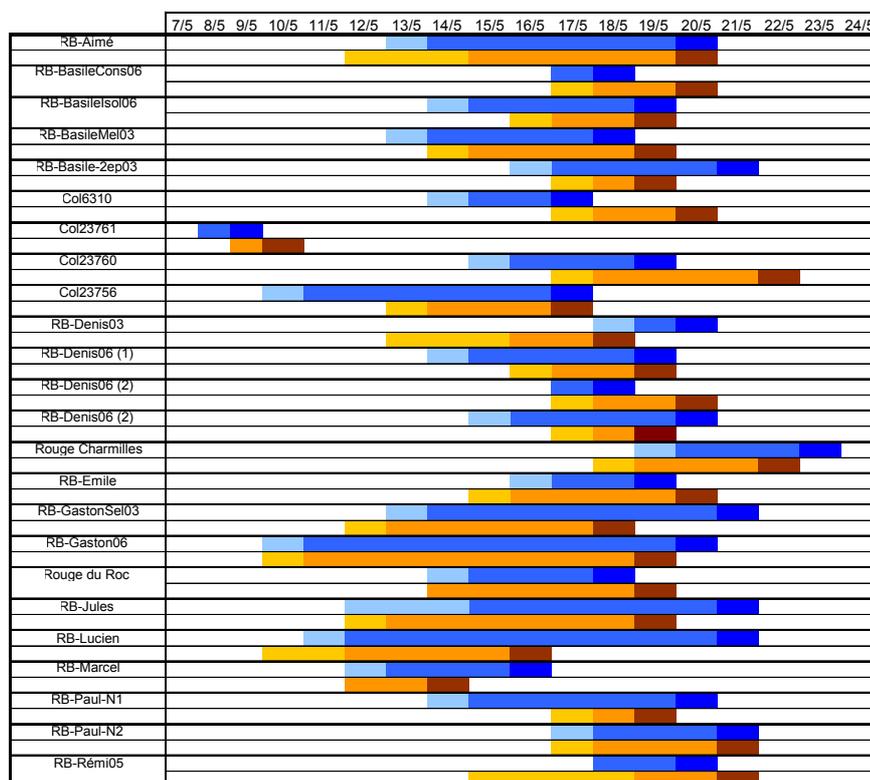


Figure 4 : Distribution des dates d'épiaison des plantes au sein des différentes populations dans le dispositif d'évaluation à la station expérimentale INRA du Moulon avec semis à l'automne 2006. Pour chaque échantillon sont données les dates d'épiaison relevées dans la répétition 1 (la ligne du dessus) et celles relevées dans la répétition 2 (ligne du dessous). En clair, la date correspondant à une ou deux plantes de la parcelle épiées, en plus sombre, l'étalement des épiaisons de 5 à 90 % des plantes, et en très sombre, la date correspondant à 100 % des plantes de la parcelle épiées.

4. DISCUSSION

L'analyse génétique des échantillons de la même variété-population, obtenus auprès de paysans et auprès des gestionnaires de la Collection Nationale montre que la diversité conservée dans les champs et dans les chambres froides est différente, à plusieurs points de vue. Cette remarque n'interdit pas des ressemblances entre les populations cultivées dans certains champs et celles conservées à la collection, en témoigne la ressemblance génétique entre les lots fournis par Basile, Denis et la référence Col6310.

D'une part, les niveaux d'hétérogénéité génétique diffèrent. Les populations caractérisées par une faible variabilité génétique (accessions de la collection, mais également Basile, Denis) sont gérées dans un souci de conservation à l'identique. Cela se traduit par des précautions à différents moments de la mise en culture (tous les dix ans pour la collection lors de la régénération des graines conservées) : isolement des parcelles, évitement des mélanges, élimination des hors-types à la récolte. Le matériel génétique utilisé par ces trois acteurs, issu originellement de la collection Verneuil, héritière de la collection historique Vilmorin, avait par ailleurs pour caractéristique une base génétique relativement faible liée à sa gestion statique conservatrice sur plusieurs décennies.

D'autre part, les différentes populations étudiées se structurent majoritairement en deux groupes génétiquement distincts. Cette structuration s'explique relativement bien par la circulation de semences qui a eu lieu entre le groupe Denis-Basile-Collection INRA d'une part et le groupe Gaston, Jules, Rémi, Paul d'autre part. Il est à noter (ce que ne montre pas la carte des échanges) que le transfert de semences de l'INRA Moulon vers Gaston, ainsi que celui de Basile vers Gaston est relativement récent, en tout cas ultérieur à la diffusion des semences dans le réseau en aval de Gaston – si bien que les semences diffusées par Jules ou Rémi n'ont pas d'origine commune avec celles issues de la collection Vilmorin-Verneuil. Ces semences sont issues de la Communauté de l'Arche (notée Arche 47), une communauté d'inspiration gandhienne créée dans les années soixante-dix. Cette observation nous amène à formuler l'hypothèse du rôle de passeur intergénérationnel qu'auraient joué les communautés de néo-ruraux, du fait de l'intérêt précurseur qu'elles ont porté aux variétés paysannes et locales et de leur position centrale et visible au sein de certains réseaux agricoles [18].

L'adaptation locale aux conditions du milieu et aux pratiques paysannes peut également expliquer une part de la structuration de la diversité. La précocité d'épiaison, ainsi que les besoins en vernalisation, sont des caractères majeurs de l'adaptation climatique, et leur diversification au sein de populations d'origine commune est à mettre en relation avec les résultats de l'expérimentation Gestion dynamique [14], [31]. Les pratiques paysannes sont elles aussi diversifiantes, comme en témoigne par exemple le semis accidentel par Aimé de son RB au printemps, qui a contribué à la contre-sélection dans sa population des types « hiver ». La sélection, ici induite, est parfois délibérée : ainsi la population Rouge du Roc est-elle une sélection d'épis « barbus » réalisée par Gaston dans une parcelle de RB qui aboutit à une certaine différenciation par rapport au génotype de la population d'origine. Ainsi la diversité des populations de blé cultivées est à mettre en relation à la fois avec la diversité des conditions pédoclimatiques et avec la

diversité des stratégies paysannes [2], [23], [22], [28], [29]. Chercher à optimiser la conservation à la ferme en préconisant aux agriculteurs un type donné de gestion pourrait conduire à un effet paradoxal, à savoir l'uniformisation des choix et des pratiques, alors que la propriété diversifiante de ce dispositif repose précisément sur la diversité des environnements et des stratégies de culture.

La grande hétérogénéité de la variété-population Rouge de Bordeaux, constatable ici à travers l'éclatement génétique des populations qui la représentent, nous interroge sur la délimitation de cette catégorie variétale. La majorité des personnes apparaissant dans le réseau d'échange identifient le lot de semences qu'ils reçoivent grâce aux indications qui leur sont simultanément fournies. Par le nom, attribut essentiel des semences, mais aussi par les informations associées lors de l'échange (usages, origine, histoire), une certaine continuité génétique et culturelle des variétés est ainsi assurée. La dénomination des semences est d'autant plus importante de nos jours que la disparition des variétés anciennes dans le paysage agricole s'est accompagnée de l'oubli de leurs caractères phénotypiques. Mais l'identification des populations majoritairement par leur filiation n'exclut pas le recours ponctuel aux critères phénotypiques. Cette étude nous en fournit trois exemples : (i) la « variété » Rouge des Charmilles est issue d'une opération de re-nomination par un paysan, d'une population initialement qualifiée de RB, mais qu'il jugeait atypique (elle est de fait éloignée génétiquement des autres populations de RB) ; (ii) la population de Marcel a été constituée sur critères phénotypiques à partir d'un mélange de 20 variétés qui contenait initialement du RB ; (iii) du côté de la collection, deux des accessions notées « type RB » (Col23760 et Col23756) se sont avérées après enquête être des blés de Haute-Loire, à épis rouges, rebaptisés « type RB » lors de leur entrée dans la collection**.

Cette définition multiple de la *variété*, aux contours imprécis et laissés à l'appréciation de chacun, pose la question de la nature de ce qui est conservé/cultivé sous un nom variétal donné : le type qui correspond le plus fidèlement possible au type conservé dans la collection historique ? Une diversité génétique issue de la variété-population d'origine, mais qui aurait évolué au cours depuis sous l'effet des pressions de sélection rencontrées pendant des décennies de cycles de culture ? Un ensemble génétiquement

** L'absence apparente de filiation de ces deux accessions avec la variété-population RB explique leur position excentrée dans l'arbre par rapport aux autres échantillons de RB. C'est grâce aux informations recueillies lors de la mise en collection de l'accession que nous avons pu revenir à la personne qui avait fourni l'échantillon, et parachever l'identification de la variété-population d'origine. D'où l'importance de noter lors du recueil de semences, à la ferme, le maximum d'informations permettant d'en reconstituer la généalogie, en partenariat avec les détenteurs de savoirs vivants que sont les agriculteurs collectionneurs.

hétérogène mais exprimant des phénotypes convergeant vers la description phénotypique de la variété ? Une reprise des analyses avec plus de marqueurs et une analyse des génotypes multilocus nous permettraient d'établir plus précisément les relations entre les individus appartenant à des échantillons différents et de savoir s'il s'agit de recombinants ou de mélanges de génotypes d'origine hétérogène. Eet ainsi d'évaluer la part de variabilité issue de recombinaisons génétiques au sein de la variété-population RB ancestrale, et celle associée à des mélanges avec des populations d'origine différente. Il faut noter cependant que les échanges et mélanges de semences ainsi que les flux de gènes non contrôlés font partie intrinsèque de la dynamique des systèmes traditionnels de gestion à la ferme et ne remettent pas en question le « périmètre » phénotypique des variétés-populations ainsi maintenues, comme cela a été décrit notamment dans le cas des variétés de maïs au Mexique [28], [29]. Les flux de gènes (force homogénéisante) sont alors contrebalancés par la sélection divergente (force diversifiante). Dans ces contextes, les variétés cultivées à la ferme dérivent bien d'une variété ancestrale, tout en incorporant une part de diversité exogène.

Même si la caractérisation des mécanismes en jeu et la connaissance de la nature de la diversité conservée dans la gestion à la ferme sont encore partielles, il nous semble que nous présentons ici les premiers éléments indiquant que des réseaux d'agriculteurs en France apportent à la gestion de la diversité génétique une contribution positive et originale, complémentaire de celle, centrale des collections *ex situ*. Ces éléments nous conduisent à inciter l'ensemble des acteurs de la gestion du vivant à développer, pour plus de sécurité et d'efficacité des dispositifs de conservation, un passage plus régulier des ressources d'un compartiment à l'autre :

- **Des champs à la collection**, afin d'assurer la conservation de ressources encore menacées d'érosion du fait de leur insuffisante diffusion dans le paysage agricole français. L'analyse génétique a montré la façon dont l'accession Col-23761 et la population Rouge du Roc dont elle était issue avaient divergé, alors que la première était censée représenter la deuxième. Il conviendra pour éviter cet écueil de veiller à prélever des échantillons sur un volume significatif et de procéder à des prélèvements réguliers rendant compte de l'évolution des populations dans leur milieu agricole.
- **De la collection aux champs**, afin de permettre aux ressources génétiques conservées de se rediversifier en évolution conjointe avec les agrosystèmes, et de permettre aux agriculteurs l'usage de variétés reléguées par le marché et, de ce fait, difficilement accessibles ailleurs.

Ainsi, il conviendrait pour une meilleure gestion des ressources génétiques, de renforcer les partenariats pour l'heure informels entre agriculteurs et gestionnaires de collection.

La reconnaissance de cette gestion évolutive par les acteurs institutionnels de la conservation des ressources génétiques nous apparaît aujourd'hui incontournable, ce qui nous conduit des questions d'ordre stratégique. Entre deux objectifs complémentaires mais parfois contradictoires, la conservation à l'identique des variétés patrimoniales et le développement du potentiel évolutif des ressources génétiques, quelle priorité les acteurs institutionnels de la conservation se donnent-ils ? Sachant Au regard de la biodiversité cultivée dans les champs, comment envisagent-ils d'organiser et de coordonner leurs efforts avec ceux des paysans ? Par ailleurs, ces résultats conduisent à des considérations sur la nécessité de faire évoluer la réglementation. Actuellement, le règlement UPOV en vigueur en France interdit toute transaction (marchande ou non) de semences autres que celles correspondant aux variétés inscrites au catalogue officiel des obtentions végétales. Si l'Europe permet des mesures d'exception pour la conservation *in situ* des variétés dites « de conservation » (directive européenne 98/95CE), si d'autres pays d'Europe (l'Italie par exemple) semblent aujourd'hui plus avancés dans la prise en compte réglementaire de la conservation à la ferme, la France dispose d'un nombre d'expériences et d'un recul sur celle-ci qui la place en bonne position pour contribuer significativement à l'aménagement de cette réglementation. Afin d'avancer dans ce sens, une étude basée sur des expérimentations et l'analyse de cas est en cours actuellement dans le projet européen STREP « FarmSeedOpportunities » (<http://www.farmseed.net/home/>).

REMERCIEMENTS

Nous remercions les agriculteurs qui ont eu la patience et la gentillesse de répondre à nos questions, ainsi que ceux qui se sont prêtés à cette recherche en participant à l'enquête ou en intégrant leur variété au protocole expérimental. Nous remercions Xavier Raffoux, Delphine Madur, Valérie Combes et Fabrice Dumas pour l'extraction des ADN, ainsi que Charles Poncet et la plate-forme de Clermont-Ferrand pour le génotypage microsatellite. Merci également à Julie Dawson, qui a assuré les traductions en anglais. Cette recherche s'est déroulée dans le cadre du programme bisannuel (2005-2007) « Étude des complémentarités entre modes de gestion dynamique (à la ferme ou en station) et statique à partir du cas du

blé tendre », piloté par I. Goldringer et financé par le Bureau des Ressources Génétiques.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Almekinder C. J. M., de Boef W., Engels J., 2000. - Synthesis between crop conservation and development. In: Almekinders C., de Boef W. (Eds), Encouraging Diversity. The conservation and development of plant genetic resources. Intermediate Technology Publications, London, pp. 330-338.
- [2] Alvarez N., Garine E., Khasa C., Dounias E., Hoassaert-McKey M., McKey D., 2005. - Farmer's practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroun. *Biological Conservation* 121: 533-543.
- [3] Belkhir K., Borsa P., Chikhi L., Raufaste N., Bonhomme F., 2004. - GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France), <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>.
- [4] Berthaud J., Clément J. C., Emperaire L., Louette D., Pinton F., Sanou J., Second G., 2001. - The role of local level gene flow in enhancing and maintaining genetic diversity. In: Cooper H.D, Spillane C, Hodgkin T (Eds) Broadening the genetic diversity of crop production, CABI Publishing in association FAO and IPGRI, Rome, Italy, pp. 81-103.
- [5] Bonneuil C., Demeulenaere E., 2007. - Vers une génétique de pair à pair ? L'émergence de la sélection participative, in Charvolin F., Micoud A., Nyhart L.K.. (Eds.), *Des sciences citoyennes ? La question de l'amateur dans les sciences naturalistes*. L'Aube, La Tour d'Aigues, pp. 122-147
- [6] Bonneuil C., Thomas F., 2008. - Gènes, pouvoirs et profits. La recherche publique dans les transformations des régimes de production des savoirs en génétique végétale de Mendel aux OGM. à paraître, Quae, Paris.
- [7] Bustarret J., 1944. - Variétés et variations. AA: 336-362.
- [8] Carrascosa Garcia M., 2003. - Enquêtes sur les expériences et savoir-faire "Semences paysannes, biologiques et biodynamiques" dans les campagnes françaises. [En ligne sur : http://www.semencespaysannes.org/images/FCK/file/publications/etude_carrascosa.pdf]
- [9] Elias M., McKey D., Panaud O., Mc Anstett, Robert T., 2001. - Traditional management of cassava morphological and genetic diversity by the Makushi Amerindians (Guyana, South America): perspectives for on-farm conservation of crop genetic resources. *Euphytica*, 120: 143-157.
- [10] Emperaire L., 2000-2004. — La biodiversité agricole en Amazonie brésilienne : ressource et patrimoine. *JATBA*, 42: 113-126.
- [11] Enjalbert J., Goldringer I., Paillard S., Brabant P., 1999. - Molecular markers to study genetic drift and selection in wheat populations. *Journal of experimental botany*, 50: 283-290.

- [12] Goldringer I., Enjalbert J., Raquin A.-L., Brabant P., 2001. - Strong selection in wheat populations during ten generations of dynamic management. *Genetics, Selection, Evolution*, 33: S441-S463.
- [13] Goldringer I., Enjalbert J., David J., Paillard S., Pham J.-L., Brabant P., 2001. - Dynamic management of genetics resources a 13-year experiment on wheat. In: Cooper H.D., Spillane C., Hodgkin T. (Eds), *Broadening the Genetic Base of Crop Production*. IPGRI/FAO, Rome, Italie, pp. 245-260.
- [14] Goldringer I., Prouin C., Rousset M., Galic N., Bonnin I., 2006. - Rapid differentiation of experimental populations of wheat for heading time in response to local climatic conditions. *Annals of Botany*, 98: 805-817.
- [15] Goudet J., 2002. - FSTAT version 2.9.3.2. A program to estimate and test gene diversities and fixation indices. Institut of Ecology, Lausanne, Switzerland, Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995).
- [16] Guyomarc'h H., Sourdille P., Charmet G., Edwards K.J., Bernard M., 2002. - Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. *Theoretical Applied Genetics*, 104: 1164-1172.
- [17] Henry J.-P., Pontis C., David J.-L., Gouyon P.-H., 1991. - An experiment on dynamic conservation of genetic resources with metapopulations, In: Seitz A., Loeschcke V. (eds), *Species Conservation: a Population Biological Approach*, Basel, pp. 185-198.
- [18] Hervieu-Léger D., Hervieu B., 2005. — *Le retour à la nature*. L'Aube, La Tour d'Aigues (1^{ère} éd. 1979, Seuil).
- [19] Lavigne C., Reboud X., Lefranc M., Porcher E., Roux F., 2001. - Evolution of genetic diversity in metapopulations: *Arabidopsis thaliana* as an experimental model. *Genetics, Selection, Evolution* 33: S399-S423.
- [20] Le Boulc'h V., David J., Brabant P., de Vallavieille-Pope C., 1994. Dynamic conservation of variability: responses of wheat populations in different selective forces including powdery mildew, *Genetics Selection Evolution*, 26: S221-S240.
- [21] Louette D., Charrier A., Berthaud J., 1997. - *In situ* conservation of maize in Mexico genetic diversity and maize seed management in a traditional community. *Econ. Bot.* 51: 20-38.
- [22] Louette D., Smale, 2000. - Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. *Euphytica*, 113: 25-41.
- [23] McKey D., Emperaire L., Elias M., Pinton F., Robert T., Desmoulière S., Rival L., 2001. - Gestions locales et dynamiques régionales de la diversité variétale du manioc en Amazonie. /*Genetics, Selection et Evolution* 33 (supplement 1): S465-S490.
- [24] Nei M., 1987. - *Molecular Evolutionary Genetics*. Columbia University Press, New York, 512 p.
- [25] Paillard S., Goldringer I., Enjalbert J., Doussinault G., de Vallavieille-Pope C., Brabant P., 2000a. - Evolution of resistance against powdery mildew in winter wheat populations conducted under dynamic management. I- Is specific seedling resistance selected? *Theoretical Applied Genetics*, 101: 449-456.
- [26] Paillard S., Goldringer I., Enjalbert J., Trottet M., David J., De Vallaveille-Pope C., Brabant P., 2000b. - Evolution of resistance against powdery mildew in win-

- ter wheat populations conducted under dynamic management. II- Adult plant resistance. *Theoretical and Applied Genetics*, 101: 457-462.
- [27] Porcher E., Giraud T., Goldringer I., Lavigne C., 2004. - Experimental demonstration of a causal relationship between heterogeneity of selection and genetic differentiation at quantitative traits. *Evolution*, 58: 1434-1445.
- [28] Pressoir G., Berthaud J., 2004a. - Pattern of population structure in maize landraces from the central valley of Oaxaca in Mexico.. *Heredity*, 92: 88-94.
- [29] Pressoir G., Berthaud J., 2004b. - Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity*, 92: 95-101.
- [30] Reynolds M.P., Borlaug N.E., 2006. - Applying innovations and new technologies for international collaborative wheat improvement. *Jour. Agri. Sci.*, 144: 95-110.
- [31] Rhoné B., Remoué C., Galic N., Goldringer I., Bonnin I., 2008. - Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. *Molecular Ecology*, 17: 930-943.
- [32] Röder M., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., Ganal M., 1998. - A microsatellite map of wheat. *Genetics*, 149: 2007-2023.
- [33] Roussel V., Koenig J., Beckert M., Balfourier F., 2004. - Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. *Theoretical and Applied Genetics*, 108: 920-930.
- [34] RSP, 2003. -- Cultivons la biodiversité dans les fermes, Actes des premières rencontres Semences Paysannes, Auzeville, 27-28 février 2003. [En ligne : http://www.semencespaysannes.org/images/imagesFCK/file/publications/actes_auzeville.pdf]
- [35] Smith M.E., Fernando C.G., Gomez F., 2001. - Participatory plant breeding with maize in Mexico and Honduras. *Euphytica*, 122: 551-565.
- [36] Weir B.S., Cockerham C.C., 1984. - Estimating F-statistics for the analysis of population structure. *Evolution*, 38: 1358-1370.

Diversité agricole et patrimoine dans le moyen Rio Negro (Amazonie brésilienne)

Laure EMPERAIRE^{(1)*}, Pascale de ROBERT⁽²⁾, Juliana SANTILLI⁽³⁾,
Ludivine ELOY⁽¹⁾, Lúcia van VELTHEM⁽⁴⁾, Esther KATZ⁽¹⁾,
Cláudia LOPEZ⁽⁴⁾, Anne-Elisabeth LAQUES⁽⁵⁾,
Manuela CARNEIRO DA CUNHA⁽⁶⁾, Mauro ALMEIDA⁽⁷⁾

⁽¹⁾IRD – UR 200, CP 7091 – Lago Sul, 71619-970, Brasília, DF, Brésil,

⁽²⁾IRD – UR 200, MNHN, HNS, CP026, 57 rue Cuvier,
75231 Paris cedex 05, France

⁽³⁾ISA – SCLN 210, Bloco C, sala 112, Brasília, 70346- DF, Brésil

⁽⁴⁾MPEG-MCT, Av. Magalhães Barata 376, 66040-170, Belém, PA, Brésil

⁽⁵⁾Dpt Géographie, Univ. d'Avignon, 74 avenue Louis Pasteur,
84029 Avignon Cedex, France

⁽⁶⁾Dpt Anthropology, Univ. Chicago 1126 E 59th Street Chicago IL 60637, USA

⁽⁷⁾IFCH, Unicamp, CP 6166, 13803-970 Campinas, SP, Brésil

Abstract: Agriculture diversity and heritage in the mid-Rio Negro region (Brazilian Amazon). Local norms and practices about perception and management of agrobiodiversity are seen as a biological cum cultural heritage. We develop a methodological approach for interdisciplinary study of agrobiodiversity in multiple scales (domestic unit, village, region). We offer some results of its application on the mid-Rio Negro region, in a pluriethnic context. The agricultural system is based on slash and burn practices in a low density area. Three study areas were chosen : two of them are forest villages and one is an urban area (7,000 inhabitants). Agricultural diversity was studied with 4, 9 and 5 families respectively. The results show a broad agrobiodiversity (243 cultivated plants and 73 varieties of manioc for 18 families). Under current conditions, this diversity is maintained in the urban area as well. We stress the different plant collections managed by each woman agriculturalist. Such high diversity is linked to an extremely active and wide circulation of plants between spaces and people at different scales. Resources were socially and geographically mapped, and they showed a flux that extends 500 km up as well as 500 km down river. However, most of these exchanges take place in the village and its vicinity and are associated with kinship. In town, the flux is determined by neighbourhood and economic factors rather than by family links. In both cases, plant circulation takes place outside of agricultural development institutions, markets and economic relations. They belong to a larger system of exchange in which phylogenetic resources are not individually owned. Furthermore, agricultural diversity is related to a very rich handicraft culture and to a complex food system. Finally, we argue that legal tools currently being built in Brazil around phylogenetic resources are not conducive to the preservation and valorisation of an

* Correspondance et tirés à part : laure.emperaire@ird.fr

agricultural system which produces diversity such as the one we describe in the mid-Rio Negro region. As things stand now, registering that system as cultural heritage could provide a better acknowledgement of its worth, give it visibility and ensure its conservation. How can they be articulated with Brazilian recent juridical norms on biodiversity and traditional knowledge?

agrobiodiversity/ Amazonia/ networks/ legislation/ cultural heritage

Résumé : Nous mettons en regard la construction patrimoniale de l'agrobiodiversité avec les différents instruments de conservation discutés aujourd'hui au Brésil. La question posée est d'identifier comment les normes et pratiques locales de perception et gestion de l'agrobiodiversité s'articulent aux instruments juridiques de conservation et valorisation de la biodiversité et savoirs associés récemment mis en place au Brésil. L'article comprend deux volets, une proposition méthodologique pour une approche interdisciplinaire de l'agrobiodiversité et l'application de cette approche à la région du moyen Rio Negro. Les résultats portent sur l'amplitude de la diversité agricole, le rôle fondamental de la circulation des variétés et espèces cultivées dans le maintien d'une diversité élevée, sa dimension culturelle et régionale.

agrobiodiversité/ Amazonie/ réseaux/ législation/ patrimoine culturel

1. INTRODUCTION

Des traités internationaux aux certifications, un faisceau d'instruments juridiques et économiques se met en place en vue d'assurer des droits de propriété intellectuelle sur le vivant, notamment les ressources phytogénétiques, et de contribuer à la conservation de ces ressources via leur valorisation économique. Ces instruments procurent un nouvel éclairage aux modes classiques de conservation des plantes, *ex situ*, *on farm* ou *in situ*. En même temps, la diversité agricole, toujours produite dans un contexte socioculturel et écologique particulier, est en passe de s'articuler avec un système dont les rouages sont en grande part inconnus des populations concernées. De manière symétrique, les normes et les formes de gestion locales des ressources phytogénétiques sont *terra incognita* pour les acteurs institutionnels de la conservation. La question posée est donc celle de la construction d'une interface susceptible d'assurer la pérennité d'un patrimoine agricole, c'est-à-dire d'une diversité agricole porteuse de sens et de références. Nous proposons ici une approche pluridisciplinaire qui permet d'analyser les processus associés à la production et à la conservation locales de la diversité agricole à partir de l'étude de la gestion de cette diversité dans la région du moyen Rio Negro.

2. UN NOUVEAU CONTEXTE LÉGAL POUR LA CONSERVATION DE L'AGROBIODIVERSITÉ ?

La conservation de l'agrobiodiversité s'inscrit aujourd'hui au Brésil dans une double dynamique, l'une qui tend à renforcer des droits de propriété intellectuelle sur le matériel biologique, l'autre qui relève de politiques de valorisation de la diversité culturelle. Ainsi, en marge des deux principaux instruments de protection de l'agrobiodiversité que sont la loi de protection des cultivars (9.456/97) et celle des semences (10.771/2003), deux instruments sont depuis peu mobilisés pour assurer une protection de l'agrobiodiversité traditionnelle. Il s'agit de la mesure provisoire 2186/16 qui régit l'application au Brésil de la Convention sur la Diversité Biologique et du décret 3551/2000 qui vise à la protection des biens culturels immatériels.

La première régit l'accès à la biodiversité et aux connaissances traditionnelles associées, qu'il s'agisse de recherche scientifique, technologique ou de bioprospection. Elle constitue une avancée dans la construction d'une recherche davantage participative et dans la protection des savoirs associés à la biodiversité [20]. Le Ministère de l'Environnement est chargé de son application. Cette mesure est toutefois aujourd'hui en cours de révision et un nouveau projet de loi, à caractère plus restrictif, est actuellement soumis à une consultation générale [4]. Issu de négociations laborieuses entre différents ministères, il vise sous sa forme actuelle à faciliter l'accès aux ressources phytogénétiques en vue de la création variétale. Sous prétexte de réglementer le traité de la FAO, il instaure de fait une différence de traitement juridique pour la diversité génétique en général et pour l'agrobiodiversité en particulier. Cette dernière devrait relever du Ministère de l'Agriculture et non, comme la première, du Ministère de l'Environnement. Ce nouveau système traduit une vision de la diversité agricole dissociée de la diversité culturelle qui la produit. Une reformulation de la loi de protection des cultivars est également en cours avec pour objectif la consolidation des droits de propriété intellectuelle sur ceux-ci. Si elle est approuvée, elle créera des restrictions à la circulation et à l'échange des variétés et savoirs agricoles, et ira à l'encontre des processus producteurs de diversité tels que nous avons pu les observer.

A l'opposé de cette dynamique restrictive, le décret 3551 permet de mettre l'accent sur l'ancrage culturel de diverses formes d'expression. Il s'appuie sur la Constitution qui inclut, dans le concept de biens culturels, des biens matériels et immatériels (formes d'expression, de création, connaissances, ...) et permet d'enregistrer au patrimoine culturel brésilien des manifestations d'ordres divers, religieux, esthétique, artistique, L'enregistrement d'un bien est de nature déclaratoire et est mis en œuvre par l'Institut du Patrimoine

Historique et Artistique National (Iphan [5]), organisme relevant du Ministère de la Culture. S'il n'assure pas de protection au sens juridique du terme, l'enregistrement d'un bien immatériel au patrimoine de la nation entraîne cependant l'application de mesures de protection et de valorisation définies d'un commun accord entre la population locale, via ses associations représentatives, et l'Iphan. Appliquée à la diversité agricole, cet instrument de patrimonialisation opère selon une logique proche de celle des GIAHS (*Globally Ingenious Agricultural Heritage Systems*) de la FAO. L'existence de cet instrument, qui s'inscrit dans la logique des conventions de l'Unesco de 2003 et de 2005, ouvre de nouvelles perspectives pour penser la conservation et la valorisation de la diversité agricole, en général objet de mesures centrées sur sa dimension productive ou génétique ou sur son intérêt économique.

3. ASPECTS MÉTHODOLOGIQUES : RÉSEAUX SOCIAUX ET FLUX DE PLANTES

Pour aborder la diversité agricole en tant que construction à la fois biologique et culturelle nous avons privilégié une entrée par une analyse systémique en nous intéressant à l'ensemble des plantes cultivées en un lieu par un groupe donné et non à ses seules espèces majeures. Nous entendons par cultivée toute plante, domestiquée, en cours de domestication ou sauvage, qui fait l'objet d'une intentionnalité de présence en un lieu. La méthode, construite avec les apports de plusieurs disciplines scientifiques, part du constat selon lequel les connaissances sur la diversité des plantes cultivées ne peuvent être comprises hors du cadre des dynamiques spatiales et sociales associées et ce, à des échelles d'espace et de temps variées. Dans ce sens, un protocole a été défini afin de (i) relever la diversité agricole cultivée en un lieu (parcours commenté avec les agriculteurs, photographies, relevé des noms, usages et origine des plantes) ; (ii) caractériser les espaces associés (relevés GPS des parcelles cultivées et de leurs accès, report sur les images satellites, cartographie, histoire de la gestion des espaces) ; (iii) relier les histoires de vie des familles ou des individus, leurs principales étapes, et les réseaux de circulation des plantes cultivées, en identifiant en particulier les liens de parenté réelle ou symbolique, et les bases spatiales des trajectoires des individus.

Les flux de germoplasme, modélisés par les prises de décision des agriculteurs, révèlent l'ancrage non seulement social et territorial, à différentes échelles, de la diversité agricole mais aussi intergénérationnel. Les réseaux de circulation de plantes cultivées ont déjà été abordés par différents auteurs qui insistent plus sur les liens entre cette circulation et la diversité génétique ([2], [3], [12], [16], [18], [21]) ou sur l'approvisionnement en germoplasme ([1], [8], [15]) que sur leur dimension socioculturelle. Les

données de base - quelle est l'origine de chaque plante cultivée dans un espace donné, de qui a-t-elle été obtenue et d'où? – ont été traitées sous forme de tableau et formalisées avec le logiciel Pajek [17] permettant une interprétation visuelle des flux.

Le projet de recherche, la divulgation de ses résultats et le contexte légal de l'accès à la biodiversité et savoirs associés ont été discutés avec les communautés concernées. Cette démarche, outre son aspect déontologique, est une obligation légale (MP 2186) pour l'obtention de l'autorisation de recherche du Conseil du Patrimoine Génétique. Nous avons procédé à l'identification des plantes *in situ* et à partir de photographies sans collecte de matériel. L'étude a été menée sur trois terrains dans le cadre du programme de recherche BRG : dans le moyen Xingu auprès d'un village amérindien Kayapó, dans le haut et moyen Rio Negro auprès de populations de différentes ethnies, et dans le haut Juruá, en Amazonie péri-andine, auprès d'agriculteurs descendants de Nordestins. Pour des raisons de place, nous n'exposons dans cet article que les résultats obtenus sur le Rio Negro où l'étude a été menée de manière exhaustive dans deux villages, Tapereira et Espírito Santo avec environ 25 (4 familles) et 50 personnes (9 familles) et à Santa Isabel (7 000 habitants) en peu plus en amont, auprès de 5 familles.

4. LES PARTICULARITÉS DE LA ZONE D'ÉTUDE: LE MOYEN RIO NEGRO

A l'écart des fronts de colonisation, la région du moyen Rio Negro appartient à l'Amazonie des fleuves (fig. 1). De Barcelos au haut Rio Negro, soit sur un parcours de 600 km, l'environnement forestier est omniprésent, seulement interrompu par des villages formés de quelques familles et les trois petites villes de la région, Barcelos, Santa Isabel et São Gabriel. La densité de population (0,2 hab/km²), les voies d'accès limitées au réseau fluvial, l'histoire et la position excentrée de la région par rapport au reste du pays expliquent que l'élevage, l'agrobusiness et les violents conflits fonciers, désormais paradigmatiques de l'Amazonie dite des routes, y soient minimes. On y connaît par contre le même processus d'urbanisation (autour de 40 % des 65 000 habitants vit en ville) associé à une importante mobilité de la population et au maintien de liens complexes entre le rural et l'urbain [11].

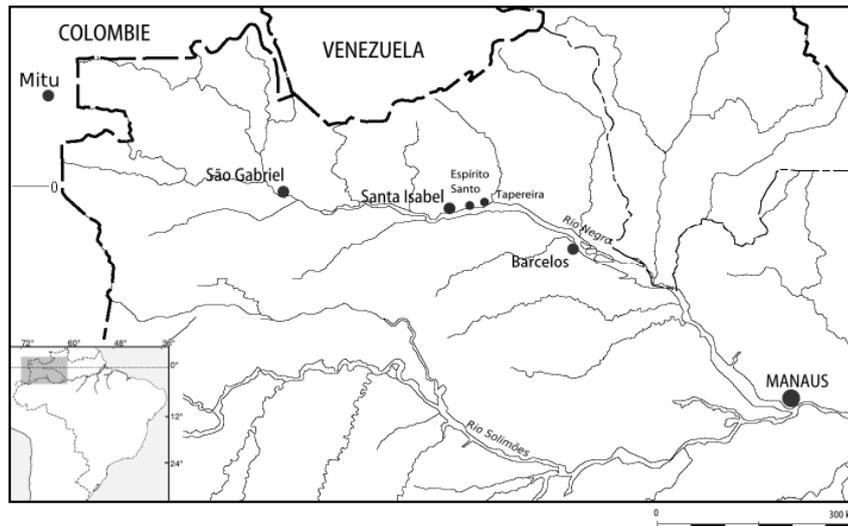


Figure 1 : La région du Rio Negro, localisation des sites d'étude.

L'agriculture sur brûlis est l'activité dominante. Elle est centrée sur le manioc et destinée à la consommation familiale et à la vente de produits à petite échelle. La gestion des abattis, le choix et la connaissance des variétés cultivées relèvent des activités féminines tandis que la préparation de l'abattis est un travail masculin. Une famille déboise ou défriche moins d'un demi-hectare par an pour subvenir à ses besoins et gère deux à trois parcelles à différents stades de production.

La population, majoritairement amérindienne, est multiethnique (familles linguistiques tukano oriental et arawak) et traversée par des revendications identitaires. Dans le haut Rio Negro, en amont de São Gabriel, celles-ci se sont concrétisées depuis une quinzaine d'années par la reconnaissance de Terres Indigènes[†] et la multiplication de projets économiques et culturels menés par les associations locales, relayées ou non par des ONG. En aval, cette dynamique est plus récente avec une demande de reconnaissance de Terres Indigènes en cours.

[†] La Constitution brésilienne de 1988 attribue des droits spécifiques aux peuples autochtones et notamment l'usufruit exclusif et collectif de leurs Terres Indigènes. Les Terres Indigènes, territoires reconnus d'occupation ancestrale, occupent aujourd'hui 20 % de la surface de l'Amazonie légale.

5. RÉSULTATS

5.1. L'amplitude de la diversité

La région du Rio Negro est connue comme étant un centre secondaire de diversité agricole [7] et nos résultats confirment cette importance pour sa conservation. Sur les trois sites, qui concernent au total 18 familles, on a recensé 242 espèces de plantes cultivées (tabl. I) avec 75 variétés nommées et différenciées pour le manioc dont une de manioc doux et une ornementale. Les abattis et le jardin entretenus par une agricultrice et/ou sa famille comportent de 17 à 96 espèces et de 6 à 20 variétés de manioc amer. La diversité infraspécifique est également élevée dans les piments, les ananas, les ignames et les bananiers. De manière générale, une plus large gamme de plantes cultivées va de pair avec une ancienneté de résidence ou/et l'âge de l'agricultrice. Les trois sites, urbain et forestiers, renferment une diversité comparable tant au niveau des espèces qu'à celui des variétés : on recense 35 variétés de manioc amer à Espírito Santo, 28 à Tapereira et 44 à Santa Isabel. Le développement d'une agriculture péri-urbaine à Santa Isabel n'est donc pas aujourd'hui synonyme d'une érosion de la diversité agricole en terme quantitatif global : la perte éventuelle de variétés locales lors des migrations est compensée par l'obtention de nouvelles variétés, locales ou extra-locales.

Tableau I : Nombre d'espèces cultivées dans les trois sites par catégorie d'usage.

Catégorie d'usage	Espírito Santo (N=9)	Tapereira (N=4)	Santa Isabel (N=5)	TOTAL
Alimentaire annuelle	27	26	28	38
Fruitier	39	41	46	61
Médicinale	37	30	75	99
Ornementale	9	3	23	34
Technique	6	2	1	6
Protectrice des abattis	4	1	4	4
TOTAL	122	103	177	242

(N= nombre de familles)

Toutefois, même si la diversité des plantes alimentaires demeure remarquable dans l'agriculture péri-urbaine, l'augmentation de diversité constatée en ville concerne surtout les plantes poussant aux alentours des maisons et correspondant aux catégories d'usage ornementale et médicinale. Les ornementales sont insignifiantes dans les deux villages mais prennent de l'importance dans le creuset qu'est la ville, où se mêlent apports esthétiques, alimentaires et thérapeutiques des groupes amérindiens et ceux issus du modèle occidental d'urbanisation, de la mission salésienne et des pouvoirs publics. La catégorie médicinale compte une vingtaine d'espèces issues de la

pharmacopée généralisée brésilienne qui ne se retrouve pas nécessairement dans les sites forestiers.

Il faut souligner que l'ensemble des plantes entretenues par une agricultrice présente une dimension individuelle et locale forte. Chaque famille, voire chaque femme, constitue et remanie continuellement son propre assemblage. En moyenne, moins d'un tiers des plantes est commun aux différentes familles d'un même site et moins d'un quart en ce qui concerne les maniocs. L'observation est confirmée par le calcul des indices de similitude de Jaccard [9] dont la moyenne est de $0,30 \pm 0,034$ à Espírito Santo, de $0,24 \pm 0,064$ à Tapereira et de $0,24 \pm 0,027$ à Santa Isabel ($p=0,05$). Pour le manioc, on obtient des indices moyens respectivement de $0,27 \pm 0,04$, $0,17 \pm 0,13$ et $0,14 \pm 0,04$. Le caractère singulier des collections se retrouve à l'échelle intersite avec des indices de l'ordre de 0,35 (51 des 242 espèces sont communes aux trois sites), et de 0,26 à 0,27 pour les variétés de manioc (10 variétés communes). La singularité de ces ensembles de plantes cultivées reconnue à l'échelle de l'unité domestique et du site n'est pas signe d'une gestion statique mais résulte de l'intérêt remarquable dont les habitants font preuve pour la circulation et l'échange de plantes. Ainsi, la diversité agricole du Rio Negro ne saurait se résumer à celle d'une unité domestique : chacune d'entre elles offre une version singulière, à la fois même et différente, de la diversité régionale. En somme, les agriculteurs du Rio Negro se comportent comme des collectionneurs qui s'appliquent à constituer, entretenir et gérer une famille originale de plantes. Ceci confirme l'importance d'une perspective régionale pour l'étude de la diversité agricole.

5.2. Flux, réseaux sociaux et espace géographique

On distingue les plantes selon quatre sources d'obtention possibles : celles issues d'une germination spontanée *in loco*, celles issues « des anciens » qui occupaient antérieurement la terre, celles provenant du propre stock de l'agricultrice, et celles - plus nombreuses et seules à faire l'objet d'une circulation au sens géographique du terme - obtenues auprès d'une tierce personne (parente, voisine, patron de l'extractivisme) ou institution urbaine. Ces quatre grandes catégories correspondent respectivement à 5,9 % (70), 1,1 % (13), 13,1 % (156) et 79,9 % (951) des 1 190 faits élémentaires[‡] à l'origine de la présence des plantes cultivées qui ont pu être qualifiés (259 non identifiés sur un total de 1 449). Ces valeurs mettent bien en évidence l'importance de l'échange dans l'obtention de germoplasme. Chaque relation

[‡] Nous avons considéré comme fait élémentaire, la présence d'une espèce ou variété dans un espace particulier (jardin, abattis, ...) provenant d'une même source chez une agricultrice donnée. Ainsi, la présence d'une même espèce fruitière par exemple peut résulter de plusieurs événements si les personnes donatrices sont différentes.

entre agricultrice et donateur est porteuse d'une à 28 plantes cultivées, les valeurs les plus élevées étant essentiellement liées à l'obtention de variétés de manioc (tabl. II).

Tableau II : Importance des différentes sources d'obtention de plantes dans les trois sites étudiés.

Source	Site	Espírito Santo (N=9)	Tapereira (N=4)	Santa Isabel (N=5)	Total
Parenté		267 (55,3 %)	127 (48,4 %)	138 (31,0 %)	532
Voisinage		115 (23,8 %)	56 (21,4 %)	148 (33,3 %)	319
Stock personnel		54 (11,2 %)	28 (10,7 %)	74 (16,6 %)	156
Spontané		19 (3,9 %)	18 (6,9 %)	33 (7,4 %)	70
Institutions urbaines		10 (2,1 %)	3 (1,1 %)	37 (8,3 %)	50
Patron		6 (1,2 %)	29 (11,1 %)	12 (2,7 %)	47
Anciens et autres		12 (2,5 %)	1 (0,4 %)	3 (0,7 %)	16
Total		483	262	445	1190

(N= nombre de familles)

La circulation du germoplasme opère selon deux logiques. La première est constitutive du processus même de l'agriculture sur brûlis qui exige le transfert continu du matériel propagatif des anciens abattis vers les nouveaux, dans un espace restreint. La seconde se fonde sur des réseaux d'échange structurés par les normes sociales et par les histoires de vie des individus inscrites dans des espaces plus vastes. Il importe de souligner que sous sa forme actuelle, la circulation de plantes ne relève pas de relations marchandes et se réalise en marge des institutions de développement agricole. Elle s'intègre dans un système généralisé de flux de ressources phytogénétiques, non appropriables individuellement, dont le fonctionnement est gage de pérennité et de sécurité alimentaire pour la collectivité. Cette circulation repose sur un ethos collectif de solidarité et de maintien d'un patrimoine commun.

L'analyse du réseau d'obtention des plantes souligne son caractère individuel (fig. 2). Chaque agricultrice, soit en ville ou en zone forestière, a son propre réseau de donneurs. Une diversité élevée peut être interprétée comme le signe d'une sociabilité accomplie. De manière générale, les réseaux les plus larges sont ceux des agricultrices les plus anciennement établies. Les échanges font appel à des liens basés sur la parenté (environ 45 % des obtentions), les relations de voisinage (27 %), celles avec des institutions urbaines (préfecture, mission, écoles, marché, commerces ou encore dépotoir, 4 %) ou encore celles entretenues avec les anciens patrons de l'extractivisme (4 %) qui ont un accès privilégié à d'autres sources de plantes (mission, pouvoirs publics).

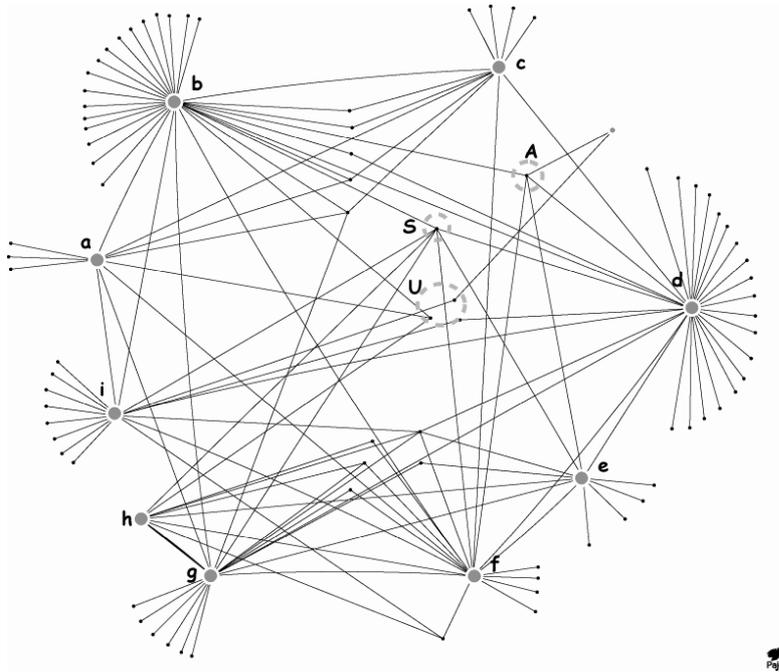


Figure 2 : Exemple de formalisation des sources mises en jeu dans l'obtention des plantes cultivées à Espírito Santo (cercle gris de a à i : informateur-récepteur ; point noir : donneur ; A : anciens ; S : spontané ; U : urbain).

Le développement de l'agriculture autour de la ville s'accompagne de deux modifications majeures. La part du voisinage se renforce au détriment des liens familiaux (tendance à laquelle contribue la configuration de l'habitat urbain) et passe de 23 à 33 %. La transmission d'ascendants à Ego diminue en ville (43 à 28 %) au profit d'une transmission d'enfants à Ego (12 à 38 %). Cette transformation peut être interprétée comme une perte de la dimension patrimoniale attachée à la diversité et préfigure selon toute vraisemblance une évolution régionale. De plus, en ville, les réseaux de circulation des plantes se calquent sur ceux d'obtention d'un espace forestier où ouvrir un abattis. L'obtention de germoplasme devient alors en quelque sorte un produit dérivé de l'accès à la terre. Cette configuration s'oppose à celle relevée en zone forestière où l'accès à la terre est soumis à une simple entente au sein du village.

Les plantes circulent dans un espace régional structuré par le fleuve avec pour points extrêmes Mitu en Colombie et Manaus. De 70 à 90 % des flux se réalisent dans l'espace du village et de ses alentours. Les flux secondaires, d'origine plus lointaine, sont cependant originaux car ils drainent des plantes issues de niches culturelles bien spécifiques. Ainsi la ville de Santa Isabel est

marquée par un apport de plantes du haut Rio Negro (souvent des médicinales) et extra-régional. Des curiosités botaniques, des fruitiers nouveaux ou des ornementales (un pied de vigne, du ramboutan, un grenadier, une nouvelle variété de piment, ...) viennent du Minas Gerais ou des commerces de Manaus.

5.3. Le traitement de la plante

Le statut des plantes cultivées dans l'abattis relève davantage de règles d'usufruit d'un bien collectif que de règles de propriété individuelle. Ce caractère de bien collectif, qui s'applique aussi en zone forestière à la terre, se traduit *a contrario* par le regard défavorable porté sur celle qui esquive l'échange : donner des plantes, en particulier des boutures de manioc, est une obligation à laquelle on ne saurait se soustraire. De même, négliger ce patrimoine en abandonnant par exemple un abattis au feu ou à la friche sans en avoir auparavant soustrait les dernières récoltes et boutures, relève d'un manquement aux règles élémentaires de gestion et de traitement des plantes cultivées. Dans ce cas, ce n'est pas tant l'espace cultivé qui est abandonné mais les plantes présentes, c'est-à-dire des êtres que l'agricultrice dit devoir « élever » au-delà de les avoir plantés. Comme cela a déjà été décrit ailleurs ([10], [14] notamment) et ne peut être développé ici, le mode de traitement des plantes doit être rapporté au mode de traitement d'autrui et du même, les plantes et les hommes partageant finalement des modèles comparables de relations avec les autres êtres. Cette conception, qui place la pratique agricole au sein des pratiques sociales, joue certainement un rôle clé pour la conservation de la diversité dans le Rio Negro.

La notion de bien collectif ne s'étend cependant pas à toutes les plantes. Les médicinales, objets de savoirs spécialisés, et celles qui assurent la protection des abattis relèvent davantage d'une propriété, familiale ou individuelle. De même, chaque arbre fruitier planté autour de la maison appartient à un membre de la famille qui bénéficie de sa production, peut la vendre à titre personnel mais en assume l'entretien. S'il s'absente, il peut céder ce bien à un tiers selon des modalités définies en commun. En tout état de cause, il apparaît que la gestion de la diversité agricole s'encadre dans un système de normes et valeurs davantage centré sur les ressources biologiques que sur les règles du foncier habituellement prises en considération par les organismes de développement.

L'importance des noms ne peut être dissociée de celle des objets biologiques qui les supportent. Ainsi, les noms de manioc sont construits à partir des éléments de l'environnement naturel, végétal et animal : une collection de variétés est aussi une collection de mots qui confirment, en l'élargissant encore, cet intérêt pour la diversité biologique [13].

5.4. Diversité agricole, culture matérielle et système alimentaire

Enfin, les artefacts et objets sont à la charnière des registres agricole et alimentaire. La diversité culinaire observée dans la région est tout autant liée à celle des plantes cultivées qu'à celle des objets utilisés pour leur culture et leur transformation. Il faut une douzaine d'artefacts aux différentes fonctions (râpe, platine à torrifier, hotte de portage, couleuvre à manioc, tamis, éventail à feu, récipients...) pour que le manioc devienne l'aliment de base qui est consommé, comme ailleurs dans l'Amazonie des fleuves, avec du poisson. L'enquête menée à Santa Isabel montre combien les aliments et les boissons issus du manioc sont diversifiés : « farines » de couleur et consistance différentes, galettes aux tailles et compositions variées, bouillies, bières, condiments, etc. Chaque préparation associe généralement plusieurs variétés de manioc, et plusieurs produits différents dérivent de la transformation du tubercule (farine de manioc, fécule, suc, ...). Pareillement, les piments se consomment surtout en mélange que ce soit en poudre ou macérés. Un grand nombre de fruits et de tubercules sont encore consommés tels quels, crus ou cuits, en bouillies, bières ou jus.

Si la recherche de la diversité semble présider aux pratiques agricoles, artisanales ou alimentaires, elle ne saurait être réduite à une simple stratégie face au risque. Rien n'impose à l'agricultrice de mélanger plusieurs variétés différentes au moment de fabriquer sa farine, si ce n'est l'intérêt qu'elle porte à la constitution et à l'entretien d'une collection particulière de plantes.

6. CONCLUSION : DE L'AGROBIODIVERSITÉ AU PATRIMOINE AGRICOLE

Les résultats confirment l'importance de la région du Rio Negro comme centre de diversité. Mais cette importance ne peut être évaluée à la seule aune d'un hotspot. La diversité agricole est étroitement reliée aux autres registres socioculturels et matériels des populations locales. Les travaux de terrain, menés à une échelle détaillée, témoignent aussi de changements dans ces différents registres : baisse de la diversité des manioc[§], changements dans les relations sociales impliquées, abandon d'objets et aliments traditionnels au profit de leur version industrialisée. En revanche, la ville, initialement pensée comme facteur d'érosion, apparaît comme un espace de reconstruction de la diversité. À l'échelle régionale, il n'y a pas dans

§

La comparaison des données sur la diversité des manioc relevés à Tapereira entre 1996 [12] et 2006 montre que celle-ci a toutefois été réduite de moitié (66 variétés cultivées par 5 agricultrices en 1996).

l'immédiat de pertes d'objets biologiques, d'artefacts ou de savoir-faire, celles-ci demeurent ponctuelles. Mais les indicateurs relevés annoncent-ils une rupture qui mènerait vers un modèle généralisé amazonien de diversité agricole ou relèvent-ils d'une adaptation continue à des nouveaux contextes qui préserverait des spécificités locales ? En d'autres termes comment penser le devenir de cette diversité ? Quels intérêts locaux et nationaux mobilise-t-elle ?

L'intérêt local pour la diversité des plantes cultivées est manifeste par l'ampleur des acteurs qu'elle mobilise, par la fierté qu'elle engendre et par la sécurité alimentaire qu'elle offre. Les trois ans de travaux de terrain avec les discussions qu'ils ont générées ont sans aucun doute contribué à donner une nouvelle légitimité aux plantes cultivées et savoir-faire associés. À la suite des travaux sur le manioc, les indications géographiques semblaient une option intéressante pour conserver et valoriser une diversité locale [19]. Aujourd'hui, en approfondissant la gestion de la diversité agricole dans le Rio Negro, elle nous semble difficile à mettre en œuvre par la multiplicité des groupes sociaux qui participent de l'existence de cette diversité, par l'immensité du territoire en jeu et par le risque de spécialisation qu'elle engendre. La valorisation des produits locaux passe davantage par leur insertion sur les marchés locaux ou régionaux et leur juste rémunération que par la mise en place d'une organisation lourde.

En revanche, la conception de l'Iphan ouvre des perspectives pour l'enregistrement d'éléments plus vastes du patrimoine culturel brésilien tel que le système agricole du Rio Negro, caractérisé par une multitude d'éléments interdépendants et dynamiques. Elle permet de s'attacher à conserver des processus plus que des produits [6]. C'est cette possibilité qui a été discutée, et retenue, par l'association locale du moyen Rio Negro (Acimrn**) sensibilisée au devenir du système agricole régional qui a fait la demande d'enregistrement auprès de l'Iphan. L'objectif est de donner à cette diversité une visibilité et une reconnaissance sur le plan régional et national. Le dossier est aujourd'hui en cours de montage. Cet enregistrement, s'il se concrétise, servira de support à d'autres actions traduisibles en termes de reconnaissance d'une identité culturelle, de services environnementaux et de conservation de l'agrobiodiversité, aujourd'hui garantis par l'existence de cette agriculture traditionnelle.

** Association des Communautés Indigènes du Moyen Rio Negro.

REMERCIEMENTS

Les auteurs remercient le BRG pour son soutien financier. Les recherches ont été réalisées dans le cadre du programme PACTA, *Populations locales, Agrobiodiversité et Connaissances Traditionnelles Associées em Amazonie brésilienne*, CNPq - Unicamp / IRD, n° 492693 / 2004-8 avec des financements IRD, CNPq, ANR-Biodivalloc et BRG. Elle a compté dans le moyen Rio Negro avec la participation de l'Association des Communautés Indigènes du Moyen Rio Negro (Acimrn).

Autorisation n° 139 publiée au DOU (04/04/2006) : « Tout usage des informations exposées à des fins de bioprospection ou de développement technologique doit être soumis à l'obtention d'un accord préalable informé et d'un contrat d'utilisation du patrimoine génétique et de répartition des bénéfices établi avec les populations concernées, ainsi qu'à l'autorisation du CGEN. »

Ont participé à la recherche Juventina Oliveira, Maria Oliveira, Jorgina da Silva, Nilza Rosende Lúcio, Edilene Lúcio, Maria Angélica Reis, Neuza da Silva Lúcio, Zulmira Oliveira et Isabel Silva de Espírito Santo, Debania Dias, Conceição Carvalho, Eduarda Murilo, Cleomar Dias Costa de Tapereira et Angelina da Silva Gervásio, Angelina Sousa Dias, Conceição Dias, Hilda Teixeira, Jurema Carvalho et Lucrécia Avelino de Santa Isabel. Nous remercions Pascal Cristofoli (INED/EHESS) pour son initiation à la formalisation des réseaux et Ana Gita de Oliveira de l'Iphan.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Badstue L.B., Bellon M.R., Berthaud J., Ramirez A., Flores D., Juarez X., 2007. - The dynamics of farmer's maize seed supply practices in the central valleys of Oaxaca, Mexico. *World Dev.*, 35: 1579-1593.
- [2] Barnaud, A. 2007.- Savoirs, pratiques et dynamique de la diversité génétique : le sorgho (*Sorghum bicolor* ssp. *bicolor*) chez les Duupa du nord Cameroun. Thèse de Doctorat. Univ. Montpellier II, Montpellier.
- [3] Bazile D., Abrami G., Dembele S., Le Page C., Dionnet M., Coulibaly H., 2006. - Les paysans au coeur de la construction de règles collectives de gestion de l'agrobiodiversité. In : Colloque Gestion concertée des ressources naturelles et de l'environnement, Saint Quentin en Yvelines, 26-28/06/2006. http://www.c3ed.uvsq.fr/cdgecorev/fr/theme_6.html
- [4] Brasil, 2007. – <http://www.mma.gov.br>
- [5] Brasil, 2007. – <http://www.iphan.gov.br>
- [6] Carneiro da Cunha, M. 2005. - Introdução. *Revista do IPHAN*, Brasília, 32:15-27.
- [7] Clement C.R., 1999. - 1492 and the loss of Amazonian crop genetic resources.I. The relation between domestication and population decline. *Econ. Bot.*, 53: 188-202.

- [8] Coomes O.T., Ban, N. 2004. - Cultivated plant species diversity in home gardens of an Amazonian peasant village in northeastern Peru. *Economic Botany*, 58: 420-434.
- [9] Daget P., Godron M., 1982. - Analyse de l'écologie des espèces dans les communautés, Masson, Paris.
- [10] Descola P. 1986. - La nature domestique. Symbolisme et praxis dans l'écologie des Achuar. Maison des Sciences de l'Homme, Paris.
- [11] Eloy L., 2005. - Entre ville et forêt : le futur de l'agriculture amérindienne en question - Transformations agraires en périphérie de São Gabriel da Cachoeira, nord-ouest amazonien, Brésil. Thèse de doctorat, Univ. Paris 3, Paris.
- [12] Emperaire L., Pinton F., Second G., 1998. - Gestion dynamique de la diversité variétale du manioc (*Manihot esculenta*) en Amazonie du Nord-Ouest. *Nat. Sci. Soc.*, 6: 27-42.
- [13] Emperaire L. 2005. - A biodiversidade agrícola na Amazônia brasileira: recurso e patrimônio. *Revista do IPHAN, Brasília*, 32: 23-35.
- [14] Haudricourt A. 1962. - Domestication des animaux, culture des plantes et traitement d'autrui. *L'Homme* 2: 40-50.
- [15] Lerch N. 1999. - Biodiversity in homegardens. Thèse de Doctorat, McGill University, Montréal.
- [16] Louette D., A. Charrier A., Berthaud J., 1997. - *In situ* conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. *Econ. Bot.* 51: 20-38.
- [17] Pajek, <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
- [18] Nagarajan L., Smalea M., Glewweb P., 2007. - Determinants of millet diversity at the household-farm and village-community levels in the drylands of India: the role of local seed systems. *Agric. Econ.* 36: 157-167.
- [19] Pinton F., Emperaire L., 2001. - Le manioc en Amazonie brésilienne: diversité variétale et marché. *Genet. Sel. Evol.*, 33: 491-512.
- [20] Santilli J., 2006 - Intangible cultural heritage and collective intellectual rights. In: Mathias F., Novion, H. de (eds), *The crossroads of modernities: debates on biodiversity, technoscience and culture*. ISA, São Paulo, pp. 81-98.
- [21] Subedi A., Chaudhary P., Baniya B.K., Rana R.B., Tiwari R.K., Rijal D.K., Sthapit B.R., Jarvis D.I., 2003. - Who maintains crop genetic diversity and how?: Implications for on-farm conservation and utilization. *Cult. Agric.* 25: 41-50.

Détection par EcoTILLING du polymorphisme nucléotidique de gènes de résistance chez le peuplier

Patricia FAIVRE-RAMPANT^{(1)*}, Aloïs BRESSON⁽¹⁾, Isabella PAOLUCCI⁽¹⁾,
Erwan CAILLEUX⁽¹⁾, Marion DALMAIS⁽¹⁾, Florence PIRON⁽¹⁾,
Arnaud DOWKIW⁽²⁾, Vanina GUERIN⁽²⁾, Abdelhafid BENDHAMANE⁽¹⁾,
Catherine BASTIEN⁽²⁾, Véronique JORGE⁽²⁾

⁽¹⁾INRA/UMRGV, 2 rue Gaston Crémieux, 91057 Evry cedex, France

⁽²⁾INRA UAGPF, BP 20619 Ardon, 45166 Olivet cedex, France

Abstract: EcoTILLING for identification of allelic variation in poplar resistance genes. For this pilot project, we screened 92 *P. trichocarpa* individuals for both detection of SNPs in heterozygous and homozygous variants. The individuals come from different populations of natural areas. Taking advantage of the localization of rust resistance genes on genetic map and the alignment onto the available genome sequence, 14 positional candidate genes for rust resistance were selected to be analyzed through EcoTILLING technique for nucleotide polymorphism. This method has been developed from TILLING method (Targeting Induced Local Lesions in Genomes) which was set up in *Arabidopsis* for analyzing natural variability in plants. EcoTILLING has been used and set up in *P. trichocarpa*, species used as parent in inter-specific crosses to obtain high-productive clones for wood production. Due to its origin as method for variation detection in homozygous species, it was adapted to heterozygous species with some modifications. Each experimental DNA is analyzed two times: singly and mixed with an homozygous reference DNA. After heating to separate DNA strands and slow decreasing of temperature to form heteroduplex molecules between alleles differing in SNP and in/del positions, experimental samples are partially digested with enzyme *EndoI* that cut DNAs in mismatch points. Digested samples are then separated in an acrylamide gel. The analysis of single samples permit to discover mutations for which analyzed individuals are heterozygous. In this case, *EndoI* enzyme recognizes mismatches between alleles within one individual. While when experimental DNAs mixed with reference DNA are digested, mutation points can be due to difference between the two alleles of one individual or to a difference between the DNA of analyzed individual and the reference DNA. In this second case all mutation points should be detected, while in the first one only mutations for which samples are heterozygous are detected. In this study a single allele from a BAC library of a *P. trichocarpa* individual has been used as reference. Our results show that EcoTILLING method is suitable for SNPs detection and genotyping of heterozygous individuals. This is possible also for high polymorphic genes such as resistance genes.

* Correspondance et tirés à part : faivre@evry.inra.fr

Nevertheless for routine applications, capillary systems for fragments separation could speed up the process and permit also a faster lecture of data.

EcoTILLING/ Poplar/ heterozygous/ resistance genes/ allelic variant

Résumé : La technologie EcoTILLING est testée comme méthode de génotypage de moyen à haut débit pour sa capacité à détecter le polymorphisme nucléotidique des gènes chez une espèce hétérozygote. Cette méthode est développée chez le peuplier baumier, espèce américaine introduite dans les schémas de création variétale en Europe. Les gènes choisis sont des gènes candidats positionnels, contrôlant la résistance à la rouille foliaire. La démarche développée, d'abord chez les plantes autogames fortement homozygotes, a ensuite été adaptée aux contraintes imposées par l'hétérozygotie, chez les plantes allogames. Deux digestions partielles des hétéroduplexes avec l'enzyme *EndoI* sont conduites parallèlement, l'une avec un ADN dit de référence, l'autre sans cet ADN de référence. La première permet d'identifier les points de mutation, la seconde permet de révéler le polymorphisme entre allèles d'un même individu. L'ADN de référence correspond à un seul allèle issu d'une banque BAC d'un individu *P. trichocarpa*. Nos résultats montrent que l'EcoTILLING permet la détection de SNP ainsi que le génotypage d'individus hétérozygotes, même pour des gènes cibles très polymorphes. Cependant, pour une utilisation en routine, il serait nécessaire de séparer les produits de digestion par électrophorèse capillaire afin d'optimiser la capture des SNP.

EcoTILLING/ peuplier/ hétérozygote/ gènes de résistance/ polymorphisme allélique

1. INTRODUCTION

Le développement de la génomique structurale et fonctionnelle donne accès à un nombre important de gènes impliqués dans l'adaptation des organismes à leur environnement. Avec l'aide des nouvelles technologies de génotypage moyen-débit nous sommes maintenant à même de procéder à la détection des variations alléliques des ressources génétiques. Dans ce travail nous proposons de tester l'une de ces technologies appelée EcoTILLING. L'EcoTILLING est une méthode de détection de polymorphisme simple, SNP et courtes insertion/délétion, et de génotypage moyen-débit. La technique dérive de celle du TILLING développée pour caractériser des mutations induites chimiquement, dans la séquence de gènes cibles [2], [4], [20], [21]. L'EcoTILLING consiste en la recherche systématique de polymorphisme au sein de populations naturelles dans les gènes cibles [3]. Chez les plantes allogames, deux expériences sont réalisées en parallèle, l'une correspond à l'amplification d'un mélange d'ADN, constitué d'une séquence de référence et d'un ADN à géotyper, la seconde correspond à l'amplification de l'ADN d'un seul individu. La première permet de localiser

les points de mutation qui peuvent exister entre la séquence de l'individu et la séquence de référence, la seconde permet d'identifier si l'individu est homozygote ou hétérozygote aux points de polymorphismes révélés en présence de la référence. De par ses caractéristiques l'EcoTILLING doit se révéler très performante pour la détection directe du polymorphisme nucléotidique après PCR chez les organismes allogames. Cette méthode a été utilisée avec succès chez le peuplier [7]. Les auteurs ont montré ainsi que l'EcoTILLING était adapté à la détection d'un fort taux d'hétérozygotie.

L'intégrité des peupleraies européennes est menacée depuis plusieurs dizaines d'années par le champignon *Melampsora larici populina* (*Mlp*), agent causal de la rouille foliaire. Depuis le début des années 80, on assiste aux contournements des résistances successivement sélectionnées [18]. Les premières expérimentations ont montré un contrôle génétique de la résistance totale et de la résistance partielle observées [12]. Depuis lors, des efforts particuliers ont été faits pour identifier et localiser des gènes de résistance durable et comprendre les mécanismes moléculaires sous-jacents. Un gène de résistance qualitative R1 contourné par la virulence 1, hérité de *P. deltooides* a été localisé sur le chromosome XIX. Ce locus est très étroitement lié à un QTL de résistance partielle. Un gène majeur (R_{US}), a ensuite été localisé sur la carte génétique de *P. trichocarpa*, il explique de 35 à 63 % de la variation associée à la résistance partielle à plusieurs isolats de *Mlp* mesurée à la fois par la variation de sporulation du champignon en laboratoire et par le niveau de sensibilité exprimé en pépinière [5], [6], [10]. Depuis R_{US} a été cartographié plus finement, et, nous avons pu assigner R_{US} au chromosome XIX de *P. trichocarpa*. La séquence cible révélée après alignement de la carte génétique fine et de la séquence du génome (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>, [23]) montre une composition riche en gènes de résistance arrangés en super-cluster [11]. L'exploitation de la synténie entre *P. deltooides* et *P. trichocarpa* montre que R1 et R_{US} sont localisés dans la même région du chromosome XIX. Un autre locus de résistance (*Mer*), hérité de *P. deltooides*, contourné par la virulence 7, a été cartographié également dans la même région génomique [1], [13], [18], [24]. Cette région chromosomique comporte donc, les gènes les plus importants responsables de la variation de l'expression de la résistance du peuplier vis-à-vis de la rouille foliaire. Nous avons maintenant pour objectif de trouver dans cette région chromosomique les variants moléculaires influant sur la résistance à la rouille foliaire par une approche dite de gènes candidats positionnels. Les gènes sont choisis d'après leur position dans le génome, dans notre démarche, les gènes sélectionnés se trouvent dans l'intervalle des marqueurs génétiques encadrant R_{US} .

2. MATÉRIEL ET MÉTHODES

2.1. Matériel

2.1.1. Collection de *P. trichocarpa*

Quatre vingt douze génotypes différents appartenant à l'espèce *P. trichocarpa* sont utilisés dans le cadre de cette étude. Les individus sont issus de différentes provenances naturelles situées dans différents bassins fluviaux du nord-ouest des Etats-Unis (états de Washington et de l'Oregon) et du Canada (Colombie Britannique), partie centrale de l'aire naturelle de cette espèce (tabl. I).

Tableau I : Liste et coordonnées géographiques des individus *P. trichocarpa* utilisés dans cette étude. Les nombres entre parenthèses sont les nombres d'individus testés par population.

Provenances	États	Coordonnées des provenances		
		Latitude	Longitude	Altitude (m)
003 (1)	Oregon	45°20'N	121°40'O	1000
010 (25)	Oregon	45°45'N	122°50'O	15
19 (2)	Oregon	44°25'N	123°10'O	180
036 (1)	Oregon	45°07'N	122°50'O	320
037 (20)	Oregon	45°07'N	122°50'O	320
101 (1)	Washington	47°00'N	122°45'O	30
112 (27)	Washington	48°30'N	121°36'O	90
113 (5)	Washington	48°29'N	122°15'O	30
202 (5)	Colombie Britannique	48°46'N	124°44'O	260
220 (5)	Colombie Britannique	49°05'N	123°50'O	150

L'ADN génomique a été extrait à partir de jeunes feuilles, à l'aide du kit DNeasy 96 de chez Qiagen (Hilden, Allemagne) en utilisant le protocole du fournisseur.

2.1.2. Banque BAC d'un individu *P. trichocarpa*

La banque BAC a été constituée à partir du génotype *P. trichocarpa* 101-74, parent mâle d'une famille modèle *P. deltoides* x *P. trichocarpa* utilisée en cartographie génétique pour la détection de QTL. Ce clone a été choisi car il est porteur de l'allèle R_{US}. La banque est constituée de 52 224 clones répartis dans 136 plaques de 384 puits. Les inserts issus d'une digestion partielle avec l'enzyme *Hind*III ont une taille moyenne de 150 kb. Les contaminations par l'ADN chloroplastique et par l'ADN mitochondrial sont estimées respectivement à 3 % et 0,3 %. La représentativité de cette banque est de 15 fois le génome du peuplier. Pour optimiser le criblage, la banque est organisée en pools plaque (1 pool plaque contient les 384 clones d'une plaque) et en pools lignes (1 pool ligne contient les 24 clones BAC d'une

ligne d'une plaque). La banque est donc organisée en 136 pools plaques et 3264 pools lignes.

2.2. Méthodes

2.2.1. Conception des amorces

Les gènes candidats utilisés pour cette étude sont des gènes candidats positionnels identifiés après alignement des cartes génétiques sur la séquence génomique (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>). Ce sont des gènes de résistance, groupés en clusters localisés à une extrémité du chromosome XIX et sur le scaffold 117 (tabl. IIA). Pour l'EcoTILLING les amorces doivent être spécifiques d'un locus et conduire à l'amplification d'un fragment n'excédant pas 1000 bases. Du fait de la redondance des gènes de résistance, les risques d'amplification multilocus sont importants. C'est pourquoi nous avons développé une stratégie fondée sur l'utilisation d'amorces emboîtées pour amplifier spécifiquement chaque locus. Nous avons généré deux couples d'amorces. Un couple (1) spécifique de locus et un deuxième couple (2) permettant l'amplification d'un fragment de taille compatible avec l'EcoTILLING. Les amorces (1) ont été conçues avec le logiciel Primer 3 (v. 0.4.0) [19], d'après la séquence du génome et après ré-annotation des gènes, avec les critères suivants : température d'hybridation comprise entre 55 et 60°C et une longueur des amorces comprise entre 19 et 30 nucléotides. La spécificité des locus amplifiés a été vérifiée par ePCR (PCR *in silico*, [15]) et par alignements Blast. Les tests d'amplification sont réalisés sur l'ADN génomique du clone 101-74. Les différentes formes alléliques sont ensuite isolées de la banque BAC (10 équivalents génome) par PCR afin d'identifier la séquence d'ADN qui sera utilisée comme ADN référence. L'ensemble de ces amplicons a été séquencé pour permettre le dessin des amorces (2). Elles ont été dessinées à l'aide du logiciel Primer 3, avec une température d'hybridation comprise entre 58 et 60°C et une longueur de 20 à 24 nucléotides (tabl. IIB). L'amorce sens est marquée à l'aide du fluorochrome IRDye 700 et l'amorce anti-sens avec le fluorochrome IRDye 800 (MWG-Biotech, Ebersberg, Allemagne).

Tableau IIA : Liste des gènes candidats, leur position sur le génome séquencé (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>) et nom des couples d'amorces correspondants.

Modèles de gènes	Localisation	Amorces
<i>Poptr</i> NBS-LRR80	LG_XIX: 806257	NBS80 BED80
<i>Poptr</i> NBS-LRR83	LG_XIX: 837344	BED83 NBS83
<i>Poptr</i> PK87	LG_XIX: 873510	gV87a
<i>Poptr</i> NBS-LRR90	LG_XIX: 899682	BED90 LRR90
<i>Poptr</i> TIR-NBS-LRR92	LG_XIX: 927237	TIR92
<i>Poptr</i> PK98	LG_XIX: 979779	gV98a
<i>Poptr</i> NBS-LRR102	LG_XIX: 1028156	LRR102 NBS102
<i>Poptr</i> TIR-NBS-LRR107	LG_XIX: 1073646	TIR107
<i>Poptr</i> TIR-NBS-LRR109	LG_XIX: 1095002	TNL109
<i>Poptr</i> TIR-NBS-LRR115	LG_XIX: 1147114	LRR115
<i>Poptr</i> TIR-NBS-LRR65	Sca_117: 654568	NBS65
<i>Poptr</i> TIR-NBS-LRR67	Sca_117: 673368	NBS67
<i>Poptr</i> TIR-NBS-LRR72	Sca_117: 725442	NBS72
<i>Poptr</i> TIR-NBS-LRR73	Sca_117: 734004	NBS73

Liste des amorces définies dans le cadre de cette étude et taille des produits d'amplification

pb1 : taille des produits d'amplification issus de la PCR1 exprimée en paire de bases

pb2 : taille des produits d'amplification issus de la PCR2 exprimée en paire de bases

F : amorce sens

R : amorce anti-sens

Tableau IIB : Séquence des couples d'amorces (1) et (2) et taille des amplicons.

Amorces	Séquences des couples 1	pb_1	Séquences des couples 2	pb_2
NBS80	F-GGGTCAAGAATTTAAAGTCATTAAGG R-GGACACCTTGGTGAATGTCC	1128	F-GAAAGCATATGTTCTTCGTTAATGG R-GTCCGGAAGGAATTTCTTAATTT	1087
BED80	F-GTTAGGATCGTCAGTTTTTAGTTTCC R-GCAGCAATAAGCTTCTGCAA	1093		
BED83	F-TTTATACTTTAGGGCCTCGTTACTG R-AGTTTCTCCATGAACAAAGTCCA	1069	F-TACTTTAGGGCCTCGTTACTGAAC R-GCTTCTCTCTTAGAAAGTGGCTTC	1041
NBS83	F-TCATTTTAGATGATTTGTGGGACT R-TCTCTCTAATGCAAAGACTTG	1200		
gV87a	F-GGATGTTGCCACCTTCCTC R-CAGGACTTCCAAGAAAACAAGG	1071	F-ACCTTCTCCCTAGCGATAATTT R-GTCCCTGGCTTAATTTTGAGAACIT	810
BED90	F-TGGAGTTAGGATCATTTTCTATAGTTCC R-AACAGTGATGCTAGTCACCAAATC	641	F-TTTTATTTTCAAGTTTCAACCTTCC R-AAAGATCTTCTCGTTGTCTGTG	496
LRR90	F-TGATTTGCGATTTCTCTCAAGTAAT R-ACGATGCTTACCCATCTTATCC			
TIR92	F-TCCCATCAGAATTTACATGTATTG R-GGTACTTATGAAGTCACTAATAGTATGGAC	856	F-CTGCTCATAAACAAGCAAGTCT R-CTTGCTCCCTTGCTAATTTAATGTT	672
gV98a	F-TGAGTGACAAAGAGATGGAGAA R-TGGACGACAAAACAACCTTACC	1189	F-CTTGTGAAATACCAAGAAGGAAGA R-CGACAAAACAACCTTACCTTGC	895
LRR102	F-GCATGGACGTAATTACTGATG R-TGCCTATTAATTTCAACCAAAG			
NBS102	F-GAATAAATGAGCTAAATCTCTAAAATAAC R-GTTCCTTACCATGTGGTTGATG	1218	F-TCAGATGATGTTAGCATGGTAGG R-ATCCCAGATTTCTGTGTGTTTT	942
TIR107	F-TTCTAGTTATATTAGATAGTTAGAGGTGT R-TTGTATTATGTTGATATGTGCAG	-		
TNL109	R-ATTCATAAATGAGCATGGATAC F-AAATGGGTTACAACAATAGTTAGA R-TTGTATTATGTTGATATGTGCAG R-ATTCATAAATGAGCATGGATAC	-		
LRR115	F-GAAGTGATGAATCTAATTTGCCTAT R-GGGCTGTTTCTCGACCC	1139		
NBS65	F-GGATGGGTAAGCACTTGACTTAAT R-AAGGCATGCCAACTAAAAGAGC	828	F-GACAATTAATTTGCAGGCATGAAG R-AAGGCATGCCAACTAAAAGAGC	618
NBS67	F-TTCTCACATGATTCACAGACAACA R-AATCTTCTCAGTTTTTCCAATGC	1245	F-GACAATTAATTTGCAGGCATGAAG R-AAGGCATGCCAACTAAAAGAGC	618
NBS72	F-TATGGTAGGAATCTGTGGACTTG R-AAGCGAAGTTCTAATTTGTAGGC	1016	F-CAACTGTTGCCAAAAGCTAATTAC R-AACAATATCCTCATGAAACCACA	884
NBS73	F-ATCAAAATGATGTTTGCATGGTAG R-CACCATATTTTCGGTCAATACAT	964	F-CAACTGTTGCCAAAAGCTAATTAC R-AACAATATCCTCATGAAACCACA	887

Liste des amorces définies dans le cadre de cette étude et taille des produits d'amplification

pb1 : taille des produits d'amplification issus de la PCR1 exprimée en paire de bases

pb2 : taille des produits d'amplification issus de la PCR2 exprimée en paire de bases

F : amorce sens

R : amorce anti-sens

2.2.2. EcoTILLING

L'ensemble de la démarche appliquée dans notre étude est résumé dans la figure 1, la démarche est adaptée de celle décrite par Nieto *et al*, [16]. Les locus d'intérêt (tabl. II) sont tout d'abord amplifiés à l'aide des couples d'amorces (1) en utilisant comme matrice d'une part l'ADN génomique des 92 individus et d'autre part, l'ADN d'un pool plaque porteur de l'allèle

préalablement identifié, choisi comme référence. Après une étape de dénaturation de l'ADN matrice (20 ng) à 94 °C pendant 4 minutes, trente cinq cycles d'amplification sont réalisés dans les conditions suivantes : 94 °C 30 secondes, 55-60 °C 30 secondes, 72 °C 40 secondes et pour finir 72 °C 7 minutes. Les produits d'amplification sont vérifiés sur gel d'agarose puis dilués au centième pour être utilisés comme matrice pour la deuxième PCR, réalisée en présence du couple d'amorces (2) marqué. Les produits de cette seconde amplification sont vérifiés et quantifiés sur gel d'agarose. Deux expériences sont ensuite menées en parallèle pour détecter le polymorphisme nucléotidique :

- La première est effectuée avec le produit PCR de chaque individu mélangé à l'amplicon de référence. Le mélange est en proportion 1/1.
- La seconde est effectuée avec le produit PCR de chaque individu seul.

Les ADN sont dénaturés à 94 °C, pendant 10 minutes, puis renaturés en abaissant la température de 0,1 °C par seconde jusqu'à 8 °C, pour favoriser la formation des hétéroduplexes.

Une digestion partielle des hétéroduplexes est effectuée avec l'endonucléase ENDO1 spécifique de l'ADN simple brin. Cette enzyme, clonée à l'URGV, est un homologue de l'enzyme Cel1. Elle a été isolée à partir d'*Arabidopsis thaliana* et produite par transgénèse chez le tabac (Bendahmane, brevet PCT/EP2004/009166, [22]). La digestion est effectuée dans un volume de 30 µl contenant 6 µl d'ADN, 3 µl de tampon ENDO-1 10X (0,1M HEPES, 0,1 M MgSO₄, 1 % Triton X-100, 0,2M KCl) et l'enzyme ENDO1(3U). La réaction est effectuée à 45 °C pendant 20 mn. La réaction est arrêtée par l'ajout de 5 µl d'EDTA 0,15M pH8,0. Les produits de digestion sont ensuite purifiés à travers une colonne de Sephadex G50 (S-G50 ; GE Healthcare Life Sciences, Little Chalfont, UK). Les échantillons sont concentrés au Speed Vac pendant 1 heure après ajout de 5 µl de tampon de charge/ formamide loading dye (GE Healthcare Life sciences). La migration des échantillons est effectuée parallèlement à un marqueur de taille (Li-COR size standard 50-700 pb), sur gel d'acrylamide 6,5 % de séquence Li-COR (DNA LI-COR 4300 ; LI-COR Biosciences, Lincoln, Nebraska, USA) de 96 puits 25 cm X 25 cm (1500V, 40mA, 40W, 45 °C). Le temps de migration dépend de la taille du fragment d'ADN (4h pour 1000 pb). Les produits de digestion sont visualisés sous Adobe Photoshop et la lecture est réalisée manuellement. La digestion partielle génère des coupures à chaque point de mutation. Chaque coupure libère 2 fragments, l'un marqué avec le fluorochrome IRDye 700 et l'autre avec le fluorochrome IRDye 800. La somme des tailles de ces deux fragments doit

correspondre à la longueur totale du fragment amplifié avec les amorces (2). Un individu hétérozygote au point de mutation est caractérisé par un profil de digestion identique entre les 2 expériences. Un individu homozygote possédant un allèle différent de la séquence de référence présente un profil de digestion lors de la première expérience (en présence de la séquence de référence) et une absence de profil de digestion dans la deuxième expérience. Un individu homozygote avec un allèle identique à la séquence de référence est identifié par l'absence de produits de digestion dans les deux expériences. Les lectures des expériences 1 et 2 sont réalisées en parallèle à partir des profils de digestion marqués au fluorochrome IRDye 700, puis vérifiées par la lecture des profils marqués avec le fluorochrome IRDye 800. Seuls les profils de digestion pouvant être lus sans ambiguïté sont retenus.

3. RÉSULTATS

3.1. Conception des amorces

Les amorces définies selon le protocole proposé dans le paragraphe Méthodes sont répertoriées dans le tableau II. Les amorces TIR 107 et TNL 109 ne conduisent à aucune amplification. Les amorces LRR 115 se révèlent être non spécifiques de locus. Les autres couples d'amorces amplifient un seul locus. Les couples BED 90, LRR 102 amplifient un seul des deux allèles du clone 101-74. Le couple BED 80 amplifie deux allèles se distinguant par un polymorphisme de longueur. Le couple LRR90 amplifie un fragment d'ADN présentant une séquence constituée de motifs LRR répétés (Leucine-Rich Repeat), impossible à analyser pour générer des amorces (2). Aucun produit d'amplification n'a pu être obtenu avec les amorces (2), NBS80. Sur 20 couples d'amorces (1) dessinés, au niveau de 14 gènes candidats positionnels, 9 seront finalement utiles en test pour l'EcoTILLING.

3.2. Analyse du polymorphisme des gènes d'intérêts

La technique EcoTILLING a été appliquée à un échantillon de 92 individus *P. trichocarpa*. Après sélection des loci, nous avons obtenu pour chacun des 9 couples d'amorces (2) et pour l'ensemble des individus un produit d'amplification représentant un locus unique.

Figure 1 A : Dessin des amorces

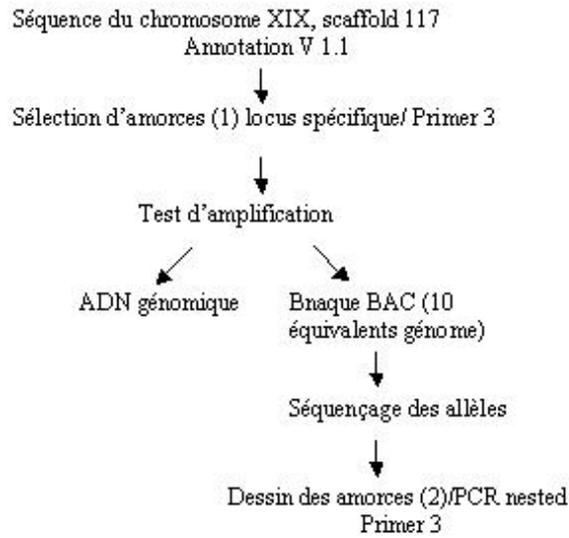


Figure 1B : EcoTILLING

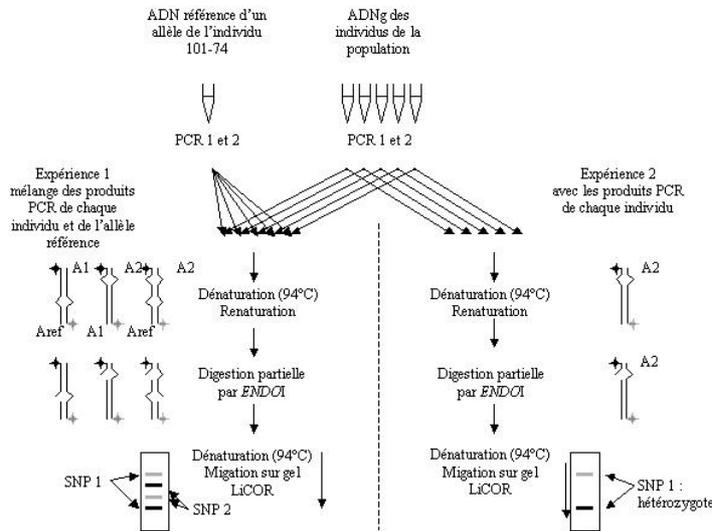


Figure1: Démarche globale de génotypage par EcoTILLING développée dans cette étude.

L'analyse des produits de digestion issus du mélange de l'ADN à génotyper et de l'ADN de référence permet de détecter et de localiser des

points de mutation comme il est schématisé dans l'expérience 1 de la figure 1. L'observation de produits de digestion issus de l'ADN d'un individu seul révèle que cet individu est hétérozygote ou non au locus étudié (expérience 2, fig. 1B).

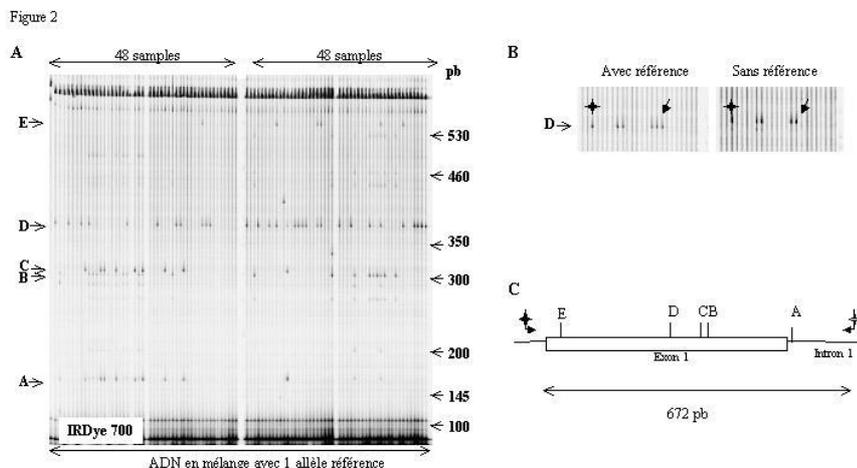


Figure 2 : Exemple de gels EcoTILLING obtenus avec le couple d'amorces TIR92. A : images obtenues après EcoTILLING sur les ADN en mélange avec l'allèle de référence. pb : échelle de poids moléculaire A, B, C, D, E représentent les points de mutation B : comparaison des profils obtenus après EcoTILLING sur les ADN en mélange avec une référence et sur les ADN des individus seuls. Chaque piste correspond à la migration d'un seul individu L'individu signalé par une étoile est hétérozygote, un des allèles est identique à l'allèle de référence l'autre possède un SNP noté D. L'individu signalé par une flèche est homozygote au locus TIR92 possédant le point de mutation D. C : distribution des SNP le long de la séquence amplifiée par les amorces TIR92

La figure 2 représente une illustration d'une portion des gels EcoTILLING obtenus pour l'étude du polymorphisme du gène *PoptrTIR-NBS-LRR92* avec le couple d'amorces TIR92. Les individus homozygotes dont les allèles sont identiques à l'allèle de référence sont caractérisés par une absence de produit de digestion (fig. 2A et 2B). Les individus homozygotes avec un allèle qui diffère de celui de la référence par 1 polymorphisme, présentent une bande dans la première expérience et aucune dans la deuxième (fig. 2B). La somme des tailles des fragments visualisés doit correspondre à la taille du fragment amplifié complet. Les individus hétérozygotes sont identifiés par la présence de produits de digestion dans les deux expériences. La comparaison avec une échelle de

taille permet de localiser les points de mutation (fig. 2C). Nous disposons de la séquence des 2 allèles de l'individu 101-74, nous avons comparé la présence de SNP et les profils de digestion obtenus : Pour les gènes *PopttrTIR-NBS-LRR92*, *PopttrTIR-NBS-LRR72*, *PopttrTIR-NBS-LRR65*, *PopttrPK87* les profils de digestion sont conformes à ceux attendus. Pour *PopttrTIR-NBS-LRR67*, deux SNP séparés par 1 base ne sont pas lisibles. Pour le gène *PopttrTIR-NBS-LRR73*, nous avons observé un artefact. Enfin pour les gènes *Popttr-NBS-LRR83* et *PopttrPK98*, les SNP ne sont pas tous détectés dans nos conditions d'expérience. Le nombre de SNP pour chaque gène ainsi que le nombre estimé d'haplotypes pour chaque fragment de gènes est présenté dans le tableau III. Des SNP sont trouvés, pour chaque fragment de gènes étudiés mais le nombre est très variable entre fragments : Pour une lecture moyenne de 617 bases, le nombre varie entre 1 et 13 SNP identifiés. Le nombre moyen de SNP par fragment de gène est égal à 7, la longueur totale des séquences analysées est de 5560 bases, nous observons en moyenne 1 SNP toutes les 84 bases. Pour chacun des SNP nous avons observé des individus hétérozygotes et des individus homozygotes.

Tableau III : Polymorphismes observés par EcoTILLING pour 9 gènes candidats positionnels, potentiellement liés à la rouille foliaire.

Genes	individus	No pb	Sequence	SNP	Haplotype	Hobs
<i>PopttrNBS-LRR83</i>	92	700	exon 1	8	15	0,81
<i>PopttrPK87</i>	84	510	introns 1, 2 exons 2, 4	12	18	0,78
gv98	92	800	Intron 4	1	2	0,34
<i>PopttrNBS-LRR102</i>	91	360	Exon 2	7	8	0,32
<i>PopttrTIR-NBS-LRR65</i>	92	500	Exon 2	4	5	0,61
<i>PopttrTIR-NBS-LRR67</i>	92	460	Exon 2	13	24	0,68
<i>PopttrTIR-NBS-LRR72</i>	92	850	Exon 2	5	6	0,5
<i>PopttrTIR-NBS-LRR73</i>	92	800	Exon 2	11	18	0,85
<i>PopttrTIR-NBS-LRR92</i>	92	580	exon 1	5	6	0,46
Total	819	5560		66		

No pb : nombre de paire de bases lisible sur le gel

Individus : nombre d'individus analysés

Haplotypes : le nombre d'haplotype a été estimé en utilisant le logiciel Haplotyper [15], [17]

Hobs : hétérozygotie observée par gène

4. DISCUSSION

Nos résultats montrent que l'EcoTILLING est une technique pour détecter les groupes d'haplotypes dont un représentant peut être séquencé pour identifier les génotypes. Ce résultat corrobore les premières conclusions de Gilchrist *et al.*, [7] qui ont montré que l'EcoTILLING pouvait être utilisé pour évaluer la diversité génétique de l'espèce *P. trichocarpa*. Dans notre étude 9 couples d'amorces (1) sur 20, initialement dessinées se révèlent utiles pour ce travail. La faible efficacité du dessin des amorces est probablement inhérente à plusieurs facteurs : - La séquence du génome est encore à l'état de brouillon qui possède des erreurs de séquence et des erreurs d'assemblage. - Le polymorphisme élevé entre individus appartenant à la même espèce. - La redondance du génome et en particulier la duplication des gènes de résistance. - La structure des gènes de résistance. L'introduction dans le protocole de deux PCR successives permet dans notre cas, de sélectionner des amorces spécifiques de locus mais aussi de limiter l'impact de la variation de qualité et de concentration entre les différents échantillons d'ADN sur les résultats. En effet, avec la miniaturisation des protocoles d'extraction d'ADN en plaque 96 puits, on observe des rendements et des qualités d'ADN très différents entre échantillons. La lecture des points de mutation qui peuvent exister à environ 50 bases des amorces est impossible, nos observations sont conformes à celles observées sur peuplier, melon et orge [7], [16], [14]. L'utilisation d'un allèle comme ADN de référence pour une espèce hétérozygote permet la détection sans ambiguïté des individus homozygotes et hétérozygotes contrairement aux travaux de Gilchrist *et al* [7], qui proposent d'utiliser l'ADN de Nisqually, génome dont la séquence est connue.

La lecture des gels peut être un obstacle au développement de l'EcoTILLING. Si on prend en compte le nombre d'individus et le nombre moyen de SNP par gène, chaque gel peut comporter plusieurs centaines de points de polymorphisme. Il est alors très difficile de différencier les SNP entre eux. En théorie, un point de mutation doit être visualisé par les 2 bandes complémentaires attendues, l'une marquée à la fluorescence IRDye 700, l'autre à la fluorescence IRDye 800. La séparation de produits de digestion par électrophorèse capillaire est probablement la solution pour le déploiement de l'EcoTILLING face à l'optimisation récente du séquençage direct de produits PCR [14]. La qualité des résultats de séquençage direct chez les espèces hétérozygotes dépend de la présence ou l'absence d'évènements d'insertion/délétion. La présence d'insertion/délétion provoque un décalage des pics des chromatogrammes, ce qui empêche la lecture des séquences. Les premiers résultats sont prometteurs dans le cas du séquençage de régions exoniques présentant peu d'insertion/délétion, mais pour les régions introniques, régulatrices ou en dehors de ces régions

les résultats sont beaucoup plus aléatoires. Dans ces cas l'EcoTILLING représente une alternative intéressante. Triques *et al.* [22] ont montré que l'activité de l'enzyme EndoI était suffisamment sensible pour reconnaître un mésappariement sur un hétéroduplex en mélange avec 7 homoduplexes [22]. Par conséquent, l'EcoTILLING doit aussi être efficace pour la détection des SNP rares.

Le taux de bases polymorphes observé est de 1 base polymorphe pour 84 bases dans la collection de 92 *P. trichocarpa*. Ce taux est plus élevé que celui observé par l'étude de Ingvarsson *et al.*, [8] chez *P. tremula* (une base polymorphe pour 93 bases) et chez *P. nigra* (Giusi Zaina, communication personnelle) (une base polymorphe pour 124 bases). L'étude faite par Gilchrist *et al.*, [7], sur 9 gènes d'intérêt de *P. trichocarpa* en EcoTILLING, a révélé un taux de bases polymorphes équivalent à celui observé chez *P. nigra* (une base polymorphe pour 124 bases). Le taux de base polymorphe plus élevé que nous observons est probablement du au niveau élevé de polymorphisme, généralement observé dans le cas des gènes de résistance [9].

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Cervera, M.T., Storme, V., Ivens, B., Gusmao, J., Liu, B.H., Hostyn, V., Van Slycken, J., Van Montagu, M. and Boerjan, W., 2001. Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* 158 : 787-809.
- [2] Colbert T., Till B.J., Tompa., Reynolds S., Steine M.N., Yeung A.T., McCallum., Comai L., Henikoff S., 2001. High-throughput screening for induced point mutations, *plant physiology*. 126: 480-484.
- [3] Comai L., Young K., Till S.H., Reynolds S.H., Greene E.A., Codomo C.A., Enns L.C., Johnson J.E., Burtner C., Odden A.R., Henikoff S., 2004. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling, *The Plant Journal* 37: 778-786.
- [4] Comai L., Henikoff S., 2006. TILLING: practical single nucleotide mutation discovery, *Plant J.*, 45: 684-694.
- [5] Dowkiw A., Bastien C., 2004. Characterisation of two major genetic factors controlling quantitative resistance to *Melampsora larici-populina* leaf rust in hybrid poplars: Strain specificity, field expression, combined effects, and relationship with a defeated qualitative resistance gene, *Phytopathology*, 94: 1358-1367.
- [6] Dowkiw A., Husson C., Frey P., Pinon J., Bastien C., 2003. Partial resistance to *Melampsora larici-populina* leaf rust in hybrid poplars: Genetic variability in inoculated excised leaf disk bioassay and relationship with complete resistance, *Phytopathology*, 93: 421-427.
- [7] Gilchrist E.J., Haughn G.W., Ying C.C., Otto S.P., Zhuang J., Cheung D., Hamberger B., Aboutorabi F., Kalynyak T., Johnson L., Bohlmann J., Ellis B.E., Douglas C.J., Cronk Q.C.B., 2006. Use of Ecotilling as an efficient SNP discov-

- ery tool to survey genetic variation in wild populations of *Populus trichocarpa*, Molecular Ecology, 15: 1367-1378.
- [8] Ingvarsson K., Garcia M.V., Hall, Luquez D., Jansson S., 2006. Clinal variation in *phyB2*, a candidate gene for day-length induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*), Genetics, 117: 1848-1853.
- [9] Jiang H., Wang C., Ping L., Tian D., Yang S., 2007. Pattern of LRR nucleotide in plant resistance genes, Plant science, 173: 253-261.
- [10] Jorge V., Dowkiw A., Faivre-Rampant P., Bastien C., 2005, - Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar : genetic mapping and QTL detection, New Phytologist, 167: 113-127.
- [11] Kohler A., Rinaldi C., Duplessis S., Baucher M., Geelen D., Hellsten U., Di-Fazio S., Boerjan W., Martin F., 2008. Genome-wide identification of NBS resistance gene in *Populus trichocarpa*. Plant Molecular Biology, 66(6): 619-636.
- [12] Lefèvre F., Goué-Mourier M.C., Faivre-Rampant P., Villar M., 1998. A single gene cluster controls incompatibility and partial resistance to various *Melampsora larici-populina* races in hybrid poplars, Phytopathology, 88: 156-163.
- [13] Lescot M., Rombauts S., Zhang J., Aubourg S., Mathé C., Janson S., Rouzé P., Boerjan W., 2004. Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements, Theor. Appl. Genet., 109: 10-22.
- [14] Mejlhede N., Kyjovska Z., Backes G., Burhenne K., Rasmussen S.K., Jahoor A., 2006. EcoTILLING for the identification of allelic variation in the powdery mildew resistance genes *mlo* and *Mla* of barley, Plant Breeding, 125: 461-467.
- [15] Murphy K., Towfique R., Winters R., White P., 2004. me-PCR: a refined ultra fast algorithm for identifying sequence-defined genomic elements, Bioinformatics, 20: 588-590.
- [16] Nieto C., Piron F., Damais M., Marco C.F., Moriones E., Gomes-Guillamon M. L., Truniger V., Gomez P., Garcia-Mas J., Aranda M. A., Bendhamane A., 2007. EcoTILLING for the identification of allelic variants of melon eIF4E a factor that controls virus susceptibility, BMC Plant Biology, 7 : 34
- [17] Niu T., Qin X., Xu X., Liu J., 2002. Bayesian haplotype inference for multiple linked single nucleotide polymorphisms. Am. J. Hum. Genet. 70: 157-159.
- [18] Pinon J., Frey P., 2005. Interactions between poplar clones and *Melampsora* population and their implications for breeding for durable resistance. In : Pei M., McCracken AR., eds. Rust diseases on willow and poplar. Wallingford, UK, CAB international. pp. 139-154.
- [19] Rozen S., Skaletsky H., 2000. Primer 3 on the www for general users and for biologist programmers, In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386.
- [20] Slade A.J., Knauf V.C., 2005, - TILLING moves beyond functional genomics into crop improvement, Transgenic Research, 14: 109-115.
- [21] Till B.J., Reynolds S.H., Green E.A., Codomo C.A., Enns L.C., Johnson J.E., Burtner C., Odden A.R., Young K., Taylor N.E., Henikoff J.G., Comai L.,

- Henikoff S., 2003. Large scale discovery of induced point mutations with high-throughput TILLING., *Genome Res.*, 13: 524-530.
- [22] Triques K., Sturbois B., Gallais S., Dalmais M., Chauvin S., Clepet C., Aubourg S., Rameau C., Caboche M., Bendahmane A., 2007. Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea, *The Plant J.*, 51:1116–1125.
- [23] Tuskan G.A., DiFazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A., Schein J., Sterck L., Aerts A., Bhale Rao R.R., Bhale Rao R. P., Blaudez D., Boerjan W., Brun A., Brunner A., Busov V., Campbell M., Carlson J., Chalot M., J. Chapman, G. Chen L., Cooper D., Coutinho P.M, Couturier J., Covert S., Cronk Q., Cunningham R., Davis J., S. Degroeve S., Déjardin A., de Pamphilis C, Detter J., Dirks B., I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, Kangasjärvi J., Karlsson J., C. Kelleher, R. Kirkpatrick, M. Kirst, Kohler A., Kalluri U., F. Larimer, Leebens-Mack J., Leplé J. C., Locascio P., Lou Y, S. Lucas, F. Martin, B. Montanini, C. Napoli, D. R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, Philippe R., Pilate G., Poliakov A., J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, Schmutz J., J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C.-J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer, Rokhsar D., 2006, - The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray), *Science*, 313: 11596-11604.
- [24] Yin T.M., DiFazio S.P., Gunter L.E., Jawdy S.S. Boerjan W., Tuskan G.A., 2004. Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence, *New Phytologist*, 164: 95-105.

Apports du séquençage multiloci à la phylogénie et à la taxonomie de deux genres majeurs de bactéries phytopathogènes : *Pseudomonas* et *Xanthomonas*

Marion FISCHER-LE SAUX^{(1)*}, Olivier PRUVOST⁽²⁾, Emilie FARGIER⁽¹⁾,
Nathalie AH-YOU⁽²⁾, Sophie BONNEAU⁽¹⁾, Carolina GONZALEZ⁽³⁾,
Lionel GAGNEVIN⁽²⁾, Valérie VERDIER⁽³⁾, Charles MANCEAU⁽¹⁾

⁽¹⁾INRA, UMR A77 Pathologie végétale, 42 rue Georges Morel BP60057,
49071 Beaucouzé cedex, France

⁽²⁾CIRAD, UMR C53 Peuplements Végétaux et Bioagresseurs en Milieu Tropical,
7, chemin de l'Irat Ligne Paradis, 97410 Saint Pierre, La Réunion, France

⁽³⁾IRD, UMR 5096 Laboratoire des Génomes et développement des plantes,
911, avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France

Abstract : Contribution of Multilocus sequence analysis to the phylogeny and taxonomy of two major groups of plant pathogenic bacteria : *Xanthomonas* and *Pseudomonas*. The genus *Xanthomonas* and pseudomonads of the “*syringae*” group encompass most of the plant pathogenic bacteria, among which several quarantine organisms registered by the European Council Directive 2000/29/CE and bioterror agents registered in dual-use goods regulations. However the taxonomy of these groups is still controversial or even out-of-date in the case of “*syringae*” group, which makes difficult the drafting and the revision of the statutory texts. Sequencing of protein-coding genes is more and more used to resolve phylogenetic relationships between highly related species, and within species. Multilocus sequence analysis (MLSA) is a potential alternative to DNA-DNA hybridizations to define bacterial species provided that a sufficient degree of congruence between both methods could be demonstrated.

In this study, we sequenced fragments of four housekeeping genes (*atpD*, *dnaK*, *efp* et *gyrB*) for a collection representative of *Xanthomonas* genus and *gyrB* and *rpoD* genes for more than one hundred strains representative of the nine genomospecies of the “*syringae*” group. Within these two taxa, species and genomospecies are supported by extensive DNA-DNA similarity values which allowed comparison of the methods. Our results showed that MLSA is as resolute as DNA-DNA hybridizations to circumscribe species or genomospecies, and phylogenetic trees are in agreement with previous DNA fingerprints results. Within “*syringae*” group, each genomospecies is represented by a discrete cluster supported by bootstrap values above 98%. The only incongruence is represented by genomospecies 3 and 8 which form a unique cluster in MLSA. Additional hybridization experiments are needed to solve

* Correspondance et tirés à part : marion.le-saux@angers.inra.fr

this point. Partial sequencing of only one gene is sufficient to assign unknown isolates to a genomospecies but insufficient to resolve the phylogenetic relationships between the genomospecies. Within *Xanthomonas*, MLSA data evidenced the distance of *X. albilineans*, *X. hyacinthi* and *X. translucens* from the core genus, and the relationships between previously and newly described species. In both genera, MLSA resolves phylogenetic relationships between highly related species more accurately than *rrs* trees. These data will strongly support the taxonomic studies that are needed in these two major groups of plant pathogenic bacteria.

***Pseudomonas*/ *Xanthomonas*/ phylogeny/ taxonomy/ housekeeping gene**

Résumé : Le genre *Xanthomonas* et les *Pseudomonas* du groupe « *syringae* » représentent deux groupes majeurs de bactéries phytopathogènes, pour autant leur taxonomie est encore controversée, voire obsolète dans le cas du groupe « *syringae* ». Le séquençage de plusieurs gènes de ménage (ou MLSA pour Multilocus Sequence Analysis) est une approche dont l'apport à la systématique des bactéries est indéniable. Elle permet de définir avec plus de précision les relations phylogénétiques entre espèces proches, pour lesquelles le gène *rrs* avait montré ses limites. Dans le cadre de ce projet nous avons séquençé les gènes *atpD*, *dnaK*, *efp* et *gyrB* pour une collection représentative du genre *Xanthomonas* et les gènes *gyrB* et *rpoD* pour une centaine de *Pseudomonas* du groupe « *syringae* ». Pour les deux genres, nos résultats montrent que la MLSA possède un seuil de résolution équivalent à celui des hybridations ADN-ADN, technique de référence pour la délimitation de l'espèce bactérienne. Les données obtenues sont cohérentes avec celles de cette technique et celles des empreintes génétiques. La MLSA a permis d'élucider les relations phylogénétiques qui lient les espèces et fournit des données essentielles pour les remaniements taxonomiques envisagés pour ces deux groupes.

***Pseudomonas*/ *Xanthomonas*/ phylogénie/ systématique moléculaire/ gène domestique**

1. INTRODUCTION

La systématique bactérienne connaît un tournant majeur en raison de l'accès généralisé au séquençage ; la phylogénie moléculaire n'utilise plus exclusivement les gènes de l'opéron ribosomique et le concept d'espèce bactérienne est ardemment discuté, à la lumière des apports de la génétique des populations. Le séquençage de gènes de ménage (Multilocus Sequence Analysis, MLSA) est envisagé comme une alternative aux hybridations ADN-ADN (méthode lourde et maîtrisée par peu de laboratoires) pour la description de nouvelles espèces [12]. Pour valider ce transfert, le comité *Ad Hoc* pour la ré-évaluation de la définition d'espèce en bactériologie encourage l'étude de collections de souches pour lesquelles de nombreuses données d'hybridation ADN-ADN sont disponibles [22]. La plupart des bactéries phytopathogènes appartiennent à des complexes d'espèces dont la

classification et la nomenclature demeurent encore confuses en raison des divergences entre les classifications phénotypiques et génotypiques. Cette particularité en fait d'excellents modèles pour l'étude du concept d'espèce et le développement de nouvelles approches de systématique.

Les bactéries phytopathogènes sont regroupées dans un nombre assez restreint de genres. Cette homogénéité taxonomique cache une diversité très grande qui s'exprime au niveau pathologique. L'exemple des *Pseudomonas* du « groupe *P. syringae* » et des *Xanthomonas*, divisés en des dizaines de pathovars définis par l'hôte et le type de maladie causée, est caractéristique. Ainsi, plus de 400 hôtes végétaux différents sont attaqués par ces bactéries, dont plusieurs espèces de grand intérêt économique (riz, canne à sucre, banane, *Citrus*...) et l'on assiste de façon récurrente à l'émergence de populations pathogènes responsables d'épidémies dont l'incidence socio-économique peut-être majeure. En Europe, ces observations ont conduit à l'inscription de plusieurs de ces organismes dans les annexes de la directive européenne 2000/29/CE (*P. syringae* pv. *persicae*, *X. arboricola* pv. *pruni*, *X. axonopodis* pv. *phaseoli*...) et dans la réglementation des biens à double-usage (civil et militaire) : *X. albilineans*, *Xanthomonas* pathogènes des *Citrus* et *X. oryzae* pv. *oryzae*.

La taxonomie du genre *Xanthomonas* a été révisée par Vauterin *et al.* [24], qui a décrit 20 espèces sur la base de données d'hybridations ADN-ADN. Depuis de nouvelles espèces ont été décrites dont certaines remettent en cause la classification établie par Vauterin *et al.* [24] entraînant des confusions [14], [8], [20], [21], [9]. Les *Pseudomonas* du « groupe *P. syringae* » constituent une lignée monophylétique au sein du genre *Pseudomonas* [16], [4] qui regroupe dix espèces et plus de 50 pathovars. Les travaux de Gardan *et al.* [11], basés sur l'hybridation ADN-ADN ont montré qu'au moins neuf genomospecies pouvaient être distinguées chez les pathovars de *P. syringae* et les espèces apparentées, mettant ainsi en évidence des espèces synonymes et des espèces cryptiques. En raison de l'absence de caractères phénotypiques discriminants, seules deux genomospecies ont pu être élevées au rang d'espèce.

Ainsi, un écueil majeur, pour la caractérisation de ces bactéries émergentes en vue de l'établissement et de la révision des textes réglementaires, est une systématique encore à ce jour controversée (cas des *Xanthomonas*) ou non révisée (cas des *Pseudomonas*), et on constate l'absence d'outils simples d'identification tant au niveau de l'espèce que du pathovar. Les techniques actuellement utilisées (AFLP et rep-PCR) pour caractériser les pathovars [19], [18] génèrent des données difficilement reproductibles, et ne permettent pas la constitution de bases de données ouvertes à la communauté scientifique.

Le projet a pour objectif de caractériser des collections représentatives du genre *Xanthomonas* et du « groupe *P. syringae* » par séquençage de plusieurs

gènes de ménage, et de comparer les structures phylogénétiques obtenues aux données génomiques (AFLP et hybridations ADN-ADN).

Il s'appuie sur une ressource majeure dans le domaine, la Collection Française de Bactéries Phytopathogènes (CFBP), entretenue par l'UMR A77 (INRA, Beaucozéz, France).

2. MATÉRIEL ET METHODES

2.1. Collections étudiées

La très grande majorité des souches utilisées dans cette étude proviennent de la CFBP. Pour le « groupe *P. syringae* » nous avons caractérisé l'ensemble des souches des travaux de Gardan *et al.* [11] auquel nous avons ajouté des représentants des pathovars décrits depuis cette date, ainsi que les souches type de *P. congelans* et *P. caricapapayae*. La collection de souches de *Xanthomonas* représente la diversité spécifique du genre. Une emphase particulière a été donnée à l'espèce *X. axonopodis*, la plus diverse du genre. Les souches ont été cultivées sur milieu LPGA (Levure 7g/L, Peptone 7 g/L, Glucose 7g/L, agar, 15 g/L) à 27 °C.

2.2. Amplification PCR et séquençage des gènes de ménage

Pour les souches du genre *Pseudomonas*, l'amplification PCR a été réalisée à partir de suspensions bactériennes bouillies. Une portion d'environ 900 pb des gènes *gyrB* (codant pour la sous-unité B de l'ADN gyrase) et *rpoD* (codant pour l'ARN polymérase facteur sigma-70) a été amplifiée selon le protocole décrit par Yamamoto *et al.* [26].

L'ADN des souches de *Xanthomonas* a été extrait suivant le protocole décrit par Llop *et al.* [15]. Les séquences partielles (environ 500 pb) des quatre gènes domestiques suivants (répartis sur le chromosome de *X. campestris* pv. *campestris* souche ATCC33913) : *atpD* (codant pour la chaîne beta de l'ATP synthase), *dnaK* (codant pour la protéine chaperonne DnaK), *efp* (codant pour le facteur d'élongation P) et *gyrB* ont été obtenues à l'aide des amorces décrites dans le tableau I.

Tableau I : Amorces utilisées pour l'amplification de fragments de gènes dans le genre *Xanthomonas*.

Gène	Amorce PCR	Sequence (5'-3')	Amorce pour le séquençage	Sequence (5'-3')
<i>atpD</i>	P-X-ATPD-F*	GGGCAAGATCGTTTCAGAT	emiATPD1F	TTCAGATCATCGGCGCGGT
	P-X-ATPD-R*	GCTCTTGGTCGAGGTGAT	emiATPD2R	TTGGTCGAGGTGATGCGCT
<i>dnaK</i>	P-X-DNAK-F*	GGTATTGACCTCGGCACCAC	emidnaK1F	ACCAAGGACGGCGAAGTGCT
	P-X-DNAK-R*	ACCTTCGGCATACGGGTCT	emidnaK2R	CGATCGACTTCTTGACCAGG
<i>efp</i>	P-X-EFP-F*	TCATCACCGAGACCGAATA	emiefp1F	TCACCGAGACCGAATACG
	P-X-EFP-R*	TCCTGGTTGACGAACAGC	P-X-EFP-R*	TCCTGGTTGACGAACAGC
<i>gyrB</i>	emigyrB1F	TGCGCGGCAAGATCCTCAAC	emigyrB2F	CGCTACCACCGCATCATCT
	emigyrB4R	GCGTTGTCTCGATGAAGTC	emigyrB3R	AGGTGCTGAAGATCTGGTGC

Les produits PCR ont été confiés à la plate-forme OUEST-Génopole ou à Macrogen pour purification et séquençage.

2.3. Analyses phylogénétiques

Les alignements ont été réalisés à l'aide du logiciel Multalin [7]. Les données de séquences ont été concaténées et des arbres phylogénétiques ont été réalisés à l'aide de la méthode Neighbor-Joining (en utilisant le paramètre kimura 2) du package PHYLIP [10]. La robustesse de la topologie obtenue a été évaluée par une analyse de ré-échantillonnage (bootstrap, 1000 réplicats).

2.4. Hybridations ADN-ADN

Les ADN génomiques utilisés pour les hybridations ADN-ADN ont été extraits suivant le protocole de Brenner *et al.* [6]. Les hybridations ADN-ADN ont été effectuées par la technique dite nucléase S1 / acide trichloroacétique comme décrit par Gardan *et al.* [11].

3. RÉSULTATS ET DISCUSSION

3.1. Phylogénie du complexe « *P. syringae* et espèces apparentées »

L'analyse phylogénétique des séquences met en évidence des résultats très comparables avec les deux gènes *gyrB* et *rpoD* : une structuration en groupes discrets, portés par de longues branches, et validés par de fortes valeurs de bootstrap (fig. 1). Lorsqu'on compare ces résultats aux données d'hybridation ADN-ADN, on constate que la plupart des genomospecies correspondent à ces groupes monophylétiques validés par de très fortes valeurs de bootstrap. C'est le cas des espèces *P. viridiflava* (genomospecies 6, notée G6) et *P. cannabina* (G9), et des genomospecies G2, G4 et G7. La genomospecies G1 forme un groupe monophylétique sur l'arbre *rpoD*, mais pas sur l'arbre *gyrB*.

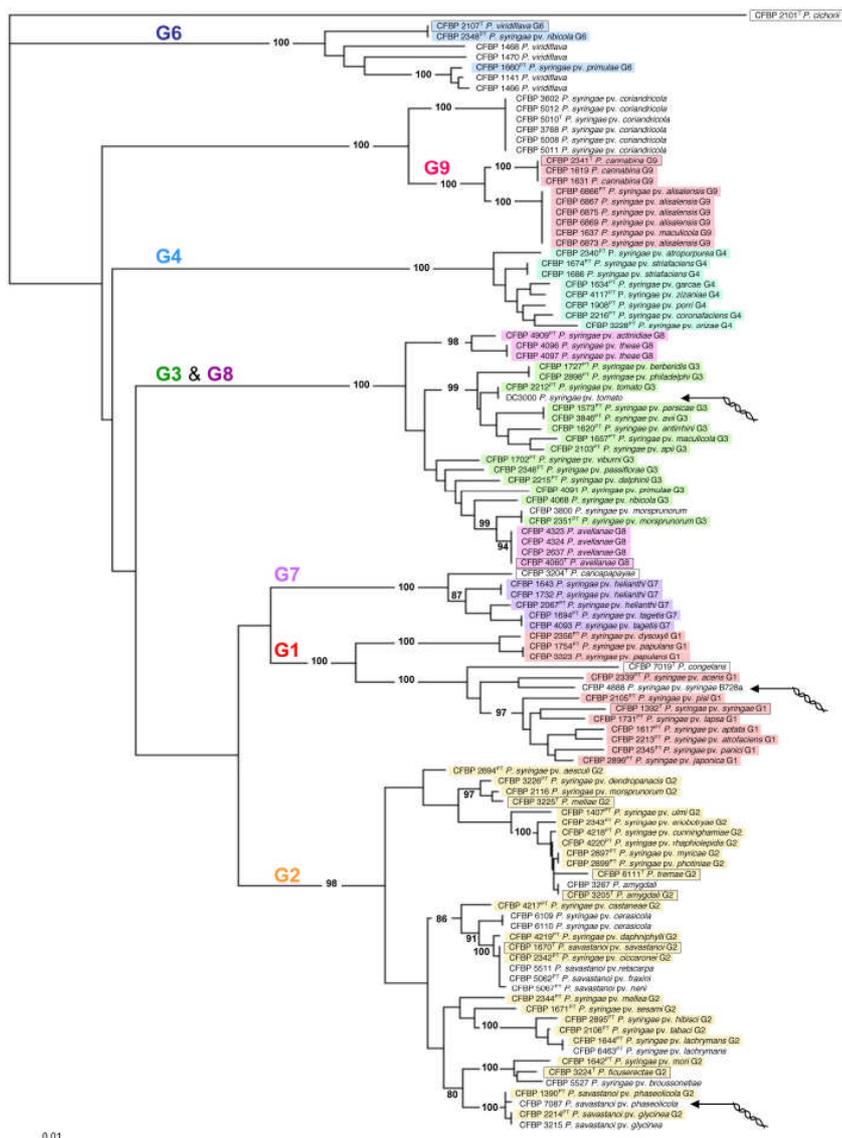


Figure 1 : Arbre phylogénétique des *Pseudomonas* du « groupe *syringae* » construit par Neighbor-Joining d'après les séquences concaténées de fragments des gènes *rpoD* et *gyrB*. Les valeurs de bootstrap significatives ($\geq 80\%$) sont indiquées (1 000 réplicats). Les souches dont l'appartenance à une genomospécies a été déterminée par hybridation ADN-ADN sont surlignées en couleur ; le nom de la genomospécies est précisé après le nom du taxon (ex. G1). Les souches type sont encadrées. Les souches dont le génome est séquéncé sont indiquées par une flèche.

En revanche, sur les deux arbres, les genomospecies G3 et G8 (pv. *theae*, *actinidiae* et *P. avellanae*) sont mélangées dans le même groupe phylogénétique. Le même résultat est obtenu sur le dendrogramme de F-AFLP (données du laboratoire non montrées). Il s'agit du seul cas d'incongruence entre les données MLSA et F-AFLP d'une part, et hybridations ADN-ADN d'autre part. Il est à noter que ces deux genomospecies ne sont distinguées que sur la base de deux valeurs d'hybridations [11]. Une étude d'hybridations ADN-ADN avec de plus nombreuses valeurs croisées entre ces deux genomospecies est en cours pour confirmer ou infirmer l'existence de deux genomospecies distinctes.

Par ailleurs, cette étude a permis de mettre en évidence plusieurs synonymies de noms d'espèce. L'espèce *P. tremae*, définie comme la genomospecies 5 par Gardan *et al.* [11] se place au sein des souches de la genomospecies 2 dans cette étude. Nous avons confirmé ce résultat par hybridations ADN-ADN : *P. tremae* hybride à 71-80 % avec les souches de la genomospecies 2 et 28-52 % avec les souches des autres genomospecies. Ainsi *P. tremae*, *P. meliae*, *P. amygdali*, *P. savastanoi* et *P. fuscisectae* appartiennent à la même genomospecies.

L'espèce *P. congelans*, qui correspond à des souches isolées de la phyllosphère d'herbe, se place clairement parmi les souches de la genomospecies 1 et ne constituerait donc pas une nouvelle espèce. Sur la base d'un arbre phylogénétique du gène de l'ARNr16S, Behrendt *et al.* [5] avaient identifié *P. savastanoi* (genomospecies 2) comme l'espèce la plus proche et n'avaient utilisé que cette espèce pour les hybridations AND-ADN. Des discordances entre les topologies des arbres déduits du gène *rrs* et des gènes de ménage ont déjà été montrées dans le genre *Pseudomonas* [26], [2]. Elles sont dues au faible pouvoir résolutif du gène *rrs* lorsqu'il s'agit d'étudier des espèces ayant divergé très récemment. Cet exemple met en exergue les limites de la phylogénie basée uniquement sur le gène de l'ARNr 16S pour révéler les relations phylogénétiques entre espèces proches et identifier les plus proches voisins phylogénétiques d'un isolat lors des études taxonomiques.

L'espèce *P. caricapapayae* se place avec les souches de la genomospecies 7. Seule souche type d'espèce de ce groupe identifiée à ce jour, elle donnera le cas échéant son nom à cette genomospecies.

Les pathovars décrits depuis 1999, ont pu être classés sans ambiguïté dans une genomospecies. *P. syringae* pv. *fraxini*, *nerii*, *castanae*, *cunninghamiae*, *daphniphylli*, *raphiolepidis*, *retacarpa*, *broussonetiae*, *cerasicola* appartiennent à la genomospecies 2. Ce résultat a été confirmé par hybridation ADN-ADN pour quatre de ces pathovars. *P. syringae* pv. *alisalensis* et pv. *coriandricola* se regroupent avec *P. cannabina* (genomospecies 9), mais ces trois taxons forment

trois branches distinctes. *P. syringae* pv. *spinaceae*, *avii* et *actinidiae* se placent dans le groupe phylogénétique incluant les souches des genomospecies 3 et 8.

Le génome de trois souches du « groupe *P. syringae* » est actuellement séquencé. Il s'agit de *P. syringae* pv. *syringae* B728a, *P. syringae* pv. *phaseolicola* 1448A et *P. syringae* pv. *tomato* DC3000. Bien qu'elles portent le même nom d'espèce, ces trois souches représentent trois genomospecies différentes (fig. 1). Afin que ce statut d'espèces différentes soit intégré dans les études de génomique comparative, il devient urgent de nommer ces genomospecies.

La MLSA est donc un outil puissant dont le niveau de discrimination est équivalent à celui des hybridations ADN-ADN. Le séquençage d'un seul gène suffit pour attribuer une souche à une genomospecies, mais est insuffisant pour décrire les relations phylogénétiques qui les lient. En perspectives, nous envisageons d'offrir aux phytopathologistes un langage commun en proposant un statut d'espèce aux genomospecies G1, G2, G3/G8, G4, et G7. L'absence de caractères phénotypiques discriminants pour identifier ces espèces sera compensée par un outil rapide d'identification des espèces par comparaison des séquences du gène *gyrB* ou *rpoD* à une base de données de référence disponible sur internet (site de la CFBP).

3.2. Phylogénie du genre *Xanthomonas*

Une représentation schématique de l'arbre phylogénétique du genre *Xanthomonas* obtenu après concaténation des séquences des quatre gènes de ménage *atpD*, *dnaK*, *efp* et *gyrB*, est présentée figure 2. Les différentes espèces décrites par Vauterin *et al.* [25] forment des lignées distinctes soutenues par de fortes valeurs de bootstrap (ex. *X. campestris*, *X. vesicatoria*, *X. arboricola*). Certaines espèces sont regroupées en groupes monophylétiques soutenus par de fortes valeurs de bootstrap. C'est le cas par exemple de *X. hortorum*, *X. populi* et *X. cynarae* (bootstrap de 100 %) et de *X. albilineans*, *X. hyacinthi* et *X. translucens* (100 %) (fig. 2). Ce dernier groupe est phylogénétiquement très éloigné des autres espèces du genre *Xanthomonas*, et l'amplification de certains gènes de souches de ce groupe a posé problème. Le génome d'une souche de *X. albilineans* sera bientôt disponible. Ces données permettront de dessiner de nouvelles amorces plus adaptées à ce groupe d'espèces très divergentes.

Cette structuration génétique des espèces de *Xanthomonas* est en accord avec les résultats de Parkinson *et al.* [17] basé sur la séquence partielle du gène *gyrB*. Cependant, l'arbre construit avec les 4 gènes de ménage souligne de façon beaucoup plus marquée, la divergence des espèces *X. albilineans*, *X. translucens* et *X. hyacinthi* des autres espèces du genre. Cette observation rejoint les résultats de Hauben *et al.* [13] basé sur le gène *rrs*. Ces auteurs soulignent le très haut degré de conservation du gène *rrs* (moyenne des valeurs de similitudes supérieure à 98 % au sein du genre *Xanthomonas*) et un niveau de résolution qui n'atteint pas

celui des hybridations ADN-ADN. Il en résulte que les différentes souches d'une même espèce ne forment pas de groupes monophylétiques. Au contraire, les gènes codant des protéines utilisés dans la présente étude, présentent un niveau de résolution comparable à celui des hybridations ADN-ADN et sont de bons marqueurs de la différenciation des espèces.

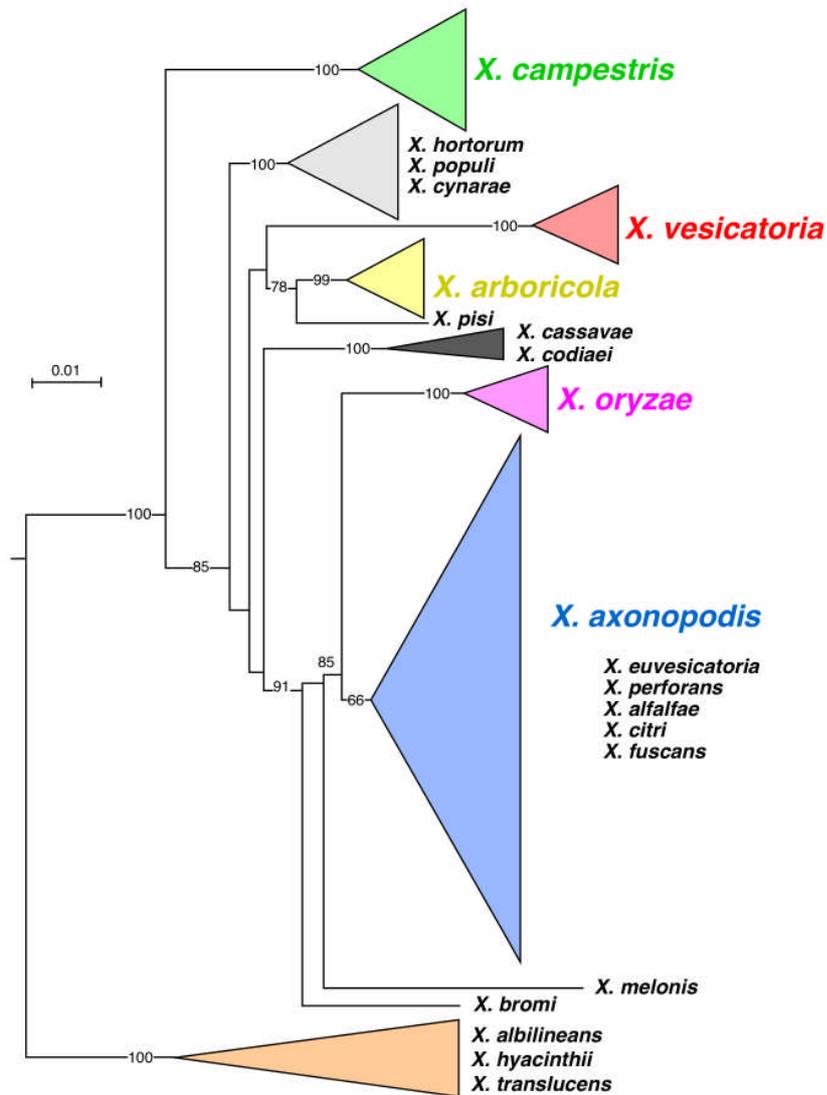


Figure 2 : Arbre phylogénétique du genre *Xanthomonas* construit par Neighbor-Joining à partir des données concaténées de séquences partielles de quatre gènes de

ménage (*atpD*, *dnaK*, *efP* et *gyrB*). Les valeurs de bootstrap (>60) après 1000 simulations sont représentées sur cette figure.

Depuis la révision du genre *Xanthomonas* par Vauterin *et al.* [24], sept nouvelles espèces ont été décrites. La caractérisation de l'agent pathogène émergent responsable de la graisse de l'artichaut a conduit à la description de *X. cynarae* [23]. L'analyse MLSA place cette espèce dans le groupe incluant *X. hortorum* et *X. populi* (fig. 2). Ce résultat concorde avec les données d'hybridations ADN-ADN, qui montrent que sur les 20 espèces testées, les plus fortes valeurs d'hybridation sont trouvées avec ces deux espèces (49 %).

Les espèces *X. euvesicatoria*, *X. perforans*, *X. alfaiae*, *X. citri* et *X. fuscans* ont été décrites suite à l'étude d'une collection de souches représentant des pathovars de l'espèce *X. axonopodis* [14], [20], [21]. Ces nouvelles espèces se placent dans l'analyse MLSA au sein du groupe constitué des souches de *X. axonopodis* dans lequel elles forment chacune des groupes monophylétiques présentant de fortes valeurs de bootstrap (fig. 2 et 3). Les propositions de Jones *et al.* [14] et Schaad *et al.* [20] ne reviennent donc pas à proposer de nouvelles espèces dans un système de classification établi, mais à proposer un nouveau système de classification qui place la limite d'espèce à un seuil inférieur. Ces deux schémas de classification se chevauchant (deux nomenclatures différentes sont possible pour les mêmes souches), ils ne sont pas compatibles et c'est à l'utilisateur de faire le choix d'adopter l'une ou l'autre des nomenclatures. Les données MLSA ne contredisent aucune des deux classifications (les espèces proposées forment toutes des lignées phylogénétiques distinctes).

L'utilisation d'une collection partielle par Schaad *et al.* [20] pose cependant problème. Par exemple, il a été montré récemment par F-AFLP que *X. axonopodis* pv. *phaseoli* est constitué de 4 lignées génétiques [3]. Les représentants d'une seule de ces lignées ont été étudiés par Schaad *et al.* [20] et reclassés dans l'espèce *X. fuscans* subsp. *fuscans*. Nos données MLSA (fig. 3) montrent que si l'on adoptait la classification proposée par Schaad *et al.* [20] les trois lignées génétiques restantes de *X. axonopodis* pv. *phaseoli* se placeraient dans des espèces ou sous-espèces différentes qui restent à décrire. Une problématique similaire a été montrée avec les souches du pathovar *mangiferaeindicae* [1].

En conclusion, nos données soulignent l'hétérogénéité génétique de *X. axonopodis* (Approved Lists 1980) emend. Vauterin *et al.* [24] par rapport aux autres espèces du genre. La structure génétique de cette espèce révélée par MLSA est cohérente avec les résultats antérieurs de rep-PCR et d'AFLP [18], [3], [1] et d'hybridations ADN-ADN [24], [14], [20]. La question fondamentale soulevée par ces données est «où placer le seuil de délimitation de l'espèce ?». Une étude plus approfondie par une approche

polyphasique permettra de clarifier la taxonomie de ce groupe qui inclut plusieurs organismes d'importance réglementaire du point de vue de la quarantaine ou du bioterrorisme.

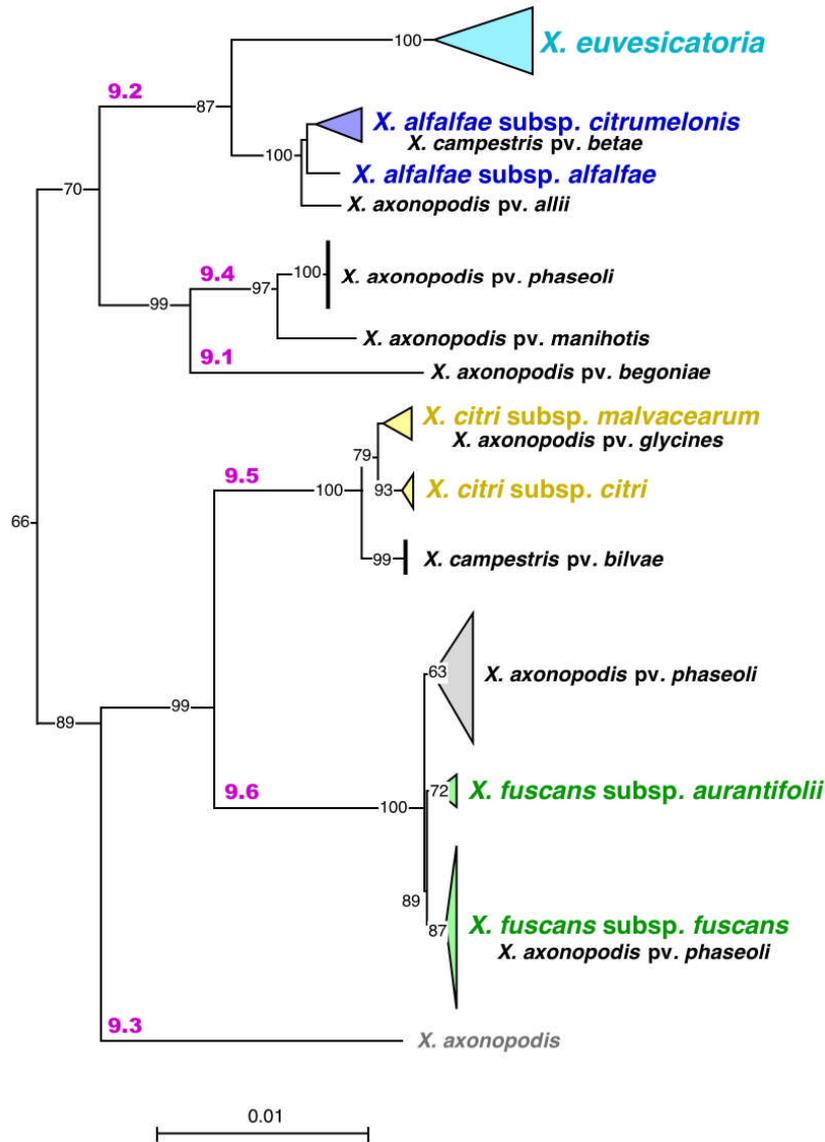


Figure 3 : Portion de l'arbre phylogénétique du genre *Xanthomonas*, correspondant à l'espèce *X. axonopodis* (Approved Lists 1980) emend. Vauterin *et al.* 1995 [24]. Cet arbre a été construit par Neighbor-Joining à partir des données concaténées de séquences partielles de quatre gènes de ménage (*atpD*, *dnaK*, *efP* et *gyrB*). Les valeurs

de bootstrap (>60) après 1000 simulations sont représentées sur cette figure. Les indications 9.1 à 9.6 désignent la correspondance avec les groupes identifiés par Rademaker *et al.* [18] par rep-PCR.

4. CONCLUSION

Nos résultats montrent que pour les deux groupes bactériens majeurs de bactéries phytopathogènes étudiés, la MLSA possède un seuil de résolution équivalent à celui des hybridations ADN-ADN, et fournit des données cohérentes avec cette technique et celles des empreintes génétiques. Cette approche, facile à mettre en œuvre, générant des données facilement échangeables et pouvant être mises à disposition dans des bases de données, représente incontestablement une alternative prometteuse aux hybridations ADN-ADN pour la délimitation des espèces bactériennes. Il reste cependant à définir quels critères utiliser pour délimiter le seuil de l'espèce bactérienne. Au-delà de règles strictes et de seuils, le plus raisonnable serait de conserver une approche polyphasique intégrant la topologie de l'arbre MLSA et un maximum de données génotypiques, phénotypiques et écologiques.

Les séquences des gènes de ménage issues du présent projet viennent enrichir la base de données liée aux accessions de la CFBP. En interne, ces séquences représentent « la carte d'identité génétique » des souches et pourront être utilisées au cours des contrôles qualité visant l'authentification de l'identité du matériel biologique détenu à la CFBP. Ces contrôles répondent aux exigences des lignes directrices de l'OCDE relatives aux pratiques exemplaires concernant les centres de ressources biologiques et aux exigences de la norme AFNOR NF S 96-900 « qualité des centres de ressources biologiques » en cours de préparation. En externe, l'objectif est de créer un service d'identification interactif des *Xanthomonas* et des *Pseudomonas* sur le site web de la CFBP (<http://www.angers.inra.fr/cfbp/>) au travers d'outils dédiés à l'identification moléculaire (BLAST en lien avec des bases de données d'intérêt taxonomique, logiciels d'alignements et de reconstruction phylogénétique).

REMERCIEMENTS

Ces travaux ont été financés par le Bureau des Ressources Génétiques (AP 2005-2006). La majorité du séquençage a été réalisée par la plate-forme séquençage et géotypage de OUEST-Génopole. Alain Huard est remercié pour son aide dans la mise en forme des figures.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Ah-You N., Gagnevin L., Chiroleu E., Jouen E., Neto J.R., Pruvost O., 2007. - Pathological variations within *Xanthomonas campestris* pv. *mangiferaeindicae* support its separation into three distinct pathovars that can be distinguished by amplified fragment length polymorphism. *Phytopathology*, 97: 1568-1577.
- [2] Ait Tayeb L., Ageron E., Grimont F., Grimont P.A.D., 2005. - Molecular phylogeny of the genus *Pseudomonas* based on *rpoB* sequences and application for the identification of isolates. *Research in Microbiology*, 156: 763-773.
- [3] Alavi S.M., Sanjari S., Durand F., Brin C., Manceau C., Poussier S., 2008. - Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations. *Applied and Environmental Microbiology*, 74: 3295-3301.
- [4] Anzai Y., Kim H., Park J.Y., Wakabayashi H., Oyaizu H., 2000. - Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *International Journal of Systematic and Evolutionary Microbiology*, 50: 1563-1589.
- [5] Behrendt U., Ulrich A., Schumann P., 2003. - Fluorescent pseudomonads associated with the phyllosphere of grasses; *Pseudomonas trivialis* sp. nov., *Pseudomonas poae* sp. nov. and *Pseudomonas congelans* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 53: 1461-1469.
- [6] Brenner D.J., McWhorter A.C., Knutson J.K., Steigerwalt A.G., 1982. - *Escherichia vulneris*: a new species of Enterobacteriaceae associated with human wounds. *Journal of Clinical Microbiology*, 15: 1133-1140.
- [7] Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16: 10881-10890
- [8] Euzéby J., 2006. - List of new names and new combinations previously effectively, but not validly, published. *International Journal of Systematic and Evolutionary Microbiology*, 56: 925-927.
- [9] Euzéby J., 2007. - List of new names and new combinations previously effectively, but not validly, published. *International Journal of Systematic and Evolutionary Microbiology*, 57: 893-897.
- [10] Felsenstein J., 1989. - PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics*, 5: 164-166.
- [11] Gardan L., Shafik H., Belouin S., Broch R., Grimont F., Grimont P.A.D. 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *International Journal of Systematic Bacteriology*, 49: 469-478.
- [12] Gevers D., Cohan F.M., Lawrence J.G., Spratt B.G., Coenye T., Feil E.J., Stackebrandt E., Van de Peer Y., Vandamme P., Thompson F.L., Swings J., 2005. - Opinion: Re-evaluating prokaryotic species. *Nature Reviews. Microbiology*, 3: 733-739.
- [13] Hauben L., Vauterin L., Swings J., Moore E.R.B., 1997. - Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *International Journal of Systematic Bacteriology*, 47: 328-335.

- [14] Jones J.B., Lacy G.H., Bouzar H., Stall R.E., Schaad N.W., 2004. - Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Systematic and Applied Microbiology*, 27: 755-762.
- [15] Llop P., Caruso P., Cubero J., Morente C., Lopez M.M., 1999. - A simple extraction procedure for efficient routine detection of pathogenic bacteria in plant material by polymerase chain reaction. *Journal of Microbiological Methods* 37: 23-31.
- [16] Moore E.R.B., Mau M., Arnscheidt A., Bottger E.C., Hutson R.A., Collins M.D., Van de Peer Y., De Wachter R., Timmis K.N., 1996. - The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas (sensu stricto)* and estimation of the natural intrageneric relationships. *Systematic and Applied Microbiology*, 19: 478-492.
- [17] Parkinson N., Aritua V., Heeney J., Cowie C., Bew J., Stead D., 2007. - Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, 57: 2881-2887.
- [18] Rademaker J.L.W., Louws F.J., Schultz M.H., Rossbach U., Vauterin L., Swings J., de Bruijn F.J., 2005. - A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology*, 95: 1098-1111.
- [19] Roumagnac P., Gagnevin L., Gardan L., Sutra L., Manceau C., Dickstein E.R., Jones J.B., Rott P., Pruvost O., 2004. - Polyphasic characterization of xanthomonads isolated from onion, garlic and Welsh onion (*Allium* spp.) and their relatedness to different *Xanthomonas* species. *International Journal of Systematic and Evolutionary Microbiology*, 54: 15-24.
- [20] Schaad N.W., Postnikova E., Lacy G.H., Sechler A., Agarkova I., Stromberg P.E., Stromberg V.K., Vidaver A.K., 2005. - Reclassification of *Xanthomonas campestris* pv. *citri* (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and *X. alfalfae* subsp. *citrumelo* (ex Riker and Jones) Gabriel *et al.*, 1989 sp. nov. nom. rev. comb. nov.; *X. campestris* pv. *malvacearum* (ex Smith 1901) Dye 1978 as *X. smithii* subsp. *smithii* nov. comb. nov. nom. rev.; *X. campestris* pv. *alfalfae* (ex Riker and Jones, 1935) Dye 1978 as *X. alfalfae* subsp. *alfalfae* (ex Riker *et al.*, 1935) sp. nov. nom. rev.; and "var. *fuscans*" of *X. campestris* pv. *phaseoli* (ex Smith, 1987) Dye 1978 as *X. fuscans* subsp. *fuscans* sp. nov. *Systematic and Applied Microbiology*, 28: 494-518.
- [21] Schaad N.W., Postnikova E., Lacy G., Sechler A., Agarkova I., Stromberg P.E., Stromberg V.K., Vidaver A.K., 2006. - Emended classification of xanthomonad pathogens on citrus. *Systematic and Applied Microbiology*, 29: 690-695.
- [22] Stackebrandt E., Frederiksen W., Garrity G.M., Grimont P.A.D., Kampfner P., Maiden M.C.J., Nesme X., Rossello-Mora R., Swings J., Truper H.G., Vauterin L., Ward A.C., Whitman W.B., 2002. - Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 52: 1043-1047.
- [23] Trebaol G., Gardan L., Manceau C., Tanguy J.L., Tirilly Y., Boury S., 2000. - Genomic and phenotypic characterization of *Xanthomonas cynarae* sp. nov., a new species that causes bacterial bract spot of artichoke (*Cynara scolymus* L.). *International Journal of Systematic and Evolutionary Microbiology*, 50: 1471-1478.

- [24] Vauterin L., Hoste B., Kersters K., Swings J., 1995. - Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology*, 45: 472-489.
- [25] Vauterin L., Rademaker J., Swings J., 2000. - Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology*, 90: 677-682.
- [26] Yamamoto S., Kasai H., Arnold D.L., Jackson R.W., Vivian A., Harayama S., 2000. - Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology*, 146: 2385-2394.

**Diversité nucléotidique pour le gène
PHANTASTICA (PHAN) le long d'un gradient
altitudinal chez le frêne commun
(*Fraxinus excelsior* L.) et chez le frêne
oxyphylle (*Fraxinus angustifolia* Valh)**

Nathalie FRASCARIA-LACOSTE^{(1)*}, Juan FERNANDEZ-MANJARRES⁽²⁾,
Jolly BASAK^{(2)§}, Pierre GÉRARD^{(2)**}, Paola BERTOLINO⁽³⁾,
Jean DUFOUR⁽⁴⁾, Corinne CRUAUD⁽⁵⁾, Arnaud COULOUX⁽⁵⁾,
Sophie MASSOT⁽⁶⁾

⁽¹⁾ Laboratoire Ecologie, Systématique, et Evolution (ESE), UMR CNRS-UPS-AgroParisTech 8079, Bâtiment 360, Univ Paris Sud, 91 405 Orsay Cedex, France

⁽²⁾ Univ Paris Sud, 91 405 Orsay Cedex, France

⁽³⁾ CNRS, 91405 Orsay Cedex France (paola.bertolino@u-psud.fr)

⁽⁴⁾ INRA, Unité d'Amélioration génétique et de physiologie forestière, BP 20619, 45166 Ardon, Olivet, France

⁽⁵⁾ Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, CP5706, 91057 Evry cedex, France

⁽⁶⁾ Institut de Biotechnologie des Plantes, UMR CNRS-UPS 8618, Bât 360, Univ Paris Sud, 91 405 Orsay Cedex, France

§ adresse actuelle : Leibniz Institute of Plant Biochemistry, Department of Stress and Developmental Biology, Weinberg 3, 06120 Halle, Germany

** adresse actuelle : Department of Biology, University of Rochester, Rochester, NY 14627 USA

Abstract: Nucleotide diversity at the *PHANTASTICA* gene (*PHAN*) along an altitudinal gradient for common ash (*Fraxinus excelsior* L.) and narrow-leaved ash (*Fraxinus angustifolia* Valh.). The extent and pattern of DNA sequence variation in natural populations can provide useful information about the evolutionary forces acting on a species. In this respect, an important point is to understand how molecular variation at a single-gene level is affected by selection at higher phenotypic level. This is a difficult task, as the effects of different evolutionary factors are confounded. Clinal variation in genetically traits can provide compelling evidence for spatially varying selection across an environmental gradient. One means of distinguish between adaptative and no adaptive causes of geographical variation is to compare relative levels of between-population divergence in quantitative traits, neutral DNA markers and candidate genes. *Fraxinus excelsior* L. (common ash) and *Fraxinus angustifolia* Valh. (narrow-leaved ash) are two widespread trees in Western Europe whose

* Correspondance et tirés à part : nathalie.frascaria@u-psud.fr

present distribution suggests an important influence of isolation and recolonisation processes shaped by different ecological preferences. Common ash grows on low and middle elevation hills while narrow-leaved ash prefers areas along rivers where flooding may occur. *Fraxinus excelsior* flowers during March-April while *F. angustifolia* begins flowering in late December.

For these two species, nucleotide variation was examined at two candidate gene, *PHANTASTICA* (*PHAN*) and *FRIGIDA* gene along eight populations of an altitudinal gradient in Oriental Pyrenees. *PHANTASTICA* gene encoded a MYB transcription factor involved in the leaves form. *FRIGIDA* gene is a major gene involved in the regulation of flowering time. These populations were also screen for variation at molecular markers (4 nuclear microsatellites and 6 chloroplastic microsatellites) and at morphological (foliar characteristics) and phenological traits (flowering period). Along this gradient, *Fraxinus angustifolia* populations were at low altitude while *F. excelsior* populations were at middle elevation.

Canonical discriminant analysis indicated that the number of leaflets and leave areas had significant differences along the altitudinal gradient. A clear temporal isolation by divergent floral phenologies is observed between the two species. No chloroplastic microsatellites variability was found and nuclear microsatellite variability is high along the gradient. Deficiency values were high but lower for *Fraxinus angustifolia* compared to *F. excelsior* indicating probable temporal Wahlund effect. Values for genetic structure were moderate. Higher F_{st} were found in the case of populations within species ($F_{st} = 0.0325$) compared to populations within total ($F_{st} = 0.0227$).

The *FRIGIDA* sequences obtained were impossible to interpret without any equivalent under Genbank. *PHANTASTICA* sequences pointed out 57 polymorphic sites along 392 bp obtained.

To infer natural selection from DNA polymorphism data, the Tajima's D test was used. The negative value obtained here concerning 103 sequences revealed either a recent rapid expansion of *Fraxinus* populations, either purifying selection. The next studies would requires others markers in the whole genome to better understand the demographic history of the populations. The idea is not to overstate the prevalence of natural selection in plant genomes but to present only the right selected genes.

***Fraxinus* / PHANTASTICA gene / nuclear microsatellites / altitudinal gradient / local adaptation**

Résumé : Un aspect important en génétique des populations est de comprendre à l'échelle nucléotidique, la base génétique de la variation adaptative gouvernée par la sélection. Le jeu entre l'histoire des populations, le système de reproduction et la sélection détermine la nature des variations et délimite les mécanismes qui permettent la diversification. Dans ce travail, nous souhaitons mieux comprendre la relation entre différents traits quantitatifs (morphologiques et phénologiques) et les variations moléculaires (microsatellites nucléaires et chloroplastiques et deux gènes candidats *PHANTASTICA* et *FRIGIDA*) dans des populations naturelles de deux espèces de frênes (*Fraxinus excelsior* L., *F. angustifolia* Valh.) le long d'un gradient altitudinal situé dans les Pyrénées orientales Pour notre échantillonnage de huit

populations, le long de ce gradient altitudinal, où des différences dans la morphologie foliaire ont été mesurées, où des décalages phénologiques induisent des barrières de flux de gènes, nos marqueurs nucléaires de type microsatellites semblent montrer une différenciation génétique récente entre les espèces, là où le seul gène candidat obtenu, le gène *PHAN* soit traduit une augmentation récente de taille des populations, soit montre une sélection positive. Ces hypothèses demandent à être confirmées et départagées par des études plus complètes sur l'ensemble du génome à l'aide d'autres marqueurs moléculaires.

***Fraxinus* / gène *PHANTASTICA* / microsatellites nucléaires / gradient altitudinal / adaptation locale**

1. INTRODUCTION

Deux espèces du genre *Fraxinus* (frênes, Oleaceae) sont rencontrées en France à l'état naturel : *Fraxinus excelsior* L. (frêne commun) et *Fraxinus angustifolia* Valh. (frêne oxyphylle). Ces deux espèces, autochtones, ont des aires de distribution séparées mais sont réellement en sympatrie dans certaines régions, notamment dans les Vallées de la Loire, de la Saône, régions qui sont en limite d'aires pour le frêne oxyphylle. La combinaison entre des analyses génétiques (marqueurs neutres de type microsatellites) et morphologiques réalisées à l'échelle de la France [2] montrent pour ces zones d'hybridation des structures très différentes. Les Fst entre ces populations ne sont pas différents indiquant que les marqueurs neutres ne rendent pas compte de l'adaptation locale.

Un aspect important en génétique des populations est de comprendre à l'échelle nucléotidique, la base génétique de la variation adaptative gouvernée par la sélection. Le jeu entre l'histoire des populations, le système de reproduction et la sélection détermine la nature des variations et délimite les mécanismes qui permettent la diversification. C'est une tâche difficile dans le sens où les différents facteurs évolutifs sont étroitement intriqués.

L'approche gènes candidats s'attache à choisir des gènes qui ont des effets majeurs sur le phénotype. La variation allélique du gène est définie par des haplotypes présentant un polymorphisme nucléotidique (SNPs). Bien que limitée à quelques gènes, ce type d'approche a été déjà testée sur des plantes modèles, *Arabidopsis* [12] et *Pinus* [3] et a permis d'appréhender l'adaptation à l'échelle locale.

Nous avons souhaité aborder ce type d'approche pour nos deux espèces de frênes. L'étude des variations adaptatives en populations naturelles étant plus aisée dans des régions de transition clinale forte entre des populations des deux espèces, par exemple dans des régions où l'élévation d'altitude est rapide, nous avons réalisé un transect en Pyrénées Orientales, où les deux

espèces semblent séparées par une étroite zone de transition, *F. excelsior* étant seul à partir de 300m d'altitude et *F. angustifolia* plutôt en bord de mer. La question que nous nous sommes posés était la suivante : est-ce que les patterns de variations nucléotidiques observés sur des gènes potentiellement candidats associés à ceux de marqueurs neutres mesurés sur les mêmes populations indiquent une sélection locale ?

Pour cela, nous avons choisi de travailler sur deux gènes candidats particuliers trouvés dans la littérature. Nous avons tout d'abord étudié un gène impliqué dans les variations des temps de floraison, appelé *Frigida*. Ce locus agit de façon synergique avec le locus *Flowering C* (FLC) qui cause une floraison tardive [16]. Des traces de sélection ont pu être détectées sur un ensemble de séquences [12], et la variation à ce locus a pu être étudiée en populations naturelles d'*A. thaliana* afin d'appréhender son potentiel adaptatif [12]. Ensuite, nous avons choisi un autre candidat intéressant, le gène *PHANTASTICA* (*Phan*), codant pour un facteur de transcription impliqué dans la forme des feuilles d'un grand nombre d'espèces, dont *Fraxinus americana* : par exemple, sa sous-expression permet de produire des feuilles palmées (type cannabis) au lieu de feuilles pennées (type frêne) [11]. Si l'expression de ce gène peut être impliquée dans le nombre de folioles formées, celui-ci peut être également un bon candidat pour l'étude de la variation adaptative chez les frênes et la recherche de traces de sélection au sein de la séquence promotrice.

2. MATÉRIEL ET MÉTHODES

2.1. Matériel

Nous avons choisi de travailler sur un cline altitudinal situé dans les Pyrénées orientales à proximité de Perpignan dans la vallée du Tech (depuis la mer, population Mas Larrieu 42°35' N, 3°01' E ; en altitude population La Preste 42°24' N, 2°24'E). Ce transect part de la mer et monte jusqu'à 1 100 mètres (fig. 1). Huit populations (dont une nommée Arles, scindée en deux) dont 3 de frênes oxyphylles (Mas Larrieu, Ortaffa, Nidolerrès, à faible altitude), 4 de frênes communs (La Forge, Arles 2, Parcigoule, La Preste, à une altitude élevée) et 2 d'hybrides putatifs (St Jean et Arles 1, altitude intermédiaire) ont été échantillonnées. Entre 300 et 800 m, nous n'avons pas trouvé de populations de frênes.

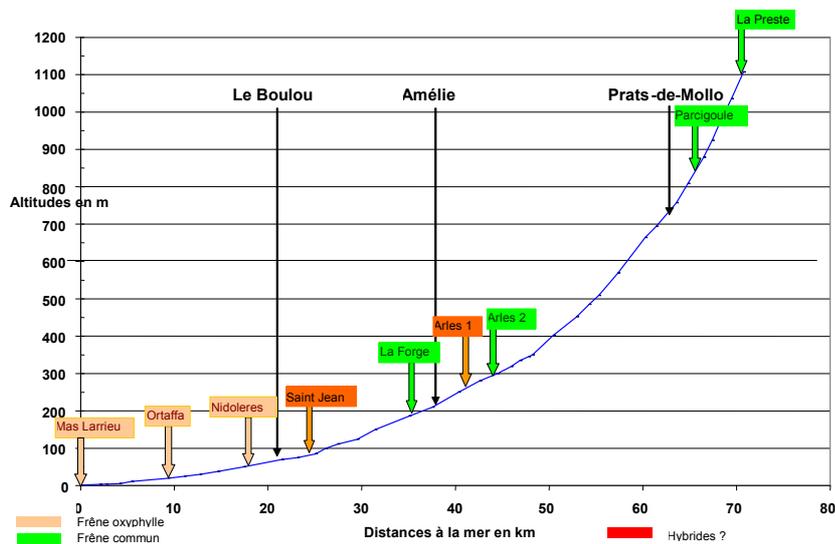


Figure 1 : Transect altitudinal des 8 populations de frêne oxyphyllé, d'hybrides putatifs et de frêne commun situé dans les Pyrénées orientales.

Sur 24 arbres cartographiés par populations, nous avons prélevé des feuilles à l'échenilloir pour l'analyse moléculaire et pour les observations morphologiques. Pour les mesures morphologiques, sur chaque arbre, nous avons mesuré le nombre maximum et minimum de folioles par feuille puis, sur une feuille, la longueur et la largeur de la feuille, puis sur chaque foliole avant terminal, la longueur, la largeur, la position de l'axe de symétrie et le nombre de dents. Par ailleurs, trois années de suite (2005, 2006, 2007) entre décembre et mars, nous avons suivi sur place, le long du transect, la floraison des deux espèces. Pour cela, nous avons observé à plusieurs reprises le stade de débourrement des fleurs selon une codification allant de 0 à 5 (de 0, pas de débourrement à 5, fleur fanée).

2.2. Analyses moléculaires

2.2.1. Marqueurs microsatellites nucléaires et chloroplastiques

L'ADN total a été extrait selon Fernandez-Manjarrès *et al.*, [2]. Nous avons choisi de travailler sur quatre marqueurs microsatellites nucléaires (M230, [1]), 1999 ; FEMSALT 4, FEMSALT 10, FEMSALT 12 [13] qui ont montré leur pertinence dans des publications précédentes [7], [8], [14]. Six marqueurs microsatellites chloroplastiques répertoriés dans deux publications ont été testés (ccmp 6, ccmp 7, ccmp 10, [14] et CPFRA2, CPFRA5, CPFRA6, [6]).

Les protocoles moléculaires de PCR suivent Morand *et al.*, [15]. La taille allélique a été quantifiée grâce à un séquenceur automatique 3700 ABI Prism et analysée grâce au logiciel Genescan (PE Applied Biosystems).

2.2.2. Gène *FRIGIDA*

Il n'existe que deux séquences d'ADNc de *FRIGIDA* sous Genebank ; l'une est sur *Arabidopsis* et l'autre sur *Oryza*. Nous avons dessiné une série d'amorces et avons testé différents couples sur différentes populations de nos deux espèces. Les températures d'hybridation des PCR étaient élevées pour une meilleure spécificité ($T = 60^{\circ}\text{C}$). Le séquençage a été réalisé sur un échantillon d'individus au préalable. L'alignement des séquences a été réalisé par CodonCode Aligner (www.codoncode.com) et par BioEdit (www.mbio.ncsu.edu/BioEdit/bioedit.html).

2.2.3. Gène *PHANTASTICA*

Nous avons recherché ici le polymorphisme dans les populations de frênes des deux espèces. L'expérience était ici plus simple puisqu'il existait en banque la séquence de ce gène sur *Fraxinus americana*. Nous avons dessiné des amorces et avons testé ce gène en populations. Nous avons séquencé pour les 24 individus des 8 populations, le gène dans les deux sens (384 séquences dans les deux sens). L'alignement des séquences a été réalisé par CodonCode Aligner (www.codoncode.com) et par BioEdit (www.mbio.ncsu.edu/BioEdit/bioedit.html).

2.3. Analyses statistiques

Pour les données morphologiques, nous avons utilisé une Analyse Canonique Discriminante (ACD) en utilisant PROC DISCRIM (SAS V8, SAS Institute). Par ailleurs, pour déterminer s'il existait un lien entre nos variables morphologiques et divers paramètres écologiques, dont l'altitude, nous avons procédé à une analyse de corrélation canonique en utilisant PROC CANCELL (SAS V8).

2.4. Analyses génétiques

Le nombre d'allèles A , le nombre d'allèles efficaces A_e , l'hétérozygotie observée H_0 et espérée H_E ainsi que le paramètre de différenciation génétique F_{is} ainsi que le F_{st} ont été estimés en utilisant le logiciel Geneticstudio (<http://www.umsl.edu/~biology/Dyer/GeneticStudio/>). Pour tester la neutralité du gène *PHAN* nous avons effectué le test statistique de Tajima [17] estimé selon MEGA4 [18]. Toutes les positions contenant des gaps et des données manquantes ont été éliminées de la base de données (Option de délétion complète).

3. RÉSULTATS

3.1. Analyses morphologiques et phénologiques

L'analyse canonique discriminante montre que les deux espèces diffèrent morphologiquement comme l'article de Fernandez-Manjarrès *et al.*, [2] le suggérait déjà. Les variables les plus significatives sont le nombre de folioles et la surface foliaire. L'analyse morphologique nous révèle bien deux groupes distincts de populations correspondant aux deux espèces (fig. 2).

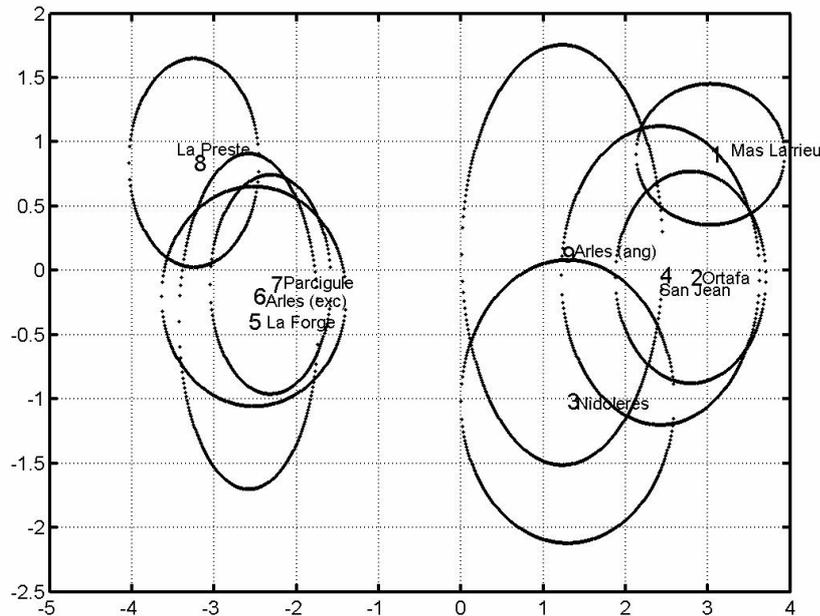


Figure 2 : Analyse canonique discriminante sur composantes principales de variables morphologiques foliaires pour les 8 populations du transect.

La morphologie de la population St Jean est bien oxyphylle et non hybride. La morphologie pour la population d'Arles 1, laisse plutôt à penser à un mélange d'individus des deux espèces, bien distincts morphologiquement et non à des hybrides. L'analyse de corrélation canonique (fig. 3) a été effectuée notamment en prenant en compte l'altitude et la surface foliaire. Elle est très significative ($F_{2,189} = 61,02$ $P < 0,0001$, $R^2 = 0,3923$, fig. 3). A 200 m, les deux espèces sont en mélange avec des morphologies très différentes (exemple d'Arles 1). Jusqu'à 800 mètres la surface foliaire augmente avec l'altitude. Ensuite les valeurs de surfaces foliaires sont moins fortes mais toujours plus élevées pour le frêne commun comparé au frêne oxyphylle.

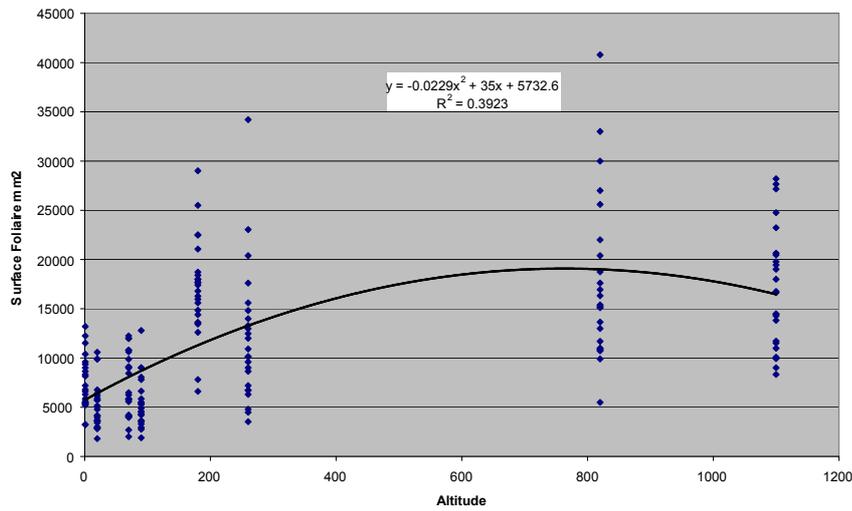


Figure 3 : Analyse de corrélation canonique de la surface foliaire des feuilles de frêne en fonction de l'altitude le long du transect altitudinal.

Les mesures phénologiques sont elles aussi très tranchées. Les populations de frêne oxyphylle fleurissent chaque année entre janvier et février alors que celles de frêne commun fleurissent entre mars et avril. Sur les 3 années de mesure nous n'avons jamais noté de recouvrement ce qui suggère que l'hybridation est un phénomène rare sur ce transect, ce dernier délimitant bien les échanges de gènes entre les deux espèces.

3.2. 3.2. Analyses moléculaires

3.2.1. Marqueurs microsatellites nucléaires et chloroplastiques

Les microsatellites chloroplastiques testés ne sont pas polymorphes et ne nous renseignent malheureusement pas, ni à l'échelle interspécifique, ni à l'échelle intraspécifique. Les microsatellites nucléaires indiquent une forte diversité allélique globale mais avec une tendance à la diminution en altitude (tabl. I). Les valeurs du Fis sont élevées comme cela a déjà été noté chez le frêne et plusieurs fois explicité [14], [4]. Les valeurs de différenciation Fst sont modérées et montrent une plus grande structuration des populations quand considérées pour chacune des espèces (le Fst moyen des populations par rapport à l'ensemble est de 0,027 et le Fst moyen des populations prises pour chaque espèce est de 0,0325).

Tableau I : Diversité génétique et déviation des proportions d'Hardy-Weinberg pour les 4 loci microsatellites nucléaires sur le transect des Pyrénées. A, nombre d'allèles ; A_e, nombre d'allèles efficaces ; H_o, hétérozygotie observée ; H_e, hétérozygotie espérée ; F_{IS}, coefficient de consanguinité intra-population.

Pop	Altitude (m)	A	Ae	Ho	He	Fis
1	1	14,25	6,5734	0,6924	0,8537	0,1848
2	20	16,75	10,9445	0,5173	0,9402	0,4486
3	70	14,5	7,7437	0,5344	0,8644	0,3791
4	90	12,5	5,9197	0,5299	0,8335	0,3602
5	180	11,25	5,2867	0,5945	0,8392	0,29
6	260	14	8,8332	0,5363	0,9203	0,416
7	820	14,5	7,4509	0,4768	0,8684	0,4552
8	1100	11	5,8231	0,4519	0,8159	0,4488

3.2.2. Gène *FRIGIDA*

Nous avons choisi de travailler sur un gène impliqué dans les variations de temps de floraison appelé *FRIGIDA*. Chez l'*Arabidopsis*, la floraison tardive est un état ancien qui était bien adapté au moment des glaciations et la floraison précoce aurait évolué récemment suite aux glaciations par sélection et mutation sur le locus *FRIGIDA*. L'examen attentif de ce gène qui code pour une protéine 609 AA montre que chez les plantes précoces, il aurait perdu sa fonctionnalité. L'inactivation de FRI s'est faite par une délétion ponctuelle qui s'avère suffisante pour faire passer une plante d'une écologie à une autre. Dans notre cas, nous nous sommes interrogés sur ce gène et sur sa pertinence dans notre cadre d'étude avec les deux espèces de frênes étudiées. Nous avons obtenu une séquence de plus de 800 paires de bases relativement identique d'une population à l'autre et d'une espèce à l'autre mais malheureusement n'ayant aucun équivalent en banque actuellement. Nous nous interrogeons sur cette séquence. Dans la mesure où ce gène représenterait la séquence de *FRIGIDA*, est-ce un intron ?

3.2.3. Gène *PHANTASTICA (PHAN)*

Nous avons obtenu des séquences de 602 pb correspondant à l'exon identifié sur *Fraxinus americana* incluant une portion du facteur MYB 2 du gène. Sur un premier examen de 103 séquences à partir d'un échantillon de chacune des 8 populations, nous avons repéré 57 sites variables sur les 392 pb lisibles dont notamment chez les frênes oxyphylles (population Ortafa) une insertion CACCCA. Le test de Tajima effectué (tabl. II) donne une valeur négative de -1,52 (15 substitutions de type non synonymes contre 42 synonymes). La valeur du test de Tajima est de -1,75 pour *Fraxinus excelsior* et de -0,3527 pour *F. angustifolia*. Ces valeurs peuvent intégrer des erreurs de séquençage qu'il nous faudra estimer. Elles indiquent néanmoins, soit une

expansion récente des populations, soit une marque pour ce gène de sélection de type purificatrice. Ceci n'est valable que pour ce gène et demande une certaine prudence quant à une interprétation plus générale.

Tableau II : Le test statistique D de Tajima [17].

Nombre de sites m	Nombre de sites ségrégeants s	$P_s = s/m$	$\Phi = P_s / a1$ (allèle $a1$)	Diversité nucléotidique π	D
103	57	0,146530	0,028140	0,014754	-1,519716

4. DISCUSSION

L'examen des données morphologiques montre une nette différenciation le long du cline altitudinal pour la morphologie des feuilles des deux espèces de frênes. Le nombre de folioles et la surface foliaire varient en fonction de l'altitude. Ce que nous ne savons pas est la part de plasticité phénotypique qu'il nous faudrait contrôler en jardin commun. Les données phénologiques montrent aussi un décalage net sur trois années entre les deux espèces qui interdit les flux de gènes entre elles.

L'analyse sur les marqueurs microsatellites nucléaires montre une diversité allélique plus faible en altitude et un coefficient de consanguinité élevé probablement dû en partie (la présence d'allèles nuls étant aussi probable) à un effet Walhund temporel comme nous l'avions suggéré par ailleurs [4]. La différenciation entre les populations est modérée mais plus forte entre les populations des deux espèces. On est bien dans un cadre d'évolution locale d'une différenciation génétique entre les deux types de frênes probablement récente. Quatre autres marqueurs microsatellites nucléaires sont à l'étude et viendront confirmer ou non ces premiers résultats.

La formation des organes végétaux dépend de la fonction d'auto-organisation des groupes de cellules non différenciées appelées le méristème, situé au sommet de la pousse et de la racine. Lors du développement de la pousse post-embryonnaire, le méristème apical de la pousse (MAP) produit continuellement des organes : tiges, feuilles et fleurs sur une base régulière. Les variations subtiles dans le primordium initié par le MAP sont donc responsables pour des différences que nous observons dans la nature. Des analyses de mutants chez l'*Arabidopsis* ont permis d'identifier les facteurs génétiques impliqués au niveau du MAP. Les gènes à homéoboîte (*KNOX*) de type « *KNOTTED* » de la classe I spécifiques aux plantes sont impliqués dans la régulation de la formation des MAP, d'autres contrôlent l'expression de ces gènes à homéoboîte [5]. Des études génétiques ont pu identifier plusieurs régulateurs négatifs qui modifient les phénotypes par une

expression erronée de gènes *KNOX*. Par exemple, le gène *PHAN* inhibe l'expression du gène *KNOX* car il codifie pour une protéine de domaine MYB qui est exprimée dans les primordia de la feuille et non pas dans le MAP, régulant ainsi de façon négative son expression. L'examen de l'expression de ce gène chez de nombreuses espèces végétales dont *Fraxinus americana* suggère une corrélation entre son expression et le développement des feuilles composées *versus* simples [11].

Dans notre étude, sur la base des séquences analysées et hors les erreurs de séquençage, le gène *PHAN* traduit une augmentation récente de taille des populations et/ou est soumis à une sélection de type purificatrice en raison de sa fonction probablement vitale pour l'arbre. Nous ne pouvons trancher pour l'une ou l'autre des hypothèses à la lumière des séquences analysées. L'expansion récente des populations est une possibilité non négligeable à considérer. Par ailleurs, le constat de sélection purificatrice est aussi un indicateur de la conservation de la séquence avec un taux de substitutions non-synonymes faible comparé à celui de substitutions synonymes. On est dans le cadre d'une sélection positive qui exprime la fixation d'une mutation favorable et qui, pour être ainsi détectée, est peut-être le signe d'un événement de sélection important récemment survenu. Les plantes domestiquées sont, par exemple, de bons modèles pour détecter des phénomènes de sélection positive car elles ont subi récemment des événements drastiques de sélection à de nombreux loci [10].

Pour notre échantillonnage, le long de ce gradient altitudinal où des différences dans la morphologie foliaire ont été mesurées, où des décalages phénologiques induisent des barrières de flux de gènes, nos marqueurs nucléaires de type microsatellites semblent montrer une différenciation génétique récente entre les espèces, là où le seul gène candidat obtenu, le gène *PHAN*, soit traduit une augmentation récente de taille des populations, soit montre une sélection positive. Ces hypothèses demandent à être confirmées et départagées par des études plus complètes sur l'ensemble du génome à l'aide d'autres marqueurs moléculaires. Une banque d'EST sera bientôt disponible pour le frêne et devrait permettre de mesurer de façon plus large l'effet de la sélection sur ces gradients.

REMERCIEMENTS

Les auteurs remercient les lecteurs anonymes qui ont corrigé l'article. Ce travail a été supporté par le « Consortium National de la Recherche en Génomique » et le service de « systématique moléculaire du Muséum National d'Histoire Naturelle (IFR 101). Il s'agit d'un agrément n°2005/67

entre le Géoscope et le MNHN (projet Macrophylogeny of life dirigé par Guillaume Lecointre) dans lequel nous sommes associés.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Brachet S, Jubier MF, Richard M, Jung-Muller B, Frascaria-Lacoste N., 1999 - Rapid identification of microsatellite loci using 5' anchored PCR in the common ash *Fraxinus excelsior*. Mol Ecol, 8: 160-163.
- [2] Fernandez-Manjarrès JF, Gérard PR, Dufour J, Raquin C, Frascaria-Lacoste N., 2006 - Differential patterns of morphological and molecular hybridization between *Fraxinus excelsior* L. and *F. angustifolia* Vahl. (Oleaceae) in Eastern and Western France. Mol Ecol, 15:3245-3257.
- [3] Garcia-Gil M.R., Mikkonen M., Savolainen O., 2003 – Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol. 12: 1195-1206.
- [4] Gérard PR, Fernandez-M. JF, Frascaria-Lacoste N. 2006 - Temporal cline in a hybrid zone population between *Fraxinus excelsior* L. and *F. angustifolia* Valh. Mol Ecol 15 : 3655-3667.
- [5] Hake, S., Smith, H.M.S., Holtan, H., Magnani, E., Mele, G., et Ramirez, J., 2004 - The role of KNOX genes in plant development. Annu Rev. Cell Dev. Biol., 20: 125-151.
- [6] Harbourne M.E., Douglas G.C., Waldren S. & Hodkinson T.K., 2005 - Characterization and primer development for amplification of chloroplast microsatellite regions in *Fraxinus excelsior*, J. Plant Res., 118 : 339-341.
- [7] Heuertz M, Hausman JF, Tsvetkov I, Frascaria-Lacoste N, Vekemans X., 2001 - Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). Mol Ecol, 10, 1615–1623.
- [8] Heuertz M., Hausman J.F., Hardy O.J., Vendramin G.G., Frascaria-Lacoste N., Vekemans X. 2004 - Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southeastern European populations of the common ash (*Fraxinus excelsior* L.). Evolution, 58, 976–988.
- [9] Heuertz M., Fineschi S., Anzidei M., Pastorelli, D. Salvini, L. Paule, Frascaria-Lacoste N., Hardy O.J., Vekemans X., Vendramin G.G. 2004 - Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. Mol Ecol, 13, 3437–3452.
- [10] Innan H., Kim Y. 2004 – Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad.Sci. USA 101: 10667-10672
- [11] Kim M., McCormick S., Timmermans M., Sinha N. 2003 - The expression domain of *Phantastica* determines leaflet placement in compound leaves Nature 424 : 438-443.
- [12] Le Corre V., Roux F., Rebboux X., 2002 - DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time Mol. Biol. Evol. 19 (8): 1261-1271.
- [13] Lefort F., Brachet S., Frascaria-Lacoste N., Edwards K.J., Douglas G.C., 1999 - Identification and characterization of microsatellite loci in ash (*Fraxinus excelsior*

- L.) and their conservation in the olive family (Oleaceae), *Mol. Ecol.* 8: 1075-1092.
- [14] Picard J.-F., 1983 - A propos du frêne oxyphylle, *Fraxinus angustifolia* Vahl. *Forêt-Entreprise* 83:2-4.
- [15] Morand, M.-E., Brachet, S., Rossignol P., Dufour J., Frascaria-Lacoste N. 2002 - A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Mol. Ecol.* 11:377-385.
- [16] Sheldon C.C., Burn J.E., Perez P.P., Metzger J.A., Edwards W.J., Peacock W.J., Dennis E.S., 1999 - The FLF MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* 11: 445-458
- [17] Tajima F.1989 - Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.
- [18] Tamura K, Dudley J., Nei M. & Kumar S. 2007 - MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 10.1093/molbev/msm092.

Analyse et exploitation de la diversité génétique des polykétides synthases de type I dans l'ADN metagénomique d'un sol

Aurélien GINOLHAC⁽¹⁾, Sandrine DEMANECHÉ⁽²⁾, Cyrille JARRIN⁽¹⁾,
Patrick ROBE⁽¹⁾, Fabrice LEFÈVRE⁽¹⁾, Renaud NALIN⁽¹⁾,
Guy PERRIÈRE⁽³⁾, Timothy M. VOGEL⁽²⁾ et Pascal SIMONET^{(2)*}

⁽¹⁾LibraGen S.A., Bâtiment Canal Biotech 1, 3 rue des Satellites,
31400 Toulouse, France

⁽²⁾Equipe « Génomique Microbienne Environnementale », Laboratoire Ampère,
UMR CNRS 5005, Ecole Centrale de Lyon, 36 avenue Guy de Collongue,
69134 Ecully Cedex, France

⁽³⁾Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Université
Claude Bernard – Lyon 1, 43 bd du 11 novembre 1918,
69622 Villeurbanne Cedex, France

Abstract: Analysis and exploitation of type I polyketide synthase genetic diversity in soil metagenomic DNA. Type I polyketide synthases (PKSI) are modular enzymes involved in the biosynthesis of many natural products of industrial interest. PKSI modules are minimally organized in three domains: ketosynthase (KS), acyl-transferase (AT), and acyl carrier protein. The KS and AT domain phylogeny of 23 PKSI clusters available in data bases was determined leading to reveal that many horizontal transfers of PKSI genes have occurred mainly between actinomycetales species. Such gene transfers may explain the homogeneity and the robustness of the actinomycetales group since gene transfers between closely related species could mimic patterns generated by vertical inheritance. The linearity and instability of actinomycetales chromosomes associated with their large quantity of genetic mobile elements could have favored such horizontal gene transfers. In addition a metagenomic approach that provides direct access to diverse unexplored genomes, especially from uncultivated bacteria was developed to detect new biosynthetic pathways among soil bacteria. The ketosynthase domain (KS) of PKSI was used to screen a large soil metagenomic library containing more than 100,000 clones to detect those containing PKS genes. Over the 60,000 clones that were screened, 139 clones containing KS domains were detected. A 700-bp fragment of the KS domain was sequenced for 40 of 139 randomly chosen clones. None of the 39 protein sequences were identical to those found in public databases, and nucleic sequences were not redundant. Phylogenetic analyses were performed on the protein sequences of three metagenomic clones to select the clones which one can predict to produce new compounds. Two PKS-positive clones do not belong to any of the 23 published PKSI included in the analy-

* Correspondance et tirés à part : pascal.simonet@ec-lyon.fr

sis, encouraging further analyses on these two clones identified by the selection process.

PKSI/ metagenomics/ horizontal gene transfer/ phylogenetic analysis/ soil

Résumé : Une banque d'ADN metagénomique de sol a été réalisée et criblée pour la présence de gènes PKSI. Plus d'une centaine de clones sur les 60 000 criblés ont donné un signal positif. Une analyse plus approfondie limitée à 39 de ces clones a permis de montrer l'intérêt de l'approche metagénomique car toutes les séquences des inserts étaient différentes les unes des autres (aucune redondance) et présentaient un niveau de similarité très bas avec les séquences des gènes de voies de biosynthèse connues. Une analyse phylogénétique basée sur les domaines KS et AT a été réalisée en intégrant trois de ces séquences issues du metagénome qui ont été positionnées dans les arbres réalisés avec les séquences disponibles dans les bases de données. Nous montrons que des analyses limitées aux seules séquences d'ADN permettent de prédire la nouveauté des produits synthétisés et les substrats incorporés et d'indiquer au moins sommairement la structure chimique du polykétide qui pourrait être synthétisé si ces gènes étaient exprimés.

PKSI/ metagénomique/ transfert de gènes/ analyse phylogénétique/ sol

1. INTRODUCTION

L'incapacité que présentent plus de 99 % des bactéries du sol à se multiplier *in vitro* et développer des colonies sur des milieux de culture a sans conteste fortement pénalisé l'étude du monde microbien [1]. La microbiologie est ainsi devenue la science des paradoxes : d'un côté des microorganismes modèles comme *Escherichia coli* ou *Bacillus subtilis* qui sont certainement les organismes les plus étudiés et les mieux connus, le génome de plusieurs isolats de ces bactéries a été séquencé et des centaines d'autres micro-organismes bénéficient maintenant des facilités offertes par les nouvelles approches de séquençage à haut débit. Mais de l'autre côté, la très grande majorité des bactéries de l'environnement et en particulier du sol, demeurent totalement inconnues sans que l'on puisse même proposer l'ordre de grandeur du nombre d'espèces bactériennes présentes dans ces écosystèmes [21], [6], [16]. L'exploitation du monde microbien relève des mêmes incongruités, avec par exemple une pénurie à venir des antibiotiques, dont plus de 70 % ont une origine bactérienne, alors que le réservoir de nouvelles molécules que représente la fraction non cultivée des bactéries du sol est quasiment inexploré. Très certainement les efforts actuels dérivant des avancées en génomique bactérienne vont permettre d'accroître le nombre de bactéries cultivées mais il est clair qu'ils ne suffiront pas à court terme pour dévoiler et exploiter une partie significative de la diversité bactérienne.

Une parade à ces limitations, qui pénalisent les domaines fondamentaux de la recherche mais aussi leurs applications industrielles réside dans l'approche metagénomique qui consiste à extraire l'ADN bactérien directement de l'environnement, le cloner dans les vecteurs appropriés et le multiplier dans des hôtes bactériens domestiqués tant pour étudier que pour exploiter les fonctions des gènes ainsi récupérés [15]. Au cours des années 90 de très nombreux travaux ont consisté à optimiser les méthodes visant à extraire et purifier l'ADN bactérien directement de l'environnement, avec comme objectifs de substituer le moins possible les biais liés à l'isolement par ceux relevant de l'extraction des acides nucléiques. Pour des environnements comme le sol et les sédiments, ces défis étaient particulièrement ardues vu la complexité physico-chimique de ces matrices. Deux types de méthodes ont été développés avec soit une lyse bactérienne réalisée *in situ* avant l'extraction et la purification de l'ADN, soit que ces étapes aient été précédées de l'extraction des cellules bactériennes et leur purification des autres constituants du sol [4]. La méthode de lyse *in situ* peut s'appliquer en routine pour le traitement de nombreux échantillons mais présente l'inconvénient de ne pas sélectionner l'ADN procaryotique de celui des micro-eucaryotes, ni de différencier ADN intra et extracellulaire. De plus les traitements lytiques employés sont généralement assez délétères pour l'ADN, dont la taille après purification ne peut guère dépasser 100 kb. Au contraire, la stratégie qui consiste à extraire les cellules bactériennes de la matrice tellurique avant leur traitement est beaucoup plus lourde à mettre en œuvre mais produit un ADN de plus grande pureté, avec une majorité de fragments d'ADN d'origine procaryotique d'une taille de plusieurs centaines de kb [3]. C'est cette dernière stratégie qui est privilégiée pour la constitution de banques d'ADN metagénomique, la qualité de l'ADN extrait permettant le clonage de fragments d'ADN de grande taille, tant pour accroître la représentativité de la diversité bactérienne que pour permettre l'expression de voies de biosynthèse à partir d'opérons complexes [3].

Ces approches metagénomiques ont permis de confirmer le très important niveau de la diversité bactérienne dans différents environnements [2], [12], [19]. Elles permettent également d'aborder d'autres questions concernant le potentiel adaptatif et évolutif bactérien et en particulier de prendre en compte les transferts horizontaux de gènes. Les travaux présentés ici ciblent les gènes qui codent les enzymes polyketide synthases de type I, qui sont des enzymes multi-domaines modulaires impliquées dans la synthèse de nombreux produits naturels d'intérêt industriel [18], [22]. Les modules d'extension des PKS sont organisés en au moins 3 domaines, une kétosynthase (KS) pour la condensation décarboxylative de l'unité d'extension sur la chaîne en formation, une acyltransférase (AT) qui sélectionne le substrat à charger et qui le transfère et une protéine porteuse qui charge la chaîne en formation. Les objectifs ont été d'étudier

l'implication des transferts horizontaux de gènes dans le modelage des voies de biosynthèse de métabolites secondaires chez des isolats connus ainsi que dans des inserts de clones d'ADN metagénomique d'un sol après criblage moléculaire d'une banque construite au laboratoire.

2. MATÉRIEL ET MÉTHODES

2.1. Extraction de l'ADN du sol et construction de la banque metagénomique

Toutes les données concernant l'extraction de l'ADN du sol dit de Montrond (50 km à l'est de Lyon) par la méthode de séparation initiale des cellules bactériennes sur gradient de nycodenz sont présentées dans Ginolhac *et al.*, [7]. Les fragments d'ADN de grande taille ont été clonés dans un vecteur fosmidique grâce au kit « EpiFos fosmid library production » utilisé selon les procédures préconisées par le fabricant (Epicentre Technologies). Les 100 000 clones recombinants obtenus ont été transférés individuellement sur plaques de microtitration 96 puits pour croissance, stockage et criblage.

2.2. Criblage moléculaire de la banque d'ADN metagénomique

Deux techniques ont été développées, la première basée sur une approche PCR requérant l'extraction de l'ADN de pools de 96 clones à l'aide du kit Nucleobond PC100 kit (Macherey Nagel) utilisé selon les instructions du fabricant, l'ADN étant utilisé comme matrice dans une réaction PCR avec les amorces KSLF (5'-CCSCAGSAGCGCSTSYTSC TSGA-3') et KSLR (5'-GTSCCSGTSCCGTGSGYSTCSA-3') définies d'après les motifs conservés des domaines KS [7]. Les pools d'ADN présentant une réponse PCR positive ont été traités par hybridation afin de localiser le ou les clones parmi les 96 à l'origine du signal positif. La seconde approche a consisté à déposer un aliquote de chaque suspension bactérienne sur une membrane à l'aide d'un spotter adapté et de traiter les membranes en vue de permettre la croissance bactérienne, la lyse cellulaire, l'adsorption de l'ADN puis leur hybridation avec les sondes choisies.

2.3. Séquençage

Les inserts fosmidiques ont été séquencés en utilisant les approches de marquage par transposition et sous-clonage dans le vecteur pBC SK (+/-) (Stratagene) [7].

2.4. Analyses phylogénétiques

Les séquences protéiques des domaines KS et AT détectées dans les clones de la banque metagénomique ont été alignées avec 23 clusters PKS présents dans les bases de données (disponibles à http://web.libragen.com/Phylogeny/sup_table.html) représentant 203 domaines KS et 207 domaines AT. Les domaines KS et AT des clones metagénomiques ont été inclus et alignés en utilisant le programme DbClustal [20] avec correction manuelle grâce à Seaview [5]. Les reconstructions phylogénétiques ont été réalisées avec Phylo_win [5] pour la méthode de distance avec Neighbor Joining (NJ) et la matrice PAM alors que le programme PhyML [9] a été utilisé pour la méthode du maximum de vraisemblance (ML) en utilisant BIONJ avec le modèle JTT de substitution. Les valeurs de bootstrap ont été déterminées pour tous les arbres (500 répliques), ces 500 valeurs étant générées par SEQBOOT du programme PHYLIP, version 3,57c, package pour la reconstruction ML. Toutes les autres informations sont disponibles dans Ginolhac *et al.* [7] et [8].

2.5. Numéro d'accèsion des séquences nucléotidiques

Les 3 séquences nucléotidiques (Lib4, Lib7 et Lib10) codant les domaines KS et AT ont reçu les numéros d'accèsion AJ639921, AJ639922 et AJ639923 dans la base EMBL.

3. RÉSULTATS ET DISCUSSION

3.1. Analyse bioinformatique des domaines KS des gènes PKS1 des bases de données

La majorité des domaines KS analysés provenant des bases de données forment un groupe phylogénétique très cohérent en accord avec leur origine taxonomique puisque ces séquences proviennent toutes d'actinobactéries. Les autres domaines KS forment un groupe plus lâche et leurs origines sont diverses (Nostocales, Chroococcales, Myxococcales, et Pseudomonadales). Dans la plupart des cas les différents domaines d'une PKS1 donnée forment un groupe monophylétique cohérent en accord avec l'hypothèse de l'évolution des gènes PKS pour laquelle l'organisation modulaire et la diversité chimique des polykétides produits par ces enzymes proviendraient d'une duplication de domaines [10], ce qui fait que la similarité de séquence devrait être plus importante entre les domaines d'une même PKS1 qu'entre ceux de différentes enzymes. L'analyse bioinformatique montre que les PKS1 chez les actinomycétales sont plus homogènes que dans les autres groupes, ce qui peut indiquer que les transferts horizontaux y sont plus

fréquents, l'homogénéité résultant d'un brassage important qui assure ainsi une plus forte cohérence génétique comme l'avaient montré Torsvik *et al.*, [21]. Cependant, nos travaux ont montré plusieurs exceptions à ce postulat, des incongruités phylogénétiques qui révèlent des événements de transfert horizontal. A titre d'exemples on peut citer des domaines KS des rapamycin et ascomycin synthases qui sont séparés de leurs KS d'origine, incongruité qui ne peut s'expliquer que par un transfert horizontal des gènes figure 1 (figure 3 dans Ginolhac *et al.*, [8]). De plus le fait que l'embranchement très cohérent des actinobactéries comporte aussi une séquence d'un domaine KS d'une myxococcale, *Sorangium cellulosum* (bactérie productrice du soraphen) indique que les transferts de gènes peuvent intervenir entre microorganismes phylétiquement très éloignés. Ces analyses phylogénétiques peuvent aussi permettre de proposer des hypothèses évolutives comme par exemple une acquisition récente des PKSII chez les actinomycétales qui expliquerait la cohérence génétique et ceci en dépit de la grande diversité des polykétides produits. Or nous observons que les domaines KS des PKSII d'ordres bactériens non actinobactéries ne sont pas résolus. Cela signifie que l'ancêtre commun des bactéries possédait des PKSII. Cette diversité aurait ensuite été perdue pour les actinobactéries puis recréée secondairement. Cette hypothèse n'est pas parcimonieuse, surtout au vu de la très grande diversité des polykétides produits par les actinobactéries.

Une autre hypothèse stipule une évolution très rapide des gènes des PKSII dans les autres groupes que les actinobactéries, qui auraient alors masqué les traces de duplications des domaines KS d'une même PKSII. L'étude des taux relatifs d'évolution n'a montré aucune différence entre les domaines KS quelle que soit leur origine taxonomique [8]. Ces travaux ont aussi permis de tenter de corréler les nombreux transferts horizontaux observés chez les actinobactéries aux particularités de leurs génomes caractérisés par leur structure linéaire et une forte instabilité aux extrémités. Les gènes des voies de biosynthèse sont localisés à ces extrémités et sont donc sujets aux événements de duplication, délétions et recircularisation. De même il a été montré que ces gènes peuvent être localisés sur des plasmides linéaires [24] ou entre des séquences d'insertion [11] autant d'éléments en faveur d'événements de transfert de gènes fréquents chez ce groupe de bactéries.

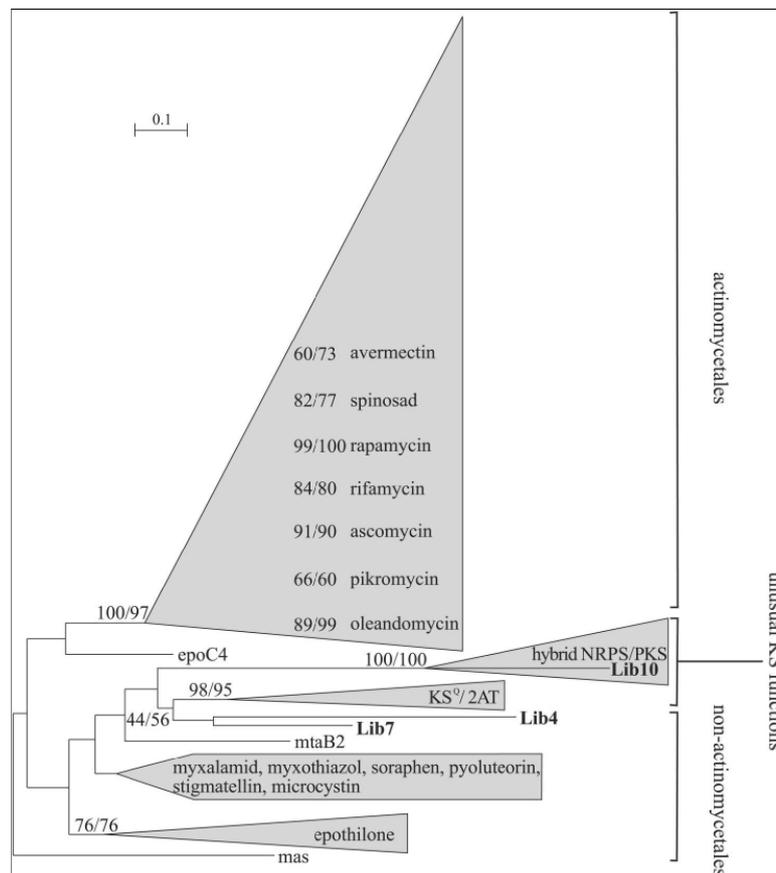


Figure 1 : Analyse phylogénétique des domaines KS basée sur 207 séquences protéiques (voir texte). Le groupe appelé hybrid NRPS/PKS inclut les domaines KS précédés par un NRPS. Les domaines KSQ et les domaines KS des modules de chargement avec l'organisation ACP-KS-AT-AT-ACP forment le groupe appelé KSQ/2AT. Les topologies d'arbres sont identiques par chacune des 2 méthodes utilisées. Les valeurs bootstrap indiquées sur la figure sont celles obtenues avec la méthode ML/NJ. Les domaines KS donnant un groupe très cohérent sont présentés sous le nom du polykétide formé avec leurs valeurs de bootstrap respectives. La figure présente le positionnement des séquences obtenues à partir des clones de la banque d'ADN metagénomique du sol.

3.2. Construction de la banque d'ADN metagénomique de sol

Une banque d'ADN metagénomique du sol de Montrond comprenant plus de 100 000 clones a été obtenue avec une efficacité de clonage de 3×10^5 clones par μg d'ADN, avec une taille moyenne des inserts variant entre 30 et 40 kpb. La taille cumulée des inserts atteint 3,5 Gpb ce qui correspond

approximativement à 850 fois la taille du génome d'*E. coli*. Si l'on considère une faible redondance de l'ADN cloné cette banque d'ADN représente 5 % de la diversité bactérienne du sol estimée sur la base de 10 000 types bactériens différents [21] ce qui donne accès à une diversité bactérienne 24 fois plus importante que par les méthodes de culture traditionnelles.

3.3. Criblage de la banque d'ADN metagénomique

Deux stratégies susceptibles d'être employées pour le criblage de banques d'ADN de cette ampleur ont été développées et sont maintenant opérationnelles. La stratégie combinant le dépôt robotisé sur membranes haute densité d'un inoculum de chaque clone recombinant, en vue d'une hybridation après croissance et traitement des colonies, est la plus prometteuse en termes de coût et d'efficacité. Les résultats présentés dans la figure 2 montrent la spécificité de l'hybridation qui permet de repérer sans difficulté les colonies produisant un signal positif. Chaque clone peut alors être traité indépendamment pour sous-clonage et séquençage des gènes d'intérêt.

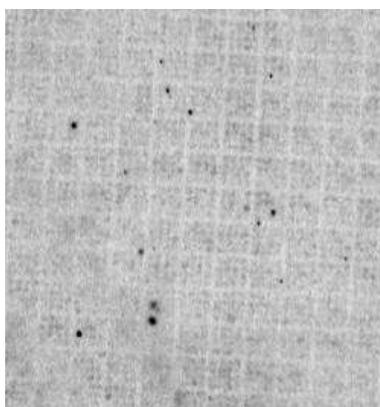


Figure 2 : Portion d'une membrane haute densité réalisée avec des clones d'ADN metagénomique de sol déposés et cultivés sur membranes. L'hybridation permet de révéler les clones dont l'insert présente une similarité de séquence avec la sonde utilisée.

C'est cependant une seconde stratégie basée sur la PCR qui a été utilisée pour cribler 60 000 des 100 000 clones de la banque d'ADN metagénomique. Les amorces utilisées avaient été définies sur les zones les plus conservées des domaines KS et de cette façon 139 clones ont pu être détectés. Dans une première étape 40 fosmidés recombinants ont été choisis au hasard parmi ces 139 clones pour être sous-clonés et séquencés. Les 40 séquences nucléotidiques ne présentaient aucune redondance entre elles et

leur niveau de similarité au niveau protéique avec les séquences disponibles dans les bases de données ne dépassaient pour aucune d'entre elles 67 % (maximum de 141 acides aminés identiques sur 210). Ces résultats confirment l'étendue de la diversité génétique bactérienne et le potentiel de l'approche metagénomique pour atteindre ces gènes bactériens.

3.4. Analyse phylogénétique des gènes PKSI détectés

Un site de domaine KS actif codant un cadre de lecture ouvert a été détecté dans chacune des séquences d'ADN amplifiées à partir du metagénome. L'analyse plus poussée a montré que certains de ces domaines KS appartenaient à des hybrides entre des NRPS (non ribosomal polyketide synthases) et des PKSI. Pour préciser l'analyse de ces gènes, la séquence complète de 3 inserts a été réalisée (Lib4, Lib7 et Lib10), représentatifs du groupe des 11 clones présentant un domaine NRPS en amont du domaine KS. De plus Lib4 et Lib7 ne possédaient pas un gène PKS à l'une ou l'autre des extrémités, argument fort en faveur de la totalité de la voie de biosynthèse sur l'insert.

3.5. Analyse des domaines KS

Un seul domaine KS a été détecté dans les séquences des clones Lib4, Lib7 et Lib10 indiquant que les produits qui pourraient être synthétisés à partir de ces gènes ne possèdent pas une structure typique de polykétides. Les analyses phylogénétiques ont été réalisées pour positionner ces domaines KS issus du metagénome au sein de l'arbre phylogénétique construit à partir des séquences dans les bases de données (voir figure 3). Les séquences des domaines KS des clones Lib4, Lib7 et Lib10 présentaient des valeurs de similarité faibles avec celles présentes dans la base GenBank (respectivement 58, 61, et 58 % d'identité BLASTP avec leurs plus proches voisins). Aucun de ces domaines KS ne se positionnait dans le groupe des actinomycétales suggérant fortement que les bactéries à l'origine de ces séquences n'appartenaient à cet ordre. On ne peut toutefois pas exclure que des contraintes fonctionnelles fortes aient entraîné des gènes à évoluer différemment comme ce peut être le cas des domaines KS catalysant une chaîne acyl initiale qui forment un groupe séparé indépendamment de la position taxonomique de l'organisme [7], [13], [14]. Il est intéressant de constater que ces domaines KS non conventionnels sont systématiquement précédés d'un module NRPS qui est aussi détecté dans la séquence du clone Lib10 qui se positionne dans ce même groupe hybride.

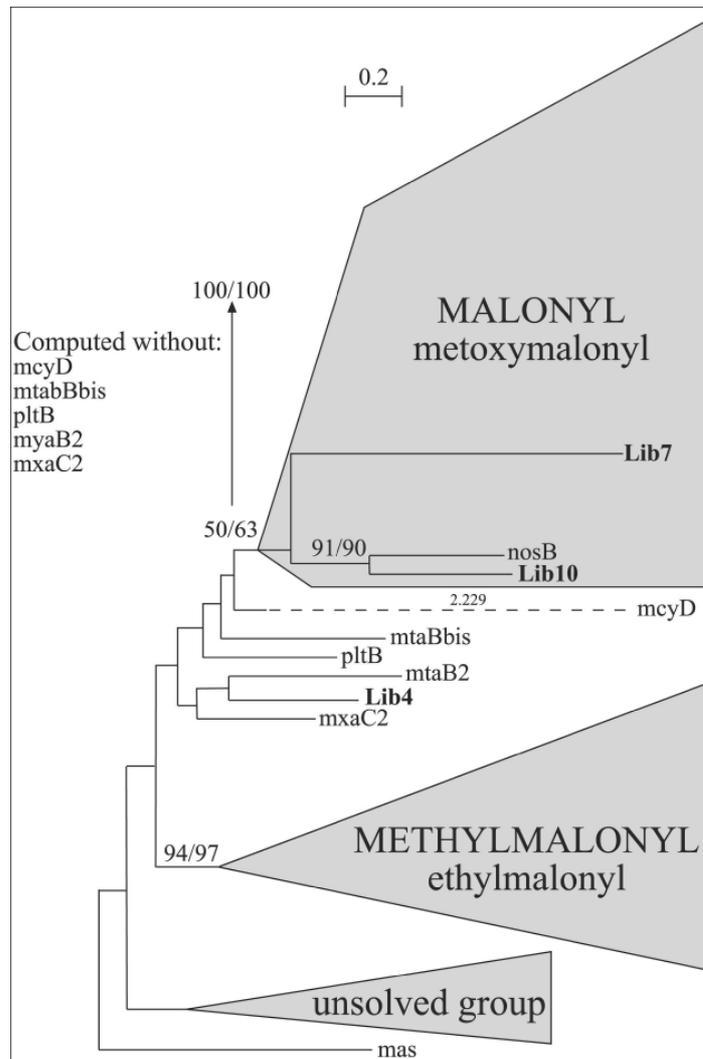


Figure 3 : Analyse phylogénétique des domaines AT basée sur 210 séquences protéiques (voir texte). Les topologies d'arbres sont identiques par chacune des 2 méthodes utilisées. Les valeurs bootstrap indiquées sur la figure sont celle obtenues avec la méthode ML/NJ. Le domaine appelé *mtaBbis* correspond au second domaine AT dans l'organisation ACP-KS-AT-AT-ACP. Les domaines AT situés à l'extérieur des deux groupes malonyl et methylmalonyl n'ont pu être regroupés par l'analyse phylogénétique. La figure présente le positionnement des séquences obtenues à partir des clones de la banque d'ADN metagénomique du sol.

Très clairement les domaines KS des clones Lib4 et Lib7 n'appartiennent pas à une des PKSI de la série utilisée dans cette étude indiquant que les 2

polykétides qui pourraient être synthétisés seraient différents de ceux référencés, tandis que le domaine KS non conventionnel de Lib10 ne permet pas de prédire la nouveauté fonctionnelle du produit potentiellement synthétisé.

3.6. Prédiction de spécificité du substrat des domaines AT

La structure chimique du composé final dépend du substrat incorporé reconnu par les domaines AT, du degré du cycle de réduction catalysé par les domaines additionnels de chaque module et du nombre de modules et de leur succession. La reconnaissance du substrat est un facteur clé qui influence la structure du polykétide. Dans la plupart des cas, l'incorporation des substrats peut être prédite par l'analyse phylogénétique des domaines AT [23] qui montre que l'incorporation du malonyl et du méthylmalonyl différencie 2 groupes distincts. Les 2 domaines AT des clones Lib7 et Lib10 se positionnent dans le groupe malonyl ce qui est un argument fort pour prédire l'incorporation de malonyl par ce domaine PKS. L'étude des signatures des sites actifs, selon l'approche utilisée par Rondon *et al.*, [17], confirme l'incorporation d'un substrat de type malonyl pour le domaine AT du clone Lib7 mais également pour le clone Lib10, même si dans ce dernier cas l'analyse était moins affirmative. Il n'en demeure pas moins que les méthodes de signature et de phylogénie donnent des résultats cohérents concernant l'incorporation de malonyl chez les 2 clones Lib7 et Lib10. En revanche, une prédiction d'incorporation d'un substrat n'est pas possible pour le clone Lib4, la séquence ne pouvant être positionnée ni dans le groupe malonyl ni dans celui du méthylmalonyl. De plus, la méthode basée sur la signature du site actif concerne principalement les voies de biosynthèse des actinobactéries et que, comme l'a déjà montré la phylogénie du domaine KS, l'insert fosmidique serait d'une autre origine taxonomique.

4. CONCLUSION

Par une approche de criblage moléculaire ciblée sur les gènes PKS, nous avons détecté 139 clones d'intérêt dans une banque d'ADN metagénomique de sol de 60 000 clones soit 0,23% des inserts. Cette fréquence de détection, l'absence de redondance et l'originalité des inserts obtenus confirment l'intérêt de cette approche pour un accès à la biodiversité bactérienne inconnue et même son exploitation en termes de synthèse de nouveaux composés. Cette dernière étape est cependant particulièrement complexe, du fait de la nécessité de transférer l'insert d'ADN à un hôte bactérien adapté et de trouver les conditions optimales de l'expression en système hétérologue de ces gènes. Même dans le cas où ces problèmes

d'expression dans un hôte bactérien seraient réglés, l'analyse chimique des composés naturels ainsi synthétisés n'est pas aisée. Nos travaux montrent que l'analyse des séquences des gènes PKSII fournit suffisamment d'informations pour sélectionner les clones présentant le plus d'intérêt, en particulier sur la nouveauté des produits synthétisés et les substrats incorporés qui permettent de prédire, au moins sommairement, la structure chimique du polykétide.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Amann R.I., Ludwig W., Schleifer K.H., 1995. - Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.*, 59:143-169.
- [2] Beja O., Suzuki M.T., Koonin E.V., Aravind L., Hadd A., Nguyen L.P., Villacorta R., Amjadi M., 2000. - Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.*, 2:516-529.
- [3] Bertrand H., Poly F., Van V.T., Lombard N., Nalin R., Vogel T.M., Simonet P., 2005. - High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J. Microbiol. Methods.*, 62:1-11.
- [4] Courtois S., Frostegard A., Goransson P., Depret G., Jeannin P., Simonet P., 2001. - Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ. Microbiol.*, 3:431-439.
- [5] Galtier N., Gouy M., Gautier C., 1996. - SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.*, 12:543-58.
- [6] Gans J., Wolinsky M., Dunbar J., 2005. - Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, 309:1387-1390.
- [7] Ginolhac A., Jarrin C., Gillet B., Robe P., Pujic P., Tiphile K., Bertrand H., Vogel T.M., Perriere G., Simonet P., Nalin R., 2004. - Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.*, 70:5522-5527.
- [8] Ginolhac A., Jarrin C., Robe P., Perriere G., Vogel T.M., Simonet P., Nalin R., 2005. - Type I polyketide synthases may have evolved through horizontal gene transfer. *J. Mol. Evol.*, 60:716-725.
- [9] Guindon S., Gascuel O., 2003. - A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, 52:696-704.
- [10] Hopwood D.A., 1997. - Genetic contributions to understanding polyketide synthases. *Chem. Rev.*, 97:2465-2498.
- [11] Jenkin G.A., Stinear T.P., Johnson P.D., Davies J.K., 2003. - Subtractive hybridization reveals a type I polyketide synthase locus specific to *Mycobacterium ulcerans*. *J. Bacteriol.*, 185:6870-6882.

- [12] Liles M.R., Manske B.F., Bintrim S.B., Handelsman J., Goodman R.M., 2003. - A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl. Environ. Microbiol.*, 69:2684-291.
- [13] Lopez J.V., 2004. - Naturally mosaic operons for secondary metabolite biosynthesis: variability and putative horizontal transfer of discrete catalytic domains of the epothilone polyketide synthase locus. *Mol. Genet. Genomics*, 270:420-431.
- [14] Moffitt M.C., Neilan B.A., 2003. - Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J. Mol. Evol.*, 56:446-457.
- [15] Robe P., Nalin R., Capellano C., Vogel T.A., Simonet P., 2003. - Extraction of DNA from soil. *Eur. J. Soil. Biol.*, 39:183-190.
- [16] Roesch L.F., Fulthorpe R.R., Riva A., Casella G., Hadwin A.K., Kent A.D., Daroub S.H., Camargo F.A., Farmerie W.G., Triplett E.W., 2007. - Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal*, 1:283-290.
- [17] Rondon M.R., August P.R., Bettermann A.D., Brady S.F., Grossman T.H., Liles M.R., Loiacono K.A., Lynch B.A., MacNeil I.A., Minor C., Tiong C.L., Gilman M., Osburne M.S., Clardy J., Handelsman J., Goodman R.M., 2000. - Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.*, 66:2541-2547.
- [18] Shen B., 2003. - Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.*, 7:285-295.
- [19] Suzuki M.T., Beja O., Taylor L.T., Delong E.F., 2001. - Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ. Microbiol.*, 3:323-331.
- [20] Thompson J.D., Plewniak F., Thierry J., Poch O., 2000. - DbClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic. Acids. Res.*, 28:2919-2926.
- [21] Torsvik V., Ovreas L., Thingstad T.F., 2002. - Prokaryotic diversity-magnitude, dynamics, and controlling factors. *Science*, 296:1064-1066.
- [22] Weissman K.J., Leadlay P.F., 2005. - Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.*, 3:925-936.
- [23] Yadav G., Gokhale R.S., Mohanty D., 2003. - Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.*, 328:335-363.
- [24] Yamasaki M., Ikuto Y., Ohira A., Chater K., Kinashi H., 2003. - Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated streptomycetes. *Microbiology*, 149:1351-1356.

Un modèle de variabilité fonctionnelle chez les arbres forestiers : le gène *CCR* d'eucalyptus

Jean-Marc GION^{(1,2)*}, Frédéric MORTIER⁽¹⁾, Eric MANDROU^(1,2,3),
Paulo Ricardo HEIN GHERARDI⁽¹⁾, Tristan COSTECALDE^(1,2),
Gilles CHAIX⁽¹⁾, Marie ÉTIENNE⁽⁴⁾, Pierre SIVADON⁽⁵⁾,
Jacqueline GRIMA-PETTENATI⁽⁵⁾, Emilie VILLAR⁽⁶⁾, Aubin SAYA⁽⁶⁾,
Brigitte POLLET⁽⁷⁾, Catherine LAPIERRE⁽⁷⁾, Philippe VIGNERON⁽¹⁾

⁽¹⁾ CIRAD UPR 39 Génétique Forestière Dpt BIOS Campus International de Baillarguet TA A/39 C, 34398 Montpellier Cedex 5, France

⁽²⁾ UMR BIOGECO 1202 INRA, Equipe de Génétique 69 route d'Arcachon 33612 CESTAS Cedex France

⁽³⁾ Vallourec, 27 avenue du Général Leclerc 92100 Boulogne-Billancourt, France

⁽⁴⁾ ENGREF –UMR 518. GRESE 19 Avenue du Maine 75732 Paris, France

⁽⁵⁾ UMR UPS-CNRS 5546 « Surfaces Cellulaires et Signalisation Végétales » Pôle de Biotechnologie Végétale 24, chemin de Borde Rouge BP 42617, 31320 Auzeville, France

⁽⁶⁾ Ur2pi, BP 1291 Pointe Noire, République du Congo

⁽⁷⁾ UMR Chimie Biologique INRA - AgroParisTech BP1 78850 Thiverval-Grignon, France

Abstract: *CCR* gene in *Eucalyptus* a model of functional variability in forest trees. Nucleotidic polymorphism of Cinnamoyl CoA Reductase (*CCR*) gene and its relation with lignin content is studied within a breeding population of *Eucalyptus urophylla* S.T. Blake ("Timor Mountain Gum"). The nearly full sequence (94%) are obtained for 15 parental trees. This gene (3220 bp) is highly polymorphic showing 131 single nucleotide polymorphism (SNP), 17 insertion-deletions (INDEL), 1 polyA sequence and a microsatellite site. Exons fragments encompass 10 non-synonymous SNPs, half of them within exon 5 (194 bp). Fifteen different haplotypes are reconstructed based on the polymorphism of exon 4 and intron 4. *CCR* promoting sequence (694 bp) including all the known regulatory sequences is described for the two alleles of one of the genitor trees displaying QTL and *CCR* gene colocalization in its genetic map. Five SNPs are presents. Lignin content was assessed within a sample of 35 full sib families (348 individuals) generated with the 15 parental trees, showing a high genetic additive control for this trait ($h^2=0.76$). A new algorithm based on Reversible-jump MCMC was developed in order to implement association studies. Half of the progeny trees (208) were genotyped using the microsatellite fragment. The results show that a significant part of the observed genetic variance of lignin content is due to the nucleotide polymorphism of the studied gene. Those preliminary results are

* Correspondance et tirés à part : gion@cirad.fr

promising for developing early gene assisted selection for eucalyptus clones used as raw material in charcoal and paper production.

CCR gene/ eucalyptus/ quantitative variation/ lignins/ association study

Résumé : La variabilité nucléotidique du gène codant la Cinnamoyl CoA Reductase (*CCR*) et ses effets sur le taux de lignine est étudiée au sein d'une population d'*Eucalyptus urophylla* S.T. Blake. La presque totalité de la séquence (94 %, 3220 paires de bases) est décrite pour 15 individus. Le gène est hautement polymorphe et présente 131 mutations ponctuelles (SNP) ainsi que divers autres types de mutations. Les fragments exoniques présentent 10 SNP non synonymes dont 5 dans l'exon 5. La séquence promotrice (694 pb) est décrite pour les deux allèles d'un des géniteurs. Elle regroupe 5 SNPs. L'analyse de la teneur en lignine de 348 arbres appartenant à 35 familles de pleins frères obtenues avec ces 15 géniteurs montre que ce caractère présente un fort contrôle génétique additif ($h^2=0.76$). Un nouvel algorithme type MCMC a été développé pour procéder aux études d'association sur 208 descendants génotypés grâce à un marqueur microsatellite présent dans le gène *CCR*. Les résultats montrent qu'une part importante de la variance du taux de lignine est due au polymorphisme du gène *CCR*. Ces résultats laissent envisager le développement d'une sélection précoce assistée par marqueurs.

gène CCR/ eucalyptus/ variabilité de caractères quantitatifs/ lignines/ test d'association

1. INTRODUCTION

La conservation et l'exploitation durable des ressources génétiques forestières nécessitent une bonne connaissance de la variabilité naturelle existante. Cette variabilité a souvent été analysée grâce aux caractères phénotypiques eux-mêmes ou à l'aide de marqueurs moléculaires anonymes permettant de décrire la variabilité neutre au sein de populations forestières. L'analyse de la variabilité nucléotidique de gènes d'intérêt reste encore limitée chez les génomes forestiers même si le développement récent de programmes de séquençage fournit une quantité croissante de séquences exprimées [14]. Ceci permet d'envisager une caractérisation de la variabilité fonctionnelle afin d'élaborer des stratégies de conservation et de sélection assistées par marqueurs chez ces espèces peu domestiquées et à long cycle de révolution [22]. Les premiers travaux sur la variabilité de populations forestières grâce à l'analyse de gènes d'intérêt ont été réalisés chez le pin et l'eucalyptus [30], [27]. Ils semblent indiquer une variabilité nucléotidique plus importante ainsi qu'un déséquilibre de liaison intra-séquences plus faible chez ces populations forestières que chez l'homme, la drosophile ou encore le maïs. Une étude sur *E. nitens* indique que le déséquilibre de liaison au sein du gène *CCR* décroît rapidement avec la distance physique et montre

une répartition hétérogène de la variabilité nucléotidique le long du gène [34]. L'analyse de la diversité de gènes pour les espèces forestières concerne, en général, un fragment du gène (EST) et donc une partie limitée de sa variabilité totale. De plus, les régions promotrices, dont la variabilité est susceptible d'affecter l'expression du gène, sont rarement prises en compte [22]. La conservation et l'exploitation de la variabilité fonctionnelle de gènes d'intérêt nécessitent un recensement des formes alléliques existantes et donc l'analyse préalable de séquences les plus complètes possibles.

Les propriétés du bois constituent un caractère majeur pour l'amélioration génétique des espèces forestières et leur adaptation au milieu. Ces propriétés dépendent en grande partie de la lignine, qui est déposée dans les parois secondaires des cellules du xylème. Ainsi, un bois avec une forte teneur en lignine nécessitera un traitement énergétique et chimique onéreux et polluant pour la production de pâte à papier alors qu'il sera adapté à la production d'énergie. La voie de biosynthèse spécifique des lignines est catalysée par deux enzymes, la cinnamoyl CoA reductase (*CCR*) et la cinnamyl alcool dehydrogenase (*CAD*). Elles utilisent comme substrat le cinnamoyl CoA produit par la voie des phénylpropanoïdes pour synthétiser les monolignols, qui sont finalement polymérisés en lignine grâce à un groupe de peroxydases et de laccases. Les connaissances approfondies de cette voie de biosynthèse, la disponibilité de séquences pour les gènes de lignification et l'impact reconnu des lignines sur la qualité du bois et son utilisation, en font un exemple de voie métabolique de choix pour mener à bien une analyse de la variabilité fonctionnelle de gènes candidats. Un des gènes candidats évident pour ce type d'approche est le gène *CCR*. En effet, la modulation de l'expression de ce gène chez des plants de tabacs transgéniques entraîne une modification de la quantité et de la qualité des lignines, avec, par exemple, jusqu'à 50 % de diminution du taux de lignines et un changement dans la composition en unités synapylique et coniférylique lorsque ce gène est réprimé [13]. De plus, ce gène est préférentiellement exprimé dans le xylème d'eucalyptus grâce à la présence de séquences promotrices spécifiques [17], [25]. Enfin, le gène *CCR* a été cartographié chez *E. urophylla* [10] et une colocalisation a été mise en évidence entre ce gène et un QTL de teneur en lignines [11].

Au sein des programmes d'amélioration des arbres, la sélection se fait classiquement par l'utilisation de modèles de génétique quantitative basés sur le phénotype. Ces modèles appartiennent à la classe des modèles à effets aléatoires, où les effets aléatoires expriment les effets génétiques. L'utilisation de marqueurs moléculaires tels que les SNP permet d'avoir une information concernant les allèles des géniteurs et de la descendance sur l'ensemble du génome ou sur des zones d'intérêt comme des gènes candidats. Une question centrale est alors de pouvoir estimer les effets génétiques parentaux mais aussi les associations entre les caractères

phénotypiques et les SNP. L'objectif des études d'association est d'identifier les formes aux sites polymorphes qui varient systématiquement avec les individus porteurs de différents états d'une caractéristique phénotypique donnée, quels que soient leurs parents [2]. Dans les études d'associations entre des caractères phénotypiques quantitatifs et un ensemble de SNP, les méthodes de régression sont classiquement utilisées. Il est néanmoins nécessaire de sélectionner les SNP significatifs parmi l'ensemble des variations nucléotidiques observées. Plusieurs méthodes ont été utilisées telles que les méthodes stepwise [7], les méthodes bayésienne (shrinkage methods [37]) ou encore les méthodes SNP tagging [3]. Mais lorsque le nombre de SNP sélectionnés reste important il est difficile, voire impossible, de tenir compte des effets d'interactions. Une dernière approche alors employée pour diminuer la dimension du problème mais aussi pour étudier les phénomènes d'épistasie consiste à fonder son analyse sur les haplotypes (haplotype-based methods) [6], [4].

Dans le cadre de cette étude, nous proposons d'étudier l'effet de la variabilité du gène *CCR* d'*Eucalyptus urophylla* sur la teneur en lignine. Le genre *Eucalyptus* est un bon modèle car il représente un des feuillus les plus plantés au monde. Ses deux principales utilisations sont la production de bois d'énergie et de bois de trituration, toutes deux directement reliées à la teneur en lignine. L'ensemble des données actuelles traduit une très grande proximité des génomes entre les quelques espèces utilisées en plantation, pourtant adaptées à des conditions écologiques spécifiques [12]. Chez *E. urophylla*, des marqueurs microsatellites révèlent une importante variabilité dont la structuration ne reflète pas celle obtenue sur la base de caractères adaptatifs [35]. L'étude de la variabilité «utile» du gène *CCR* d'*E. urophylla* constituera un modèle de référence pour l'analyse de la diversité «fonctionnelle» chez les arbres forestiers. Nous proposons une méthode statistique qui permet simultanément d'estimer les variances génétiques des géniteurs (additive et de dominance) et de sélectionner les SNP discriminant les allèles qui expliquent significativement une part de la variation phénotypique.

2. MATÉRIEL ET MÉTHODES

2.1. Matériel végétal et conditions de croissance

Le matériel végétal est issu d'un plan de croisement factoriel incomplet de huit mères et huit pères *E. urophylla* non apparentés (graines d'origine de l'île de Florès, Indonésie) ayant généré par croisement contrôlé 348 individus répartis en 35 familles de pleins frères de 9 ou 10 arbres chacune. Le test de descendance a été mis en place en janvier 1992 dans la station

forestière de Kissoko, Pointe-Noire, République du Congo (4°45'S, 12°00'E, 50m d'altitude). La pluviométrie annuelle est de 1 200 mm avec 4 mois de saison sèche. La température annuelle moyenne est de 24°C. Les sols sont chimiquement pauvres (sable > 85 %, CEC<0,5 cmolc kg⁻¹), développés sur un matériau détritique épais [16]. Une fertilisation de 150 g de NPK 13-13-21 a été apportée à chaque arbre 15 jours après plantation. La parcelle unitaire est de 4 x 4 arbres plantés à 667 tiges / hectare. Les arbres ont été abattus à 14 ans, ayant atteint 22,5 m de hauteur moyenne pour une circonférence à 1,3 m de 53 cm.

L'ADN génomique des 16 parents et des 348 descendants a été extrait à partir de jeunes feuilles séchées, suivant la méthode de Doyle et Doyle [9]. Les échantillons sont stockés à -20 °C.

2.2. Caractères de croissance et analyse des lignines

La hauteur (Ht) et la circonférence (C) individuelles ont été mesurées lors de l'abattage. Un disque de 2 cm d'épaisseur a été prélevé à 1,3 m de hauteur sur chacun des 348 arbres. Un quartier de chaque disque a été découpé, puis broyé au broyeur à couteaux (poudre 4 mm) et un aliquote de la poudre obtenue a été broyé avec un broyeur à fléau (poudre 0,5 mm).

2.2.1. Analyses NIRS

Pour chaque échantillon, plusieurs spectres d'absorption dans le proche infrarouge ont été enregistrés sur les deux types de poudres. Les poudres sont conditionnées dans une salle régulée en température (20 °C) et en humidité (humidité relative de 65 %). L'humidité d'équilibre du bois est de l'ordre de 12 %. Le matériel utilisé pour les données spectrales d'absorption est un spectrophotomètre proche infrarouge à transformée de Fourier équipé d'une sphère d'intégration (BRUKER, modèle Vector 22N-I). La résolution retenue du spectromètre est de 8 cm⁻¹ et chaque spectre est la moyenne de 32 scans. La gamme spectrale est de 12,800-3,500 cm⁻¹ (780-2,860 nm). L'acquisition des spectres a été réalisée à partir d'une coupelle dont le fond en quartz laisse passer l'énergie lumineuse. Sur la base des données spectrales des 348 poudres et des prédictions de lignines obtenues au moyen d'équations de calibration existantes, 60 échantillons ont été sélectionnés pour être suffisamment représentatifs en termes de variabilité spectrale et en termes de gamme en lignine.

2.2.2. Analyses chimiques de référence

Pour les 60 échantillons sélectionnés, 5 grammes de poudre (0,5 mm) ont été soumis à extraction exhaustive dans un appareil de Soxhlet par les solvants toluène/éthanol (2/1, v/v), éthanol, puis eau. Le résidu pariétal (RP) ainsi obtenu, constitué essentiellement des parois, a été séché et pesé.

Le dosage de la lignine acido-insoluble (dénommée Lignine Klason, LK) et de la lignine acido-soluble (LAS) a été réalisé à partir de 300 mg de RP et selon la méthode décrite dans la littérature [8]. La structure des lignines a été étudiée par thioacidolyse, à partir de 20 mg de RP, et selon la méthode décrite antérieurement [19]. Les produits monomères de type guaiacyles (G) et syringyles (S) ont été analysés sous forme de leurs dérivés triméthylsilylés et par chromatographie en phase gazeuse couplée à la spectrométrie de masse (CPG-SM) [19]. Toutes ces analyses ont été réalisées en double.

2.3. Variabilité du gène *CCR* d'*E. urophylla*

2.3.1. Amplification et séquençage

Le gène *CCR* code la cinnamoyl CoA réductase. Cette enzyme catalyse la réduction de cinnamoyl CoA en aldéhyde cinnamylique. La séquence du clone génomique caractérisé chez *E. gunnii* ([17] ; Acc : X97433) a été utilisé comme séquence de référence pour la définition d'amorces et l'analyse de séquence. Sept couples d'amorces spécifiques ont été définis grâce au logiciel OLIGO 4. Un couple d'amorces supplémentaires a été utilisé pour amplifier la partie promotrice la plus en aval du gène, située entre la position -628 pb en amont et +66 pb en aval du site +1 de l'initiation de la transcription.

L'amplification des 8 fragments du gène *CCR* a été réalisée sur un thermocycleur Applied Biosystems Gene Amp PCR system 2700 en utilisant 2 µL de tampon 10X (Invitrogen), 0,8 µL dNTPs (solution stock à 5 mM), 0,8 µL MgCl₂ (solution stock à 50 mM), 0,4 µL de chaque amorce (solutions stock à 10 µM), 20 ng d'ADN génomique (solutions stock à 20 ng.µL⁻¹), 0,8 unités de Taq DNA polymérase native (Invitrogen) complété par une quantité suffisante d'eau ultrapure pour un volume final de 20 µL. Les conditions de PCR impliquent une dénaturation initiale à 94 °C pendant 5 minutes (min) puis 35 cycles de dénaturation à 94 °C pendant 30 secondes (sec), hybridation au T_m des amorces pendant 1 min, élongation à 72°C pendant 1 min et enfin une étape d'élongation finale à 72 °C pendant 10 min. La position des amorces sur la séquence référence et le T_m utilisé pour chaque couple sont donnés dans le tableau I.

Tableau I : Caractéristiques des amorces utilisées pour l'amplification des 7 fragments du gène *CCR* [20]. La séquence de chaque amorce, leur position sur la séquence de référence *E. gunnii* (start et end) et la température d'hybridation sont données.

Fragment		Amorce 5'→3'	Start (pb)	End(pb)	Tm (°C)
CCR1	Forward	CACCTCCTGAACCCCTCT	151	168	63
	Reverse	CGCACCCCTTGATGGCTTCT	555	528	
CCR2	Forward	GCGAGGAACCGTCAGGAAC	302	320	58
	Reverse	TTTCCTCCCAATCGTCTG	920	902	
CCR3	Forward	AAGAATGTGCGATGGCGAACC	840	860	70
	Reverse	GTCCCATCACCGTGGCT	1283	1265	
CCR4	Forward	ACGTAAGAAAGAGGGACCG	1086	1104	66
	Reverse	ACTTGAGGATGTGGATGATG	1757	1738	
CCR5	Forward	GCTACGGCAAGGCAGTGG	1614	1631	66
	Reverse	AACCGACAACCCACACCTG	2244	2226	
CCR6	Forward	CTTAGATAGATAGTCCCGC	2047	2065	56
	Reverse	CAAAGGGATTCAAGACAGG	2695	2677	
CCR7	Forward	CGTCATCATCGTTCTCTCT	2496	2514	56
	Reverse	TGACAACCTCCATTCCAA	3194	3117	

Les clones des produits de PCR ont été réalisés sans purification préalable en utilisant le kit de clonage TOPO-TA Cloning kit for sequencing (Invitrogen). Les transformations ont été réalisées en utilisant les cellules One Shot TOP 10 Chemically competent *E. coli* (Invitrogen). Pour chaque produit de clonage 12 colonies ont été prélevées et ont servi de matrice pour l'amplification PCR des fragments clonés (PCR sur colonie). Les amplifications ont été réalisées sur un thermocycleur Applied Biosystems Gene Amp PCR system 2700 en utilisant 2 µL de tampon 10X (Invitrogen), 0,8 µL dNTPs (solution stock à 5 mM), 0,8 µL MgCl₂ (solution stock à 50 mM), 2 µL de chaque amorce F13 et R13 (Invitrogen) (solution stock à 2 µM), 0,3 unités de Taq DNA polymérase native (Invitrogen) complété par une quantité suffisante d'eau ultrapure pour un volume final de 20 µL. Les produits de l'amplification PCR sur colonie ont été conservés pour séquençage. Les conditions de PCR sont une dénaturation de 10 min à 94 °C, 40 cycles avec une dénaturation de 45 sec à 94 °C, une hybridation à 55 °C pendant 45 sec et une élongation à 72 °C pendant 1 min et enfin une étape d'élongation finale à 72 °C pendant 10 min.

Les produits d'amplification PCR sur colonie ont été purifiés par filtration sur membrane MultiScreen PCRµ96 Plate (Millipore). Entre 20 et 40 ng d'ADN ont été prélevés et utilisés comme matrice pour la réaction de séquence. Le séquençage a été réalisé sur séquenceur ABI 3730 DNA Analyzer en utilisant le kit de séquençage Big Dye Terminator V1.1 cycle sequencing kit (Applied Biosystems). Le séquençage a été effectué dans un seul sens pour les fragments 1, 2, 3, 4, 5 et 7 du gène *CCR* et le fragment 6 à été séquençé dans les deux sens. Les séquences ont été analysées en utilisant

le logiciel CodonCode Aligner puis alignées dans l'éditeur de séquences BioEdit.

2.3.2. Génotypage des descendants du factoriel

Pour l'amplification du motif microsatellite une seconde PCR est réalisée avec 1,5 µl de produits d'amplification, 0,3 µl de chaque amorce à 2 µM (amorce M13 marquée à l'IRD700 ou IRD800 et amorce antisens), 0,08 µl de Taq (5 U), 1,5 µl Tampon 10X « Stand Taq Buffer » Biolabs, 0,6 µl de dNTPs (5 mM) et 10,72 µl d'H₂O. Les conditions d'amplification sont une dénaturation de 4 min à 94 °C, suivie de 34 cycles avec dénaturation de 30 sec à 94°C, hybridation de 1 min à 53 °C (*T_m* spécifique de l'IRD700 et 800), une élongation de 1 min à 72 °C et enfin une élongation finale de 10 min à 72 °C. Pour la révélation sur LI-COR4300 (NE-USA) 1 µl de produits PCR est ajouté à 10 µl de bleu (78 ml formamide, 10 ml de xylène cyanol 1 %, 10 ml de bleu de bromophénol 1%, 2 ml d'EDTA [0,5 M] pH8) et 9 µl d'H₂O. Les échantillons sont dénaturés 3 min à 94 °C puis aussitôt refroidis dans la glace. Ils sont ensuite déposés sur un gel d'acrylamide non dénaturant (960 µl de TBE 10X, 10 g d'urée, 5 g d'acrylamide 19/1 et qsp 26 g avec de l'eau milliQ, 16 µl de Temed et 160 µl d'APS à 10 %). La migration est réalisée à 1 500 V, 40 mA, 35W et 35 °C pendant 6 heures. La révélation de la fluorescence se fait par excitation laser à une longueur d'onde de 700 ou 800 nm.

2.4. Analyses statistiques

2.4.1. Estimation des paramètres génétiques

L'analyse des données est faite pour chaque caractère (procédure Proc Mixed de SAS) selon le modèle $X_{ijk} = \mu_x + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$ (1) où μ_x est la moyenne de la population, α_i effet aléatoire de la mère $i \sim N(0, \sigma^2_m)$, β_j effet aléatoire du père $j \sim N(0, \sigma^2_p)$, δ_{ij} effet aléatoire d'interaction mère-père $N(0, \sigma^2_{mp})$, ε_{ijk} erreur $N(0, \sigma^2_r)$. Les variances associées aux différents effets sont déduites de l'espérance des carrés moyens (méthode REML). Les variances génétiques sont calculées comme suit : la variance additive mère $\sigma^2_{A_m} = 4\sigma^2_m$ (2), la variance additive père $\sigma^2_{A_p} = 4\sigma^2_p$ (3). Ces deux variances sont deux estimations de la même variance additive de la population parentale, les mères et pères étant issus d'une population unique. Les différences d'estimations sont attribuées aux effets d'échantillonnage. La variance génétique additive dans la population de descendants est ainsi $\sigma^2_A = 2(\sigma^2_m + \sigma^2_p)$ (4), la variance de dominance $\sigma^2_D = 4\sigma^2_{mp}$ (5), la variance génétique totale en absence d'épistasie $\sigma^2_G = \sigma^2_A + \sigma^2_D$ (6) et la variance phénotypique totale dans la population de descendants $\sigma^2_P = \sigma^2_m + \sigma^2_p + \sigma^2_{mp} + \sigma^2_r$ (7). Ces variances permettent de calculer l'héritabilité au sens strict: $h^2 = \sigma^2_A / \sigma^2_P$ (8) et l'héritabilité au sens large de la

valeur familiale $H^2 = \sigma^2 G / \sigma^2 P$ (9). L'analyse multiple de variance (Proc GLM MANOVA SAS) permet de calculer les variances associées aux différents effets pour chacun des caractères à expliquer ainsi que les covariances $COV_{m_{xy}}$, $COV_{p_{xy}}$, $COV_{mp_{xy}}$, $COV_{r_{xy}}$ associées aux différents couples de variables X et Y. Ces covariances sont décomposées en Covariance additive, Covariance de dominance et Covariance résiduelle. Les variances et covariances sont utilisées pour le calcul des corrélations entre caractères. La corrélation phénotypique entre caractères $\rho_P(x,y)$ peut avoir diverses origines, génétique quand les ensembles de gènes contrôlant les deux caractères sont identiques ou partiellement identiques (quelques gènes communs et des gènes spécifiques) et environnementale quand les variations de l'environnement affectent conjointement les deux caractères. Les corrélations génétiques additives sont estimées par $\rho_A(x,y) = CovA(x,y) / \sqrt{(\sigma^2 A_x \sigma^2 A_y)}$ (10), les corrélations génétiques de dominance par $\rho_D(x,y) = CovD(x,y) / \sqrt{(\sigma^2 D_x \sigma^2 D_y)}$ (11), et enfin la corrélation d'origine purement environnementale est déduite de la relation algébrique entre les trois types, génétique ρ_A , phénotypique ρ_P et environnementale ρ_E et sous l'hypothèse d'une corrélation de dominance nulle, est donnée par $\rho_E = (\rho_P - h_x h_y \rho_A) / \sqrt{(1-h_x^2)(1-h_y^2)}$ (12) où h_x^2 et h_y^2 sont les héritabilités au sens strict des caractères x et y (8).

2.4.2. Sélection des SNP et test d'associations

La méthodologie développée [21] peut s'expliquer de la façon suivante : si K est le nombre total de SNP observés au sein du gène *CCR*, 2^K formes alléliques sont théoriquement observables. L'objectif est de déterminer un nombre $k \leq K$ de SNP ainsi que leur effet sur la variation du taux de lignine. On notera $(\alpha_1, \dots, \alpha_k)$ ces effets théoriques. En pratique, on observe un nombre p d'allèles dans la population d'études, associés aux K SNP ($p < 2^K$). Par exemple, si un seul SNP est inclus dans le modèle (on dira qu'il est « allumé ») alors seul $p=2$ allèles sont observables. Dans le cas où 3 SNP sont allumés, $2^3=8$ allèles sont alors au plus observables. Le modèle peut donc être formulé, dans un cadre bayésien, de la façon suivante : soit k un ensemble de SNP allumés, soit y le taux de lignine observé sur n individus issus d'un plan de croisement de sorte que

$$y | m, p, \underline{\alpha}, k, \sigma^2 \sim N(\alpha_1 + \alpha_2 + m + p, \sigma^2)$$

où m et p désignent les effets mères et pères. Nous considérons les effets de dominance négligeables.

$\alpha_i, i=1,2$ sont les effets des 2 allèles (individus diploïdes) dépendant des SNP inclus dans le modèle.

Dans le cadre bayésien, pour définir le modèle il est nécessaire de stipuler les lois *a priori*. Ainsi, on suppose que les effets mères et pères sont distribués selon une loi gaussienne d'espérance nulle et de variance σ_m^2 et σ_p^2 respectivement. On suppose que, connaissant les SNP "allumés" les effets des allèles suivent une loi gaussienne centrée de variance σ_a^2 . La distribution du nombre de SNP est uniforme sur $[0, K]$. Finalement, nous supposons que σ_m^2 , σ_p^2 , σ_a^2 et σ^2 sont indépendantes et suivent une loi inverse gamma $IG(10^{-3}, 10^{-3})$. Les lois *a priori* ont donc été choisies comme non informatives. L'inférence bayésienne repose sur l'étude de la loi *a posteriori* des paramètres sachant les observations. Comme le nombre k de SNP n'est pas connu, on est conduit à changer l'espace à parcourir. On propose donc un algorithme Reversible-Jump MCMC défini de la manière suivante [31] :

- 1/ Initialiser les paramètres θ et une configuration γ de SNP « allumés ». γ est un vecteur qui contient des 1 et des 0 (1 pour les SNP « allumés » 0 sinon).
- 2/ Pour $i=1$ à N itérations
- 3/ Choisir si on augmente ou on diminue le nombre de SNP « allumés ».
- 4/ Choisir un SNP parmi les SNP allumés ($\gamma=1$), on le transforme en 0 $\rightarrow \gamma^*$.
- 5/ Déterminer les nouveaux allèles efficaces, proposer de nouvelles valeurs des paramètres selon la loi *a posteriori* sachant les SNP choisis. $\rightarrow \theta^*$
- 6/ Calculer le ratio de Métropolis-Hastings en tenant compte du changement de dimension.
- 7/ On accepte ou on rejette θ^* et γ^* proportionnellement au ratio calculé.

3. RÉSULTATS

3.1. Analyse des caractères quantitatifs

Le tableau II présente la moyenne, le coefficient de variation phénotypique ainsi que les différents paramètres génétiques associés aux variables analysées. Les différents effets du modèle (1) sont toujours très hautement significatifs (significatif pour l'effet père sur C). La variabilité observée est globalement bien structurée par le modèle : la réalisation d'une telle structure sous l'hypothèse H_0 (les variables explicatives introduites,

mère, père et interaction, n'ont pas d'effet) a une probabilité généralement inférieure à 0,0001.

Tableau II : Moyenne (Moy), coefficient de variation (CV), variances (génétique additive mère σ^2Am , génétique additive père σ^2Ap , génétique additive totale σ^2A , dominance σ^2D , génétique totale σ^2G , phénotypique σ^2P) et héritabilités au sens strict, h^2 , et large de la valeur familiale H^2 .

	Moy	CV(%)	σ^2Am	σ^2Ap	σ^2A	σ^2D	σ^2G	σ^2P	σ^2A/σ^2G	h^2	H^2
Ht (m)	21.2	17.2	5.06	3.20	4.13	1.40	5.53	13.59	0.75	0.30	0.41
C (cm)	53.0	21.2	26.20	2.18	14.19	11.90	26.10	127.11	0.54	0.11	0.21
LK (%)	28.1	4.3	1.48	0.71	1.10	0.12	1.22	1.442	0.90	0.76	0.85
S/G	2.42	12.4	0.092	0.035	0.064	0.000	0.064	0.098	1.00	0.65	0.65

L'origine parentale a un effet majeur sur les performances des descendants, ceci quel que soit le caractère considéré. Le contrôle de la variabilité phénotypique observée entre familles est essentiellement d'origine génétique pour le taux de lignine et le rapport S/G ($H^2= 0,85$ à $0,65$). Les effets environnementaux sont plus sensibles pour les caractères de croissance ($H^2= 0,41$ et $0,21$). La variance de dominance représente la moitié de la variance génétique totale pour la circonférence. Pour les autres caractères, le contrôle génétique est essentiellement additif. Il s'ensuit que les héritabilités au sens strict, $h^2 = \sigma^2A/\sigma^2P$, pour les caractères chimiques sont très élevées ($0,65$ et $0,76$), confirmant ainsi le fait que les caractères « bois » sont plus héréditaires que la croissance.

L'analyse multiple de variance est utilisée pour estimer les covariances entre caractères pris deux à deux et en dériver les corrélations génétiques et environnementales (formules 10, 11, 12). Les résultats sont présentés dans le tableau III. Les corrélations génétiques et environnementales sont positives entre caractères de croissances (C et Ht). Elles sont négatives entre quantité et qualité de la lignine (S/G diminue quand LK augmente). Par contre, Les corrélations génétiques et environnementales entre teneur en lignine et croissance sont antagonistes. En sélectionnant sur la croissance (en particulier sur la croissance en hauteur) on s'attend à diminuer la teneur en lignine et le rapport S/G. Cette diminution, favorable à la production de pâte à papier, est préjudiciable pour l'obtention de bois énergie et pourrait alors être compensée par une amélioration de l'environnement.

Tableau III : Corrélations génétiques additives (gras) et environnementales (italique) entre les 4 variables. La valeur en grisé n'est pas significativement différente de zéro.

	LK	S/G	Ht	C
LK	1	-0.15	-0.55	-0.03
S/G	<i>-0.39</i>	1	0.27	0.37
Ht	<i>0.44</i>	<i>-0.17</i>	1	0.64
C	<i>0.27</i>	<i>-0.22</i>	<i>0.82</i>	1

3.2. Variabilité du gène *CCR* chez *E. urophylla*

3.2.1. Séquençage du gène *CCR* des géniteurs du factoriel

Les 7 fragments ciblés du gène ont été générés pour 15 des 16 génotypes *E. urophylla* étudiés. Le gène n'a pas pu être amplifié chez un géniteur du fait d'une mauvaise qualité de l'ADN génomique. Au total, la description de la variabilité de *CCR* est possible sur 94 % du gène avec en moyenne 116 séquences par fragment (120 attendues avec 8 clones par parent). La totalité des introns 1, 2, 3 et 4 sont disponibles. Pour les régions codantes, on dispose d'informations de séquence pour 100 % des exons 2, 3, 4, 48 % et 88 % respectivement pour les exons 1 et 5. Ces régions correspondent aux extrémités 5'UTR et 3'UTR non traduites. Des polymorphismes de type SNP, INDEL, poly A et microsatellite ont pu être mis en évidence au sein du gène.

Tableau IV : Répartition et caractéristiques des variations en fonction des zones codantes ou non codantes du gène *CCR*. Les données suivantes sont présentées : la taille des fragments en paires de base (pb) chez *E. gunnii* (*E. g.*) et *E. urophylla* (*E. u.*), le nombre de SNP (SNP, Bi-SNP et Tri-SNP), de SNP non synonymes (SNP NS), d'insertions/délétions (INDEL), de motif polyA et microsatellite. La fréquence des SNP par paires de bases (F SNP), la fréquence totale des variations (F Tot), le nombre de transversions (Nb V), le nombre de transitions (Nb S) et le rapport transition sur transversion (S/V) [20].

Nature de la séquence	Taille (pb)		Caractéristiques des sites polymorphes											
	<i>E. g.</i>	<i>E. u.</i>	SNP	SNP NS	Bi-SNP	tri-SNP	INDEL	Poly A	Microsat	F SNP	F Tot	Nb V	Nb S	S/V
Exon 1	324	156	1	1	-	-	-	-	-	1/78	1/78	1	1	1
Exon 2	156	156	4	1	-	-	-	-	-	1/31	1/31	2	3	1.5
Exon 3	185	186	4	1	-	-	-	-	-	1/37	1/37	2	3	1.5
Exon 4	355	355	14	2	-	-	-	-	-	1/22	1/22	2	14	7
Exon 5	220	194	1	5	-	1	-	-	-	1/28	1/28	3	6	2
Total exons	1240	1047	24	10	-	1	-	-	-	1/30	1/30	10	27	2.7
Intron 1	109	117	3	-	-	-	2	1	-	1/39	1/20	2	1	0.5
Intron 2	663	748	29	-	1	-	6	-	-	1/26	1/21	12	19	1.6
Intron 3	166	166	6	-	-	-	1	-	-	1/28	1/24	1	4	4
Intron 4	1024	1142	52	-	5	-	8	-	1	1/20	1/17	23	37	1.6
Total introns	1962	2173	90	-	6	-	17	1	1	1/23	1/19	38	61	1.6
Total gène	3202	3220	114	10	6	1	17	1	1	1/25	1/22	48	88	1.8

Le tableau IV donne la répartition des sites polymorphes au sein des régions codantes et non codantes. Le type et le nombre de sites variables sont différents entre les introns et les exons. Les SNP sont en moyenne plus fréquents dans les introns (1/23) que dans les exons (1/30). Dans les exons, les sites de mutations sont de plus en plus fréquents au sein de l'exon 4 (1/22). Les fréquences des SNP sont de 1/78 pour l'exon 1 (sur 48 % de l'exon 1), 1/31 pour l'exon 2, 1/37 pour l'exon 3 et 1/28 pour l'exon 5. L'ensemble des INDEL est mis en évidence dans les zones introniques. Enfin, le ratio entre le nombre de transitions et le nombre de transversions est plus élevé dans les exons (2,7) que dans les introns (1,6). Ceci montre un biais pour le type de mutation entre régions codantes et non codantes. Sur les 131 sites SNP étudiés, on met en évidence 10 mutations non synonymes dont 5 sont portées par le dernier exon.

3.2.2. Haplotypes parentaux

Les haplotypes ont été reconstruits pour 1 161 pb, soit 36 % du gène *CCR*, en considérant pour chacun des 15 géniteurs séquencés les fragments correspondant à l'exon 4 et l'intron 4. La sélection de cette région du gène a été motivée par son niveau de variabilité plus important (1 SNP pour 21pb en moyenne) et par la présence d'un microsatellite permettant de génotyper les descendants. Un total de 15 haplotypes a été mis en évidence à partir des 30 allèles portés par les 15 parents séquencés. L'un d'entre eux n'est reconstruit que partiellement avec plus de 50 % de l'information de séquence disponible (géniteur 14-132).

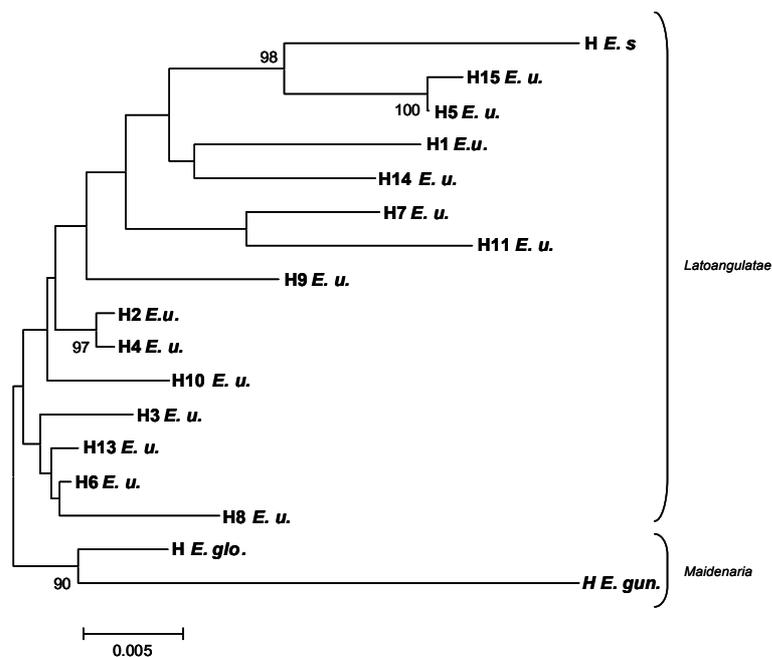


Figure 1 : Arbre des distances entre les 14 haplotypes d'*E. urophylla* (H *E. u.*) et des haplotypes pour *E. gunnii* (H *E. gun.*), *E. globulus* (H *E. glo.*) et *E. saligna* (H *E. s.*). Le calcul des distances est réalisé grâce à la méthode du plus proche voisin, pour une partie du gène *CCR* regroupant l'exon 4 et l'intron 4. Les résultats d'un test de bootstrap (1 000 répétitions) supérieurs à 80 % sont indiqués. L'unité de longueur de branche correspond au nombre de nucléotides différents entre deux séquences pondéré par le nombre de sites comparés. Tous les sites d'insertion/délétion ont été conservés.

L'arbre des distances entre les 14 haplotypes complets d'*E. urophylla* et les régions correspondantes des clones génomiques *CCR* de trois autres espèces du même sous-genre (*Symphomyrtus*), *E. gunnii* (X97433), *E. globulus* (AY656821) et *E. saligna* (AF297877) a été obtenu en utilisant la méthode des plus proches voisins [32]. Il met en évidence deux groupes d'haplotypes correspondant aux deux sections *Maidenaria* (*E. globulus* et *E. gunnii*) d'une part et *Latoangulatae* (*E. saligna* et *E. urophylla*) d'autre part (fig. 1).

3.2.3. Génotypage des descendants du factoriel

Le génotypage du microsatellite a été réalisé sur 208 individus issus de 21 familles (2 à 3 familles pour chacune des huit mères du dispositif). Le profil microsatellite des descendants permet de reconstruire leur haplotype et d'en déduire leur génotype à chacun des SNP le long des 1161pb considérées. Le tableau V indique la ségrégation des 15 haplotypes séquencés ainsi que celle

des deux haplotypes du géniteur 14-147 (H_a et H_b identifiés grâce au locus microsatellite). On dispose en moyenne de 9,5 individus par famille. L'homozygotie ou l'hétérozygotie supposées des géniteurs (données du séquençage) est confirmée par les ségrégations observées de type $\frac{1}{2} \frac{1}{2}$ ou $\frac{1}{4} \frac{1}{4} \frac{1}{4}$. La figure 2 illustre la ségrégation des haplotypes H9, H4, H7, H3 et H11 dans deux familles issues de la mère 14-138.

Tableau V : Haplotypes parentaux mis en évidence à partir des fragments 5 et 6 du gène *CCR*. La répartition des haplotypes au sein de chacune des familles est déterminée grâce au génotypage microsatellite des 208 descendants étudiés [20].

Mères	Pères	Effectif par famille	Génotype parentaux mère/père	Génotypes des descendants	Effectif pour chaque génotype
14-138	14-137	10	(H4;H7) / (H6;H11)	(H7;H6) / (H4;H6) / (H7;H11) / (H4;H11)	2 / 7 / 1 / 0
	14-130	10	(H4;H7) / (H3;H11)	(H4;H11) / (H7;H11) / (H4;H3) / (H7;H3)	6 / 3 / 1 / 0
	14-142	10	(H4;H7) / (H9;H6)	(H4;H9) / (H7;H9) / (H7;H6) / (H4;H6)	3 / 2 / 2 / 3
14-144	14-135	10	(H3;H6) / (H4;H5)	(H3;H4) / (H6;H5) / (H3;H5) / (H6;H4)	4 / 2 / 2 / 2
	14-130	9	(H3;H6) / (H3;H11)	(H3;H11) / (H3;H3) / (H6;H11) / (H6;H3)	2 / 4 / 2 / 1
	14-142	9	(H3;H6) / (H9;H6)	(H3;H6) / (H6;H9) / (H6;H6) / (H3;H9)	2 / 1 / 3 / 3
14-128	14-137	8	(H2;H1) / (H6;H11)	(H2;H6) / (H1;H11) / (H2;H11) / (H1;H6)	4 / 1 / 3 / 0
	14-135	10	(H2;H1) / (H4;H5)	(H2;H5) / (H1;H5) / (H1;H4) / (H2;H4)	4 / 2 / 2 / 2
	14-142	7	(H2;H1) / (H9;H6)	(H2;H6) / (H1;H6) / (H1;H9) / (H2;H9)	1 / 3 / 2 / 1
14-149	14-132	10	(H4;H4) / (H12;H3)	(H4;H12) / (H4;H3)	4 / 6
	14-148	10	(H4;H4) / (H4;H7)	(H4;H7) / (H4;H4)	6 / 4
14-133	14-135	10	(H13;H6) / (H4;H5)	(H6;H5) / (H13;H4) / (H6;H4) / (H13;H5)	4 / 1 / 1 / 4
	14-130	10	(H13;H6) / (H3;H11)	(H13;H11) / (H13;H3) / (H6;H11) / (H6;H3)	2 / 4 / 2 / 2
14-136	14-146	10	(H7;H5) / (H5;H5)	(H5;H5) / (H7;H5)	4 / 6
	14-148	10	(H7;H5) / (H4;H7)	(H7;H7) / (H4;H7) / (H5;H4) / (H5;H7)	2 / 4 / 2 / 2
	14-147	9	(H7;H5) / (H _a ;H _b)	(H5;H _b) / (H7;H _b) / (H7;H _a) / (H5;H _a)	2 / 4 / 2 / 1
14-140	14-132	9	(H8;H15) / (H12;H3)	(H8;H12) / (H15;H3) / (H15;H12) / (H8;H3)	2 / 1 / 4 / 2
	14-146	10	(H8;H15) / (H5;H5)	(H8;H5) / (H15;H5)	2 / 8
	14-147	10	(H8;H15) / (H _a ;H _b)	(H8;H _a) / (H15;H _b) / (H8;H _b) / (H15;H _a)	4 / 2 / 2 / 2
14-152	14-132	10	(H14;H10) / (H12;H3)	(H10;H12) / (H10;H3) / (H14;H3) / (H14;H12)	4 / 1 / 4 / 1
	14-148	8	(H14;H10) / (H4;H7)	(H10;H7) / (H14;H7) / (H10;H4) / (H14;H4)	3 / 3 / 2 / 2
	14-147	9	(H14;H10) / (H _a ;H _b)	(H14;H _b) / (H10;H _b) / (H10;H _a) / (H14;H _a)	3 / 2 / 2 / 2

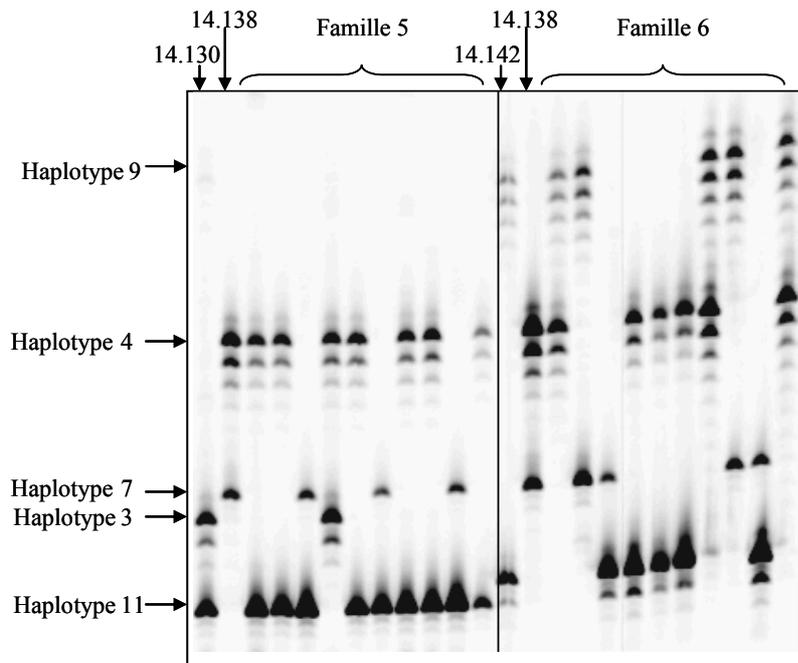


Figure 2 : Profils microsatellite obtenus pour deux familles intra-spécifique d'*E. urophylla* issues de la mère 14-138 et des pères 14-130 et 14-142. La ségrégation des 5 haplotypes portés par ces trois géniteurs (H3, H4, H7, H9 et H11) est donnée par le locus microsatellite.

3.2.4. Variabilité de la partie promotrice chez le géniteur 14-144

Un fragment de 694 pb correspondant à la partie promotrice la plus en aval du gène (-628 pb/+66 pb par rapport au site +1 de l'initiation de la transcription) a été généré à partir du génotype *E. urophylla* 14-144. Ce fragment contient toutes les séquences connues pour intervenir dans la régulation de l'expression du gène *CCR* chez *Eucalyptus* [17]. Les deux formes alléliques présentes diffèrent par 5 SNPs, dont 4 portent le même nucléotide que la séquence d'*E. gunnii*. Ils diffèrent de cette dernière par 20 SNPs et 1 INDEL de 14 pb (fig. 3).

En pratique on observe uniquement 11 haplotypes sur les 15 mis en évidence en 3.2.2.

Dans le cadre de nos algorithmes d'estimation type MCMC, nous avons lancé 20 chaînes avec des points de départ différents sur 100 000 itérations avec un burning de 10 000. Systématiquement, 3 SNP sont sélectionnés (SNP N°3, 10 et 50) avec des probabilités *a posteriori* pour l'une des chaînes égales respectivement à 0,68, 0,90 et 0,61. Les résultats pour les autres chaînes sont du même ordre. Ces SNP correspondent à deux SNP présents sur l'exon 4 et un sur l'intron 4. Le site SNP N°3 sélectionné correspond à un des deux SNP NS de l'exon 4, dont la variabilité T/G observée implique un changement d'acide aminé amidé (asparagine) en acide aminé dibasique (lysine).

Quatre groupes d'allèles sont reconstruits à partir des 3 SNP sélectionnés. La variance des effets alléliques est estimée à 0,39 avec une variance d'estimation de 0,082. Les variances des effets mères et pères sont égales à 0,236 et 0,274 avec des variances d'estimation égales à 0,025 et 0,055 respectivement. Par analogie à l'expression de l'héritabilité au sens strict utilisée dans les modèles classiques de génétique quantitative (cf section 2.4.1), nous avons calculé l'héritabilité du gène *CCR* (identifié aux fragments 5 et 6), l'héritabilité parentale et globale. Ces héritabilités sont respectivement estimées à 0,37, 0,48 et 0,85. L'héritabilité globale est la même que celle obtenue dans le cas classique, ce qui tend à valider les résultats.

On peut confirmer ce résultat à l'aide d'une analyse de variance qui montre des effets significatifs pour les groupes d'allèles (données non présentées). Lors des décompositions de type I, l'ordre d'introduction des facteurs dans le modèle est important dans le cas déséquilibré, [33]. Quel que soit l'ordre d'introduction, l'effet « groupe d'allèles » reste significatif, même si la significativité des effets est plus faible dans le deuxième cas. Néanmoins, ces résultats doivent être pris avec précaution car le plan d'expérience est fortement déséquilibré avec 132 individus pour le groupe 1, 23 pour le groupe 2, 6 le groupe 6 et enfin 5 pour le dernier groupe. Ce dernier groupe est constitué d'une seule forme haplotypique (H8) portée par un seul géniteur.

4. DISCUSSION

Différents auteurs ont mis en évidence l'importance du contrôle génétique de la qualité du bois d'eucalyptus [28], [36]. Ces caractères présentent généralement une héritabilité supérieure à celle des caractères de croissance, comme cela a pu être montré chez quelques espèces résineuses

[29], [15]. Les résultats obtenus dans cette étude vont dans le même sens. L'analyse des caractères quantitatifs indique l'importance des effets pléiotropiques des gènes (un gène peut avoir un effet sur deux caractères différents). Le taux de lignine est négativement corrélé aux autres caractères. Pour un niveau de ressources environnementales donné (eau, éléments minéraux, lumière...), les individus qui poussent rapidement semblent favoriser la synthèse de cellulose ou d'hémicellulose au détriment de celle de la lignine. Dans le même temps, ces individus favorisent la synthèse de S au détriment du G. Le matériel végétal choisi se prête particulièrement bien à l'objectif général de l'étude, notamment en ce qui concerne ses caractéristiques phénotypiques. En effet, l'observation d'une forte variabilité des caractères phénotypiques est un pré-requis à l'analyse de leur association à la variabilité moléculaire des gènes supposés en être à l'origine. L'analyse montre l'importance du contrôle génétique de la variabilité des différents caractères étudiés. L'héritabilité au sens large du taux de lignine, caractère majeur de cette étude, est supérieure à 0,8. La variance génétique additive représente une large part du contrôle génétique, conduisant à une très forte héritabilité au sens strict (0,76 pour le taux de lignine). Les gains génétiques directs attendus sont donc élevés, pouvant atteindre 2,5 points pour le taux de lignine, ceci avec un taux de sélection de 1 %. Le dispositif utilisé permet de faire la part des effets génétiques et environnementaux dans la corrélation entre caractères. L'absence de corrélation phénotypique entre caractères de croissance et taux de lignine résulte des effets antagonistes des corrélations génétiques et environnementales. La réduction du taux de lignine induit par la sélection sur la croissance peut être corrigée, au moins en partie, par une amélioration des pratiques culturales.

L'amplification de 94 % du gène *CCR* chez *E. urophylla* à partir de la séquence d'un clone génomique d'*E. gunnii* traduit un niveau de conservation important du gène chez ces deux espèces d'*Eucalyptus* appartenant au sous-genre *Symphomyrtus*. Les travaux réalisés sur ce gène chez *E. globulus* [27] vont aussi dans ce sens. S'il existe une conservation relative entre espèces pour ce gène, la variabilité du gène *CCR* mise en évidence chez *E. urophylla* apparaît plus élevée que celle décrite antérieurement chez *E. globulus* [27]. En effet, dans les introns on observe en moyenne un SNP pour 23 pb chez *E. urophylla* et 1/33 pb chez *E. globulus*. Cette tendance se retrouve aussi au niveau des exons avec un SNP pour 30 pb chez *E. urophylla* et 1/49 pb chez *E. globulus*. Trois hypothèses peuvent expliquer ces différences : i/ le pool de variabilité présent au sein des échantillons utilisés pour décrire la variabilité du gène est différent, ii/ la stratégie de séquençage, produits clonés dans notre étude *vs* produits PCR chez *E. globulus*, a un effet sur la variabilité mise en évidence, iii/ enfin, un niveau de variabilité plus important de *CCR* chez *E. urophylla* par rapport à *E. globulus*. Lorsqu'on s'intéresse aux fréquences observées chez des espèces plus éloignées et sur des gènes différents on constate aussi des

variations sensibles. Pour 18 gènes, le maïs présente en moyenne 1 SNP pour 124 pb au sein des zones codantes et 1 SNP pour 31 pb au sein des zones non codantes [5]. Les auteurs ont travaillé sur un ensemble de variétés représentant la diversité génétique des variétés améliorées aux Etats-Unis. On constate que pour une espèce ayant subi une forte pression de sélection le niveau de variabilité des introns, ne codant pas la protéine, reste proche de celui observé chez *Eucalyptus*, espèce encore peu sélectionnée. Pour les régions codantes, le niveau de variabilité au sein du gène *CCR* d'*E. urophylla* et d'*E. globulus* semble bien supérieur à celui observé pour d'autres gènes chez des espèces annuelles fortement sélectionnées ou chez d'autres espèces forestières [30]. L'impact de la variabilité nucléotidique sur la structure protéique de la *CCR* est plus fort chez *E. globulus* que chez *E. urophylla* malgré un niveau de variabilité global (exon et intron) moins important. Lacombe *et al.* [17] ont décrit chez *E. gunnii* plusieurs sites structuraux différents pouvant avoir un rôle fonctionnel chez la protéine *CCR*. Un site de huit acides aminés (KNWYCYGK) est proposé comme le site catalytique de la protéine (exon 4). Ce site catalytique, fortement conservé chez de nombreuses espèces végétales [26], n'est pas variable dans notre échantillon *E. urophylla*. Les SNPs non synonymes mis en évidence peuvent cependant avoir des effets non négligeables sur la structure tertiaire de la protéine et affecter ainsi ses propriétés physicochimiques.

Les résultats préliminaires obtenus sur le promoteur de *CCR* pour un des géniteurs *E. urophylla* semblent indiquer qu'aucun des sites polymorphes n'affecte les sites consensus de type MYB présents sur le promoteur *CCR* [18] et conduisant à une expression spécifique de ce gène dans le xylème. Malgré tout, on peut envisager que des variations de séquences dans le promoteur puissent influencer l'expression du gène. En effet, d'une part les complexes protéiques régulateurs de la transcription associent plusieurs protéines qui peuvent se fixer en plusieurs endroits sur la partie promotrice. D'autre part, leur fixation peut être potentiellement influencée par l'environnement nucléotidique. Il a donc été décidé de poursuivre ce travail *in planta* en exprimant un gène rapporteur GUS sous le contrôle de ces différentes formes alléliques. Une comparaison des résultats observés entre les deux allèles du géniteur *E. urophylla* 14-144 et les résultats décrits pour le promoteur d'*E. gunnii* [1] permettront de voir si l'expression du gène est modulée entre les génotypes.

L'analyse de la variabilité fonctionnelle du gène *CCR* réalisée sur 36 % de la séquence comprenant l'exon 4 et l'intron 4, zones les plus variables dans notre échantillon et codant le site catalytique présumé, met en évidence que le polymorphisme entre 11 haplotypes testés explique environ 40 % de la variance génétique du taux de lignine dans notre échantillon. Sur quatre groupes d'haplotypes différenciés par 3 SNP, un des groupes composé de l'allèle H8 présente des teneurs en lignine significativement supérieures aux

trois autres groupes d'allèles. Un effet de la variabilité du gène *CCR* sur des propriétés du bois a déjà été mis en évidence chez *E. nitens*. Thumma *et al.* [34] ont mis en évidence un effet significatif de deux haplotypes sur l'angle des micro-fibrilles. Le résultat du test d'association présenté dans cette étude demande à être confirmé en considérant à la fois le nombre total d'haplotypes, la séquence entière du gène, et l'ensemble des descendances du plan de croisement factoriel *E. urophylla* x *E. urophylla*.

D'un point de vue statistique, il reste à étudier l'influence des lois *a priori* pour les effets « groupes alléliques » : i) L'équivalent bayésien d'une analyse de variance avec effet fixe pour les effets groupe d'allèles consiste à choisir une loi normale de variance fixée et grande comme prior pour ces mêmes effets. Dans ce cas on s'intéresse à l'effet spécifique de chaque groupe d'allèles sur le caractère lignine. ii) Le choix d'une loi normale avec une variance inconnue que l'on cherche à estimer et sur laquelle on pose un prior Inverse-Gamma est la version bayésienne d'un modèle à effets aléatoires pour les groupes d'allèles. Cette situation propose de rendre compte de l'effet de la variabilité (qui se résume aux groupes d'allèles considérés) du gène *CCR* sur le caractère lignine. Enfin, il est nécessaire pour valider la méthode de la comparer avec les méthodes d'association classiquement utilisées [2].

Par la suite, on souhaite développer une extension de ce modèle qui prendrait en compte les données manquantes et une dépendance entre les effets des groupes d'allèles. On se demande aussi si le partage d'une même histoire évolutive conduit à des effets similaires. L'utilisation d'une similarité calculée à partir d'une distance de Nei [23] ou d'une distance fondée sur le graphe des coalescents [24] pourrait être envisagée pour valider cette hypothèse.

Ces résultats préliminaires sur l'effet de la variabilité du gène *CCR* sur la teneur en lignine chez *E. urophylla* sont très prometteurs pour envisager une sélection assistée par marqueurs pour la teneur en lignine au sein de programme d'amélioration des eucalyptus au Congo.

REMERCIEMENTS

Ce travail a été financé en partie par le Bureau des Ressources Génétiques et par une bourse CIFRE entre l'ANRT, le groupe Vallourec et le CIRAD (N° 213/2007).

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Baghdady A., Blervacq A.-S., Jouanin L., Grima-Pettenati J., Sivadon P., Hawkins S., 2006. - *Eucalyptus gunnii* CCR and CAD2 promoter activities are coordinated in lignifying cells during primary and secondary xylem formation in *Arabidopsis thaliana*. *Plant Physiology and Biochemistry*, 44: 674-683.
- [2] Balding D.J., 2006. - A tutorial on statistical methods for population association studies. *Nature Reviews; Genetics*, 7: 781-791.
- [3] Carlson C.S., Eberle M.A., Rieder M.J., Yi O., Kruglyak L., Nickerson D.A., 2004. - Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.*, 74: 106-120.
- [4] Chen J., Rodriguez C., 2007. - Conditional likelihood methods for haplotype-based association analysis using matched case-control data. *Biometrics*, 63: 1099-1107.
- [5] Ching A., Caldwell K.S., Juang M., Dolan M., Smith O.S., Tingey S., Morgante M., Rafalski A.J., 2002. - SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics*, 3: 19, 14 p.
- [6] Clark A.G., 2004. - The role of haplotypes in candidate gene studies. *Genetic Epidemiology*, 27: 321-333.
- [7] Cordell H.J., Clayton D.G., 2002. - A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: Application to hla in type 1 diabetes. *Am. J. Hum. Genet.*, 70: 124-141.
- [8] Dence C.W., 1992. - The determination of lignin. In: Lin S., Dence C.W. (eds), *Methods in lignin chemistry*. Springer-Verlag, Berlin, pp. 33-61.
- [9] Doyle J.J., Doyle J.L., 1990. - Isolation of DNA from fresh plant tissue. *Focus*, 12: 13-15.
- [10] Gion J.M., Rech P., Grima-Pettenati J., Verhaegen D., Plomion C., 2000. - Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Mol. Breed.* 6: 441-449.
- [11] Gion J.M., 2001. - Etude de l'architecture génétique des caractères complexes chez l'eucalyptus: des marqueurs anonymes aux gènes candidats. Thèse de doctorat, Université de Rennes I, 130 p.
- [12] Gion J.M., Cabannes E., Poitel M., Vignerot Ph. 2005. - Conservation of xylem preferentially expressed sequence tags in several subgenera of *Eucalyptus* L'Hérit. (Abstract) Plant and Animal Genome conference XIII, San-Diego, January 2005.
- [13] Grima-Pettenati J., Goffner D., 1999. - Lignin genetic engineering revisited. *Plant Science*, 145: 51-65.
- [14] Gupta P.K., Rustgi S., 2004. - Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genomics*, 4: 139-162.
- [15] Hannrup B., Cahalan C., Chantre G., Grabner M., Karlsson B., Bayon I. le, Jones G. L., Müller U., Pereira H., Rodrigues J.C., Rosner, S., Rozenberg P., Wilhelmsson L., Wimmer R., 2004. - Genetic parameters of growth and wood quality traits in *Picea abies*. *Scandinavian Journal of Forest Research*, 19 (1): 14-29.
- [16] Jamet R., 1975. - Internal report. ORSTOM, Cote MC 189, Centre de Brazzaville, 35 p.

- [17] Lacombe E., Hawkins S., Van Doorselaere J., Piquemal J., Goffner D., Poeydomenge O., Boudet A.M., and Grima-Pettenati J., 1997. - Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *The Plant Journal* 11 : 429-441.
- [18] Lacombe E., Van Doorselaere J., Boerjan W., Boudet A.M., Grima-Pettenati J., 2000. - Characterization of cis-elements required for vascular expression of the Cinnamoyl CoA reductase gene and for protein-DNA complex formation. *The Plant Journal*, 23: 663-676.
- [19] Lapierre C., Pollet B., Rolando C., 1995. - New insights into the molecular architecture of hardwood lignins by chemical degradative methods. *Res. Chem. Intermed.*, 21: 397-412.
- [20] *Mandrou E., Vigneron P., Plomion C., Gion J.M., 2008 – Functional variability of the CCR gene in *Eucalyptus urophylla*. *in prep.*
- [21] *Mortier F., Etienne M.P., Gion J.M., 2008. - SNP selection and association study for related individuals : a reversible jump approach. *in prep.*
- [22] Neale D.B., Savolainen O., 2004. - Association studies of complex traits in conifers. *Trends in Plant Science* 9 (7): 325-330.
- [23] Nei M., 1972. - Genetic distance between populations. *Amer. Natur.*: 106-283
- [24] Nordborg M., 2001. - Coalescent Theory. In *Handbook of Statistical Genetics*, Edited Balding D.J., Bishop M., Cannings C. John Wiley & Sons: 179-212.
- [25] Paux E., Tamasloukht M., Ladouce N., Sivadon P., Grima-Pettenati J., 2004. - Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant Mol Biol.* 55: 263-280
- [26] Pichon M., Courbou I., Beckert M., Boudet A.M., and Grima-Pettenati. J., 1998. - Cloning and characterization of two maize cDNAs encoding Cinnamoyl-CoA Reductase (CCR) and differential expression of the corresponding genes. *Plant Molecular biology*, 38: 671-676.
- [27] Poke F.S., Vaillancourt R.E., Elliott R., Reid J.B., 2003. - Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase 2 (CAD2). *Mol. Breed.*, 12: 107-118.
- [28] Poke F.S., Potts B.M., Vaillancourt R.E., Raymond C.A., 2006. - Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*. *Ann. for. sci.*, 63, 8: 813-821.
- [29] Pot D., Chantre G., Rozenberg Ph., Rodrigues J.C., Jones G.L., Pereira H., Hannrup B., Cahalan C. and Plomion C., 2002. - Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster* Ait.). *Ann. For. Sci.* 59: 563-575.
- [30] Pot D., McMillan L., Echt C., Le Provost G., Garnier-Géré P., Catoand S., Plomion C., 2005. - Nucleotide variation in genes involved in wood formation in two pine species. *New Phytol.* 167 (1): 101-112.
- [31] Richardson S. and Green P.J., 1997. - On bayesian analysis of mixtures with an unknown number of components. (with discussion). *J. R. Stat. Soc., Ser. B*, 59 (4): 731-792.
- [32] Saitou N., & Nei M., 1987. - The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
- [33] Scheffé H., 1959. - *The Analysis of Variance*. John Wiley & Sons

- [34] Thumma B.R., Nolan M.F., Evans R., and Moran G.F., 2005. - Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics*, 171: 1257-1265.
- [35] Tripiana V., Bourgeois M., Verhaegen D., Vigneron P., and Bouvet J.M., 2007. - Combining microsatellites, growth, and adaptative traits for managing *in situ* genetic resources of *Eucalyptus urophylla*. *Can. J. For. Res.*, 37: 773-785.
- [36] Vigneron Ph., Giordanengo Th., Ognouabi N., Gion J.M., Chaix G and Baillères H., 2003. - Genetic components of wood quality traits in *Eucalyptus urophylla* x *grandis* full sib families. IUFRO Congress unit 2.08.03 "Eucalyptus in a Changing World", 11 a 15 de Outubro de 2004, Aveiro, Portugal.
- [37] Wang H., Zhang Y.M., Li X., Masinde G.L., Mohan S., Baylink D.J., and Xu S., 2005. - Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics*, 170: 465-480.

Analyse du polymorphisme des α 1,2fucosyltransférases chez le lapin de garenne et recherche d'une relation avec la sensibilité à la maladie hémorragique virale

Patrice GUILLON⁽¹⁾, Nathalie RUVOËN-CLOUET^(1,2),
Béatrice LE MOULLAC-VAIDYE⁽¹⁾, Stéphane MARCHANDEAU⁽³⁾,
Jacques LE PENDU^{(1)*}

⁽¹⁾INSERM, U601, Université de Nantes, Institut de Biologie, 9 Quai Moncoussu,
44093 Nantes, France

⁽²⁾Ecole Nationale Vétérinaire de Nantes, Route de Gachet, BP40706,
44307 Nantes, France

⁽³⁾Office National de la Chasse et de la Faune Sauvage, 53 rue Russeil,
44000 Nantes, France

Abstract: Analysis of wild rabbit α 1,2fucosyltransferases polymorphism and search of a relationship with sensitivity to viral hemorrhagic disease. Rabbit Hemorrhagic Disease Virus (RHDV) is an emergent and highly virulent Calicivirus (Lagovirus), which has become the major cause of mortality in wild rabbit populations (*Oryctolagus cuniculus*). Animals die in the early 72 hours by flashing hepatitis and hemorrhages. RHDV binds to the histo-blood group antigen H type 2 (Fuc α 2Gal β 4GlcNAc) expressed by tracheal and intestinal epithelial cells. Among Caliciviridae, RHDV and Norovirus (NoV) share carbohydrate ligands of the same family. In human, polymorphism at the FUT2 locus is responsible for the genetic susceptibility or resistance to infection by various NoV strains. In rabbit, both the *Fut2* and *Sec1* genes encode α 1,2 fucosyltransferases, which synthesize the H type 2 antigen in epithelial cells, whereas *Fut1* encodes another α 1,2 fucosyltransferase expressed in brain. Nonfunctional alleles at the *Fut2* and *Sec1* loci could thus potentially determine resistance to RHDV. In this study, we looked for the presence of the H type 2 antigen on buccal epithelial cells of wild rabbits from 2 distinct geographic French areas which are under RHDV pressure and from one area where RHD was never recorded. Some animals expressing either weakly or not at all the H type 2 antigen were found in the 3 populations ("nonsecretor-like" phenotype). Their frequency increased according to the impact of RHDV, suggesting that RHD outbreaks selected survivors with low expression of the virus ligand. The polymorphism of the *Fut1*, *Fut2* and *Sec1* coding regions of 138 wild rabbits from 2 areas under RHDV pressure was then analysed. A high polymorphism was observed for the *Fut2* and *Sec1* genes, with 14 and 26 variable positions, respectively. In contrast,

* Correspondance et tirés à part : jlependu@nantes.inserm.fr

the *Fut1* gene showed a very low polymorphism. No introduction of either a non-sense codon or a frameshift that would obviously generate nonfunctional alleles was found. Gene conversion between the 2 rabbit *Fut2* and *Sec1* genes was observed. Conversion is known to enhance the genetic polymorphism while contributing to homogenization of paralogues, consistent with a potential role of *Fut2* or *Sec1* in relation with environmental factors. The frequency of a *Sec1* haplotype was increased among survivors relative to dead animals in one area hit by RHDV outbreaks. *Sec1* enzyme variants showed either moderate, low or undetectable catalytic activity. In contrast, all variant *Fut2* enzymes showed strong catalytic activity. Thus, although the genetic polymorphisms of the coding regions of rabbit α 1,2 fucosyltransferases described here cannot explain the “nonsecretor-like” phenotype, a *Sec1* variant may be associated with resistance to RHD.

**rabbit hemorrhagic disease virus/ calicivirus/ fucosyltransferase/ O(H)
blood-group/ disease resistance**

Résumé : La maladie hémorragique virale du lapin (RHD) est une des causes principales de mortalité des lapins de garenne. Le virus de la RHD (RHDV) se fixe sur l'antigène H type 2 dont la synthèse nécessite une α 1,2 fucosyltransférase. Nous avons recherché un déficit de synthèse de l'antigène chez le lapin qui pourrait conduire à une résistance au RHDV. Son absence ou sa faible expression sur les cellules épithéliales buccales de certains lapins a permis de définir un phénotype déficient, “nonsecréteur-like”. La fréquence de ce phénotype s'est avérée associée à l'impact de la RHD dans des populations de lapins de garenne. Trois gènes d' α 1,2 fucosyltransférases existent chez le lapin : *Fut1*, *Fut2*, *Sec1*. Seuls les deux derniers sont exprimés dans les tissus épithéliaux. Le polymorphisme de ces trois gènes a été caractérisé chez 138 lapins de deux sites sous pression du RHDV. Au contraire de *Fut1*, un fort polymorphisme a été observé pour *Fut2* et *Sec1* associé à une importante conversion génique entre ces deux gènes. De plus, sur un des deux sites, la fréquence d'un variant de *Sec1* est nettement augmentée dans le groupe des lapins survivants comparé au groupe des morts lors d'épidémies. L'analyse fonctionnelle des variants enzymes correspondant à l'ensemble des haplotypes *Fut2* et *Sec1* n'explique pas la survie de certain lapins. L'association d'un haploype *Sec1* avec la survie pourrait s'expliquer par une association avec un déficit d'expression d' α 1,2 fucosyltransférase, plutôt qu'un déficit d'activité catalytique.

**virus de la maladie hémorragique du lapin/ calicivirus/ fucosyltransférase/
groupe sanguin O(H)/ résistance maladie**

1. INTRODUCTION

Le RHDV (rabbit haemorrhagic disease virus) est un virus à ARN positif simple brin de la famille des Caliciviridae (genre lagovirus). Il est responsable d'une maladie foudroyante du lapin (*Oryctolagus cuniculus*) caractérisée par une hépatite fulgurante accompagnée d'une coagulation

intravasculaire disséminée, la maladie hémorragique virale du lapin ou RHD (rabbit haemorrhagic disease) [3]. Les premiers cas ont été décrits en Chine en 1984 et en France en 1987-1988. Son impact, tant dans les populations sauvages de lapins de garenne que chez les lapins domestiques, est très élevé puisque les taux de mortalité peuvent atteindre 90 % [4]. Actuellement la lutte contre cette maladie n'est possible qu'en élevage, la prophylaxie étant basée sur la vaccination systématique des animaux. Dans les populations de lapins de garenne, la RHD est considérée comme l'une des causes principales de régression de l'espèce dont les prélèvements cynégétiques sont passés de 6,4 millions de lapins en 1983-1984 à 3,2 millions en 1998-1999 [5]. L'impact de la maladie est particulièrement fort dans les régions sèches au Sud de la France et en Péninsule Ibérique. Dans ces zones, la diminution drastique des populations de lapins de garenne menace plusieurs espèces qui dépendent du lapin pour leur survie. Il s'agit du lynx pardelle (*Lynx pardinus*), de l'aigle ibérique (*Aquila aldaberti*) et de l'aigle de Bonelli (*Hieraetus fasciatus*), prédateurs spécialistes du lapin. En France, on note en outre une régression du lézard ocellé (*Lacerta lepida*) dans les zones où le lapin se raréfie. Ce lézard, dont la limite Nord de l'aire de répartition est le sud de la Vendée, utilise les terriers creusés par les lapins comme sites d'hivernage et comme gîtes diurnes en cas de forte chaleur estivale.

Nous avons observé antérieurement que le RHDV se fixe sur un trisaccharide présent sur les cellules épithéliales des voies respiratoires et digestives supérieures du lapin, portes d'entrée du virus [9]. Ce trisaccharide, $\text{Fuc}\alpha 2\text{Gal}\beta 4\text{GlcNAc}\beta\text{-R}$, existe chez d'autres espèces de mammifères, par exemple chez l'homme. Sa synthèse nécessite l'action d'une fucosyltransférase qui catalyse l'addition d'un résidu fucose (Fuc) en position $\alpha 1,2$ sur le galactose (Gal) du précurseur [6]. Chez le lapin, trois gènes codent pour de telles enzymes : les gènes *Fut1*, *Fut2* et *Sec1*. Seuls *Fut2* et *Sec1* sont exprimés dans les cellules épithéliales [1].

Chez l'homme, nous avons montré que le polymorphisme du gène *FUT2*, orthologue de *Fut2* du lapin, contrôlait la sensibilité ou la résistance au virus Norwalk, prototype des calicivirus humains (genre norovirus). En effet ce virus, responsable de gastroentérites, se fixe sur les cellules épithéliales digestives humaines par l'intermédiaire d'un trisaccharide fucosylé très voisin de celui sur lequel se fixe le RHDV : $\text{Fuc}\alpha 2\text{Gal}\beta 3\text{GlcNAc}\beta\text{-R}$. L'addition du résidu fucose aux cellules épithéliales de la surface de l'intestin grêle humain requiert l'enzyme *FUT2*. Le gène *FUT2* est polymorphe, certains allèles, dénotés *FUT2+*, codent pour une enzyme fonctionnelle. D'autres, dénotés *FUT2-*, sont inactifs ou codent pour des enzymes à l'activité catalytique fortement diminuée. La présence d'un allèle *FUT2+* confère le phénotype « sécreteur » caractérisé par la présence du motif glycanique $\text{Fuc}\alpha 2\text{Gal}\beta$ sur les cellules épithéliales

ou dans les sécretions comme la salive. Le génotype *FUT2*- correspond au phénotype « nonsécreteur » caractérisé par l'absence de ce motif. Les individus nonsécreteurs sont totalement résistants à l'infection par le virus Norwalk et par d'autres souches de norovirus, montrant qu'il s'agit d'un ligand indispensable à l'infection (pour une revue voir [2]). Dans la mesure où le RHDV se fixe sur un glycanne très voisin de celui qui est utilisé par plusieurs souches de norovirus, nous avons émis l'hypothèse qu'il pourrait exister des lapins de phénotype « nonsécreteur-like » dépourvus du ligand glycanne fucosylé sur leurs cellules épithéliales. De tels lapins seraient résistants à l'infection par le RHDV par absence de récepteur. Nous avons donc recherché l'existence du phénotype « nonsécreteur-like » chez les lapins de garenne. Ce phénotype ayant pu être mis en évidence, nous avons déterminé le polymorphisme des régions codantes des trois gènes d' α 1,2fucosyltransférases *Fut1*, *Fut2* et *Sec1* du lapin et nous avons cherché à le mettre en relation avec le phénotype nonsécreteur-like et la survie à des épidémies de RHD.

2. MATÉRIEL ET MÉTHODES

2.1. Échantillonnage

Sur 3 territoires d'étude de l'ONCFS correspondant à 3 sites géographiques touchés de manière très différente par la RHD, des frottis buccaux réalisés à l'aide d'un coton-tige ont été obtenus à l'occasion de captures d'animaux : le site Dompierre, non affecté, par la RHD, le site de Cerisay où la RHD a causé une mortalité de l'ordre de 50 % en 2000 et le site de Chèvreloup, touché par la RHD en 1995 et 1996, où la mortalité a dépassé 90 %.

Par ailleurs, des prélèvements d'un morceau d'oreille (environ 1 cm²) conservé dans l'alcool ont été utilisés. Ces prélèvements avaient été effectués avant, pendant et après les épidémies de 1995 et 1996 sur le site de Chèvreloup et pendant l'épidémie sur le site de Cerisay.

2.2. Recherche du phénotype « nonsécreteur-like »

Les coton-tiges ont été mis en présence d'1 ml de PBS puis centrifugés à 1500 rpm pendant 5 minutes. Le culot cellulaire a été récupéré et incubé pendant 1 h à température ambiante en présence de la lectine UEA-I couplée au FITC (Sigma). Cette lectine reconnaît spécifiquement le motif Fuc α 2Gal β 4GlcNAc β -R reconnu par le RHDV. Les cellules ont été ensuite lavées 3 fois au PBS, puis les noyaux colorés à l'iodure de propidium avant observation sous un microscope à épifluorescence (Olympus). La

fluorescence membranaire des cellules à la morphologie de cellules épithéliales buccales a été notée fortement positive, faible ou absente.

2.3. Amplification des gènes *rFut1*, *rFut2* et *rSec1* et séquençage

Les extractions d'ADN ont été réalisées à partir des fragments d'oreille à l'aide du kit d'extraction QIAamp DNA Mini Kit (Qiagen). Les parties codantes complètes des gènes *Fut1*, *Sec1* et *Fut2* ont été amplifiées par PCR à l'aide d'amorces déduites des séquences X80226, X91269 et X80225 pour *Fut1*, *Fut2* et *Sec1*, respectivement : gen *Fut2* (CCAAGAGCT AACTAACATC CCTGCTG), et gen *Fut2* inv (GCAAGAAACG GGTCTTGCC CACGTGAC) pour le gène *Fut2*, gen *Sec1* (CTTCAGCCTG CACAGCCCCA GCCGTT) et gen *Sec1* inv (ACTCCGTGTG GGATCCCGGG TCCATGC) pour le gène *Sec1*, gen *Fut1* (CTGTCTGGAA AGCCATCATC) et gen *Fut1* inv (GTAGAATCAC TCTGGCTGTG) pour le gène *Fut1*. L'amplification a été réalisée en présence de 2,5 % DMSO, 3 minutes à 94 °C, 35 cycles comprenant 30 secondes à 94 °C, 30 secondes à 68 °C et enfin 2 minutes à 72 °C, puis de 10 minutes d'élongation finale à 72 °C. Les séquences ont été réalisées sur un séquenceur Applied Biosystems 3730 (plate-forme Ouest-Génopole /IFR26) et analysées à l'aide du logiciel BioEdit v7.0.5.3 (Hall T.A, North Carolina State University, Department of Microbiology).

2.4. Amplification et clonage

Pour le clonage, les amplifications ont été réalisées grâce au kit AmpliTaq Gold with GeneAmp 10X PCR Gold Buffer & MgCl₂ Solution (Applied Biosystems) avec les amorces *rFut* 2.3 (ATGAGCACCG CCCAGGTCCC CTTC) et *rFut* 2.2 inv (TCAGTGCTTG AGCAATGGGG ACAG) pour *Fut2*, *rSec1*.3 (ATGAGATTCG CCCCTGACTA TGTCC) et *rSec1*.2 inv (CTAGAGGCCA CTCCACAAGG C) pour *Sec1*. Les fragments obtenus correspondent à une taille de 1044 pb pour *Fut2* et 1066 pb pour *Sec1*. Le clonage a été réalisé en vecteur pcDNA 3.1 grâce au kit NT-GFP Fusion TOPO TA Expression Kit (Invitrogen) selon les recommandations du fabricant en utilisant 2 µl de produit de PCR frais pour une incubation de 20 minutes à température ambiante. Les bactéries utilisées pour la transformation sont les *E. coli* One Shot TOP10F⁺(Invitrogen). La présence de l'insert dans le vecteur a été vérifiée par PCR sur 20 à 30 colonies. Cette amplification a permis également de renseigner le sens d'insertion dans le vecteur, en utilisant l'amorce T7 Pro (TAATACGACT CACTATAGGG) avec une amorce interne aux inserts *Fut2* et *Sec1*, *rFut*3.2 Inv (GTTGAGGTGG TAGTTCTGCC).

2.5. Test d'activité des variants d' α 1,2fucosyltransférases

Pour étudier la fonctionnalité de chacun des variants alléliques de *Fut2* et de *Sec1*, les régions codantes ont été clonées dans un vecteur d'expression eucaryote de façon à obtenir des protéines de fusion avec la GFP (green fluorescent protein) du côté intracytoplasmique (N-ter). Après amplification et purification à l'aide du kit Miniprep Qiagen, les vecteurs pcDNA 3.1 recombinants ont été transfectés de façon transitoire dans des cellules CHO par la Lipofectamine 2000 (Invitrogen). Quarante huit heures après la transfection, les cellules ont été récupérées et la présence de résidus fucose liés en position α 1,2 a été testée par cytométrie en flux à l'aide de la lectine UEA-I. Pour cela $2,5 \times 10^5$ cellules viables sont incubées à 4°C pendant 20 min avec la lectine marquée par la biotine (Vector) diluée à 5 μ g/ml. Après 3 lavages, les cellules ont été incubées en présence de streptavidine couplée à la phycoérythrine (BD Biosciences) 0,5 μ g/ml. Après 3 nouveaux lavages, la fluorescence des cellules a été mesurée sur un FACScalibur (Becton-Dickinson) et analysée à l'aide du logiciel CellQuest. L'efficacité de l'expression protéique a été quantifiée par l'intensité de la fluorescence de la GFP en FL1 et la présence du ligand Fuc α 2Gal β 4GlcNAc β -R a été mesurée en FL-2.

3. RÉSULTATS

3.1. Le phénotype « nonsécreteur-like »

Chez l'homme, les phénotypes sécreteur ou nonsécreteur sont définis à partir d'échantillons de mucines salivaires. Le marquage par la lectine UEA-I ou son absence définissent les deux phénotypes. Les mucines salivaires du lapin et d'autres espèces de mammifères ne semblent pas exprimer la structure glycanique correspondante. Nous l'avons donc recherché sur les cellules épithéliales buccales recueillies à l'aide de coton-tiges. Nous avons observé une intensité de marquage variable d'un individu à l'autre et défini arbitrairement un seuil qualitatif entre faiblement positif (douteux) et clairement positif ou négatif. Les individus faiblement positifs et négatifs ont été regroupés pour former le groupe de phénotype « nonsécreteur-like ». Le tableau I montre une nette association entre le pourcentage d'animaux possédant ce phénotype et l'impact de la RHD dans trois populations de lapins de garenne.

Tableau I : Représentation du phénotype « nonsécréteur-like » dans 3 populations de lapins de garenne en fonction de l'impact des épidémies de RHD.

<i>Population</i>	<i>% nonsécréteur-like</i>	<i>% morts de RHD</i>
Dompierre	14,8 (27)	0
Cerisay	47,1 (70)	~50
Chèvrelop	71,8 (32)	~90

*** [*]

3.2. Polymorphisme des gènes d' α 1,2fucosyltransférases chez les lapins de garenne

Le séquençage des parties codantes des gènes *Fut1*, *Fut2* et *Sec1* des lapins de garenne de 2 sites français a permis de mettre en évidence un important polymorphisme. Des échantillons de 95 animaux ont été obtenus du site de Chèvrelop et de 43 animaux du site de Cerisay. Pour cet ensemble de 138 individus, sur environ 1 kb, un total de 2, 14 et 26 positions variables a été trouvé pour les gènes *Fut1*, *Fut2* et *Sec1*, respectivement. Sur ces différentes positions, on note 1 mutation silencieuse et 1 non silencieuse pour *Fut1*, 8 mutations silencieuses et 6 mutations non silencieuses pour *Fut2*, 12 mutations silencieuses et 14 non silencieuses pour *Sec1*. Ces mutations se répartissent sur 3 haplotypes pour *Fut1* et sur 7 haplotypes pour les 2 autres gènes, montrant que le gène *Sec1* présente une variabilité plus importante que le gène *Fut2* et qu'en comparaison des deux autres, le gène *Fut1* est très peu variable.

La structure de ces enzymes n'est pas encore connue. Néanmoins, par homologie avec d'autres glycosyltransférases, une structure générale a pu être déduite. Ces enzymes possèdent une courte région N-terminale intracytoplasmique, un domaine transmembranaire, une région tige possédant des sites probables de O-glycosylation et une région catalytique divisée en deux sous domaines. Le sous domaine N-terminal étant vraisemblablement impliqué dans la fixation du substrat donneur, le GDP-fucose et le sous domaine C-terminal étant plus impliqué dans la fixation du substrat accepteur. La comparaison des séquences alléliques de *Fut2* et *Sec1* révèle une région correspondant à la majeure partie du domaine catalytique où sont présentes 8 mutations partagées entre les deux gènes (fig.1). La présence de ces polymorphismes partagés signe l'existence d'une conversion génique entre *Fut2* et *Sec1* impliquant une large partie de la région codant le domaine catalytique.

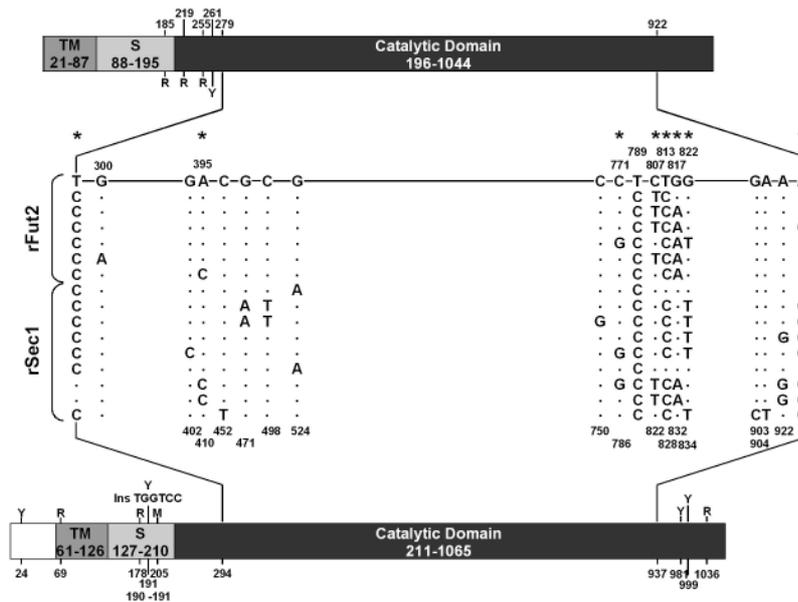


Figure 1 : Polymorphisme des régions codantes des gènes *Fut2* et *Sec1* des lapins de garenne de deux populations françaises (sites de Cerisay et Chèvreloop). Les positions variables sont positionnées le long des séquences correspondant aux différents domaines peptidiques codés par le gène *Fut2* (en haut) et le gène *Sec1* (en bas). La lettre Y indique C ou T, la lettre R indique A ou G et la lettre M indique A ou C. La région montrant des polymorphisme partagés entre les 2 gènes est montrée au centre. Dans cette zone, seules les positions présentant une variation entre haplotypes ou entre les 2 gènes sont montrées. Toutes les autres positions sont identiques. Les polymorphismes partagés sont notés par un astérisque.

3.3. Distribution des haplotypes d'α1,2fucosyltransférases dans les populations de lapins de garenne et relation avec la survie

Les variations nucléotidiques des régions codantes de *Fut2* et *Sec1* des lapins des territoires de Chèvreloop et Cerisay se regroupent suivant 7 haplotypes pour chacun des 2 gènes. Toutefois les fréquences de ces haplotypes diffèrent dans les deux populations. Ainsi l'haplotype C de *Fut2* dont la fréquence est de 0.21 dans la population de Cerisay n'est pas décelé à Chèvreloop. Inversement la fréquence de l'haplotype v9 de *Sec1* est de 0.46 à Chèvreloop et seulement de 0.01 à Cerisay. Les *Fst* calculés à l'aide du logiciel GENEPOP (<http://genepop.curtin.edu.au/>) sont respectivement de 0.085 et de 0.162 pour *Fut2* et *Sec1* suggérant que les 2 populations sont modérément distinctes génétiquement.

L'analyse du devenir des animaux à l'issue des épidémies de RHD n'a montré aucune association statistiquement significative entre la survie et un haplotype *Fut2* (résultat non montré). En revanche, à Chèvreloup, la fréquence de l'haplotype v5 du gène *Sec1* est significativement augmentée dans le groupe des survivants par rapport au groupe des animaux morts (fig. 2).

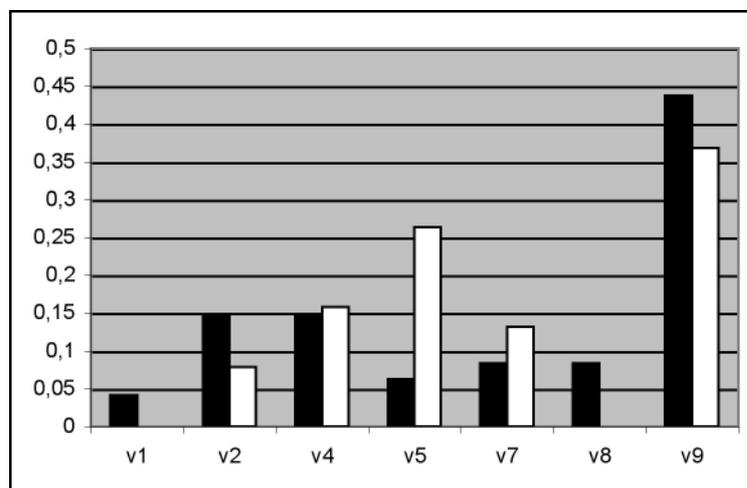


Figure 2 : Répartition des haplotypes *Sec1* entre survivants et morts lors des épidémies de 1995-1996, dans la population de Chèvreloup. Les fréquences sont indiquées en ordonnées, barres noires pour les morts et blanches pour les survivants. Les fréquences de l'haplotype v5 sont significativement différentes entre morts et survivants (test de Chi2, $p < 0.01$).

3.4. Analyse fonctionnelle des variants d' α 1,2fucosyltransférases

Après transfection dans des cellules CHO, dépourvues d'activité α 1,2fucosyltransférase endogène, la présence du ligand fucosylé du RHDV à la surface cellulaire a été mesurée par cytométrie en flux à l'aide de la lectine UEA-I. La fluorescence de la GFP permet de normaliser l'efficacité d'expression du transgène. Sur un total de 7 variants de l'enzyme *Fut2*, 6 ont pu être testés de cette façon. Les 6 se sont révélés capables d'entraîner la synthèse de motifs α 1,2fucosylés à la surface des cellules de manière identique. Tous les variants de *Sec1* ont pu être testés dans les mêmes conditions. Plusieurs se sont révélés totalement inactifs puisqu'aucun marquage spécifique n'a pu être détecté après transfection bien que la protéine était correctement exprimée. D'autres ont montré une activité plus ou moins importante mais systématiquement inférieure à celle des isoformes de *Fut2*. De manière surprenante, le variant *Sec1* correspondant à

l'haplotype v5 est actif. Ces résultats indiquent que tous les animaux que nous avons étudiés possédaient au moins un gène codant une enzyme capable de synthétiser le ligand du RHDV (fig.3).

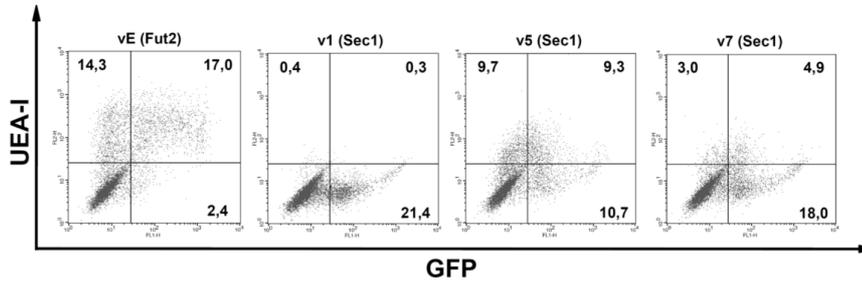


Figure 3 : Synthèse du ligand du RHDV par les produits d'allèles *Fut2* et *Sec1* du lapin de garenne. Après transfection de cellules CHO par un vecteur permettant l'expression de protéines de fusion GFP-fucosyltransférase, la synthèse de l'antigène H type 2 et mesurée par cytométrie en flux à l'aide de la lectine UEA-I. La fluorescence de la GFP permet de mesurer l'efficacité de la transfection. Les valeurs indiquées dans les quadrants supérieurs droite et gauche donnent le pourcentage de cellules exprimant le ligand du RHDV à leur surface. Les valeurs indiquées dans les quadrants de droite donnent le pourcentage de cellules exprimant les protéines chimériques (GFP-fucosyltransférases). La figure montre les résultats obtenus pour des formes enzymatiques caractéristiques : Fut2 vE (forte activité), Sec1 v5 (activité modérée), Sec1 v7 (activité faible), Sec1 v1 (activité nulle).

4. DISCUSSION

Différents calicivirus se fixent sur des glycannes de la surface des cellules épithéliales qui constituent des portes d'entrée pour les virus. Ces glycannes appartiennent à ce qu'on nomme la famille des antigènes de groupe sanguin tissulaires dont l'expression peut être polymorphe. Plusieurs souches de calicivirus humains (genre norovirus) nécessitent la présence de fucose en position $\alpha 1,2$ pour se fixer et de ce fait les individus qui en sont dépourvus (phénotype nonsécreteur) sont résistants à l'infection [2]. Nous avons antérieurement observé que le RHDV se fixait également sur une structure $\alpha 1,2$ fucosylée [9]. Par analogie, nous avons émis l'hypothèse que l'absence de ce ligand chez le lapin conduirait à une résistance au virus et que compte-tenu de la très forte pression de sélection exercée par le virus sur les populations de lapins de garenne, ce phénotype devrait être positivement sélectionné. En conséquence, nous avons recherché l'existence d'un équivalent du phénotype nonsécreteur chez les lapins de garenne. Il a effectivement été possible de le mettre en évidence, ce qui n'avait jamais

encore été décrit chez une espèce autre que l'homme. De manière très intéressante et en accord avec notre hypothèse, une nette association entre la fréquence de ce phénotype et l'impact de la maladie a été observée, suggérant que le virus sélectionne effectivement des animaux de type nonsécreteur-like. Cependant, la description du phénotype n'est que qualitative et partiellement arbitraire d'où la nécessité d'une description au niveau génotypique.

Chez le lapin existent trois gènes codant des α 1,2fucosyltransférases potentiellement fonctionnelles. Ces trois gènes sont localisés en tandem dans les génomes des espèces de mammifères chez lesquelles ils ont été localisés (homme, chimpanzé, macaque, porc, vache, rat, souris) [7]. Chez le lapin, le gène *Fut1* ne semble exprimé que dans le cerveau. Il est donc peu probable qu'il soit en cause dans le phénotype nonsécreteur-like. En revanche, les gènes *Fut2* et *Sec1* sont exprimés dans les tissus épithéliaux par lesquels le virus pénètre [1]. Le séquençage des régions codantes de ces trois gènes chez des lapins de garenne français a montré un polymorphisme important de *Fut2* et *Sec1*. De manière frappante, le gène *Fut1* n'a montré qu'un polymorphisme très limité. De plus nous avons mis en évidence une importante conversion génique entre *Fut2* et *Sec1*, indiquant que ces trois paralogues évoluent de manière tout à fait différente. Le phénomène de conversion génique conduit tout à la fois à homogénéiser les paralogues (évolution concertée) et à augmenter leur variation [10]. L'observation de ce phénomène pour les gènes *Fut2* et *Sec1* est tout à fait en accord avec leur rôle présumé dans des phénomènes d'adaptation à l'environnement et tout particulièrement à la diversité des agents pathogènes. Ainsi chez l'homme, les glycannes synthétisés par l'intermédiaire de FUT2 interviennent comme récepteur pour les calicivirus mais également pour plusieurs types bactéries comme des souches uropathogènes de *Escherichia coli* ou pour des souches d'*Helicobacter pylori*, responsables d'ulcères gastroduodénaux (pour une revue voir [6]).

L'analyse de la fréquence des haplotypes *Fut2* et *Sec1* dans deux populations de lapins de garenne français montre une différenciation modérée de ces populations, mais sensiblement supérieure à celle détectée au sein des populations françaises à l'aide de marqueurs microsatellites [8]. Pour la population de Chèvreloup, on a analysé des individus nés avant l'épidémie. D'autre part, l'impact de la RHD a été nettement plus faible à Cerisay et on n'y a pas observé de différence de fréquence significative entre le groupe des animaux mort et celui des survivants. Il est donc très vraisemblable que cette différenciation génétique observée entre les deux sites soit indépendante de la maladie (RHD). Il sera intéressant de définir la fréquence des haplotypes *Fut2* et *Sec1* dans d'autres populations françaises

car ils pourraient s'avérer utiles pour la gestion et le suivi des populations de lapins de garenne.

Dans la population de Chèvrelop, particulièrement affectée par deux épidémies de RHDV en 1995 et 1996, il a été possible de montrer qu'un haplotype *Sec1* était associé à la survie des animaux. L'analyse fonctionnelle des enzymes codées par les différents haplotypes *Fut2* et *Sec1* n'a pas permis de corréler l'existence d'un déficit d'activité enzymatique et donc de synthèse de ligand pour le virus et la survie des animaux. Les variants enzymatiques *Sec1* sont peu ou pas fonctionnels comparés aux variants *Fut2* qui sont tous efficaces pour synthétiser le ligand du RHDV. De ce fait, les deux gènes étant exprimés dans les mêmes tissus épithéliaux, le phénotype nonsécréteur-like ne peut pas être dû aux variants enzymatiques *Sec1*. Il est donc probable que ce phénotype ainsi que l'association observée entre un haplotype *Sec1* et la survie lors d'épidémies de RHDV sont dus à une ou plusieurs mutations localisée(s) dans une région régulatrice de *Sec1* et/ou de *Fut2*. Les résultats décrits ci-dessus nous encouragent à rechercher ces mutations.

REMERCIEMENTS

Ce travail a été cofinancé par le BRG, l'INSERM et l'ONCFS.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Hitoshi S., Kusunoki S., Kanazawa I., Tsuji S., 1996. - Molecular cloning and expression of a third type of rabbit GDP-L-fucose: β -D-galactoside 2- α -L-fucosyltransferase. *J. Biol. Chem.*, 271: 16975-16981.
- [2] Le Pendu J., Ruvöen-Clouet N., Kindberg E. Svensson L., 2006. - Mendelian resistance to human norovirus infections. *Seminars in Immunology*, 18: 375-386.
- [3] Marcato P., Benazzi C., Vecchi G., Galeotti M., Salda L. D., Sarli G., Lucidi P., 1991. - Clinical and pathological features of viral haemorrhagic disease of rabbits and the European brown hare syndrome. *Rev. Sci. Tech. Off. Int. Epiz.*, 10: 371-392.
- [4] Marchandeau S., Portejoie Y., Barraud S., Chaval, Y., 1998. - Impact of viral hemorrhagic disease on a wild population of European rabbits in France. *J. Wildl. Dis.*, 34: 429-435.
- [5] Marchandeau S., 2000. Enquête nationale sur les tableaux de chasse à tir saison 1998-1999. Le lapin de garenne. *Faune Sauvage*, 251: 18-25.
- [6] Marionneau S., Cailleau-Thomas A., Rocher J., Le Moullac-Vaidye B., Ruvöen-Clouet N., Clément M. Le Pendu, J., 2001. - ABH and Lewis histo-blood group

- antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie*. 83: 565-573.
- [7] Oriol R., Candelier J.-J., Mollicone R., 2000. - Molecular genetics of H. *Vox Sang.*, 78(Suppl.2): 105-118.
- [8] Queney G., Ferrand N., Weiss S., Mougel F. Monnerot M., 2001. - Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Mol. Biol. Evol.*, 18: 2169-2178.
- [9] Ruvoën-Clouet N., Ganière J. P., André-Fontaine G., Blanchard D., Le Pendu, J., 2000. - Binding of Rabbit Hemorrhagic Disease Virus to antigens of the ABH histo-blood group family. *J. Virol.*, 74: 11950-11954.
- [10] Teshima K.-M., Innan H., 2004. - The effect of conversion on the divergence between duplicated genes. *Genetics*, 166: 1553-1560.

Polymorphisme de gènes impliqués dans l'acquisition et la gestion de l'énergie chez l'huître creuse *Crassostrea gigas* : relations avec des caractères d'intérêt aquacole

Arnaud HUVET^{(1)*}, Dario MORAGA⁽²⁾, Elise DAVID^(2,3),
Anne-Leila MEISTERTZHEIM⁽²⁾, Arnaud TANGUY⁽⁴⁾,
Alain Van WORMHOUDT⁽⁵⁾, Jeanne MOAL⁽¹⁾,
Fanny JEFFROY^(1,2), Christopher SAUVAGE⁽⁶⁾,
Sylvie LAPÈGUE⁽⁶⁾, Pierre BOUDRY^(1,6)

⁽¹⁾Ifremer, UMR M100 Physiologie et Ecophysiologie des Organismes Marins,
Centre de Brest, BP70, 29280 Plouzané, France

⁽²⁾LEMAR, UMR-CNRS 6539, Institut Universitaire Européen de la Mer, Université
de Bretagne Occidentale, Place Nicolas Copernic, 29280, Plouzané, France

⁽³⁾Ecologie Microbienne, UMR-CNRS 5557, USC-INRA 1193, Université Lyon 1,
Bâtiment André Lwoff, 43 Bd du 11 Novembre 1918,
69622 Villeurbanne Cedex, France

⁽⁴⁾UMR CNRS 7144, Evolution et Génétique des Populations Marines,
Station Biologique, BP 74, Place Georges Teissier, 29682 Roscoff, France

⁽⁵⁾UMR 5178, CNRS-Muséum National d'Histoire Naturelle,
Station de Biologie Marine, 29900 Concarneau, France

⁽⁶⁾Ifremer, Laboratoire de Génétique et Pathologie, 17390 La Tremblade, France

Abstract: Polymorphism of nutrition and energy metabolism related-genes in the cupped oyster *Crassostrea gigas* : implication in important traits for aquaculture. Knowledge about bivalve feeding strategies and energetic allocation is important for aquaculture, as are the effects of these factors on different phenotypic traits such as growth, reproduction or survival. In the present study, polymorphism of ten genes implicated either in digestion (amylase A and B) or in energetic metabolism (glutamine synthetase, delta-9 desaturase, phosphoglucomutase, hexokinase, glucose-6-phosphatase, glycogen synthase and phosphorylase, sodium glucose transporter) was characterized. A very high level of DNA polymorphism was revealed in the order of magnitude of the highest levels reported in animals to date. 80 single nucleotide polymorphisms were characterized. 72 % of them appeared localized in exons and 36 % of them appeared to be non synonymous mutations leading to modification of the protein. Average density was estimated to be one single nucleotide polymorphism every 40 base pairs with the highest value observed for hexokinase: one every 14 base pairs. Polymorphism analysis revealed significant allelic frequency variations between

* Correspondance et tirés à part : ahuvet@ifremer.fr

wild populations showing contrasted levels of contamination along the French Atlantic coast, suggesting a potential selective effect of environmental stress, especially noticed for the phosphoglucosyltransferase marker. The use of phosphoglucosyltransferase as a possible marker to monitor stress exposure in disturbed ecosystems is therefore suggested. Polymorphism was also analyzed in successive generations of families selected for high or low survival to summer survival. For the glutamine synthetase and the delta-9 desaturase markers, some alleles were counter-selected in families showing high mortality rates. Any selective effect of summer mortality on glutamine synthetase C allele or delta-9 desaturase B allele could be mediated directly or via linkage to other loci involved in pathways related to susceptibility. Lastly, neutrality of amylase genes polymorphism in oyster was tested by examining relationships between amylase genotypes and growth within five families bred to be heterozygote at two amylase markers. Significant differences in growth were observed between genotypes suggesting that this polymorphism is not neutral and might be under selection due to its role in digestive function. Estimated daily yields were different between amylase genotypes, indicating the potential value of amylase markers in selective breeding programs to improve oyster growth. Investigating ecophysiological behavior, biochemical and molecular expressions of amylase in these genotypes revealed suggestive associations between growth, food-consumption related traits and amylase expression with amylase genes polymorphism. These results illustrate the great potential of studying gene polymorphism to further document selective processes in the wild and for potential application in marker assisted breeding programs.

bivalve/ energy metabolism/ gene/ growth/ pollutant/ polymorphism/ SNP

Résumé : Nous avons recherché du polymorphisme sur dix gènes impliqués dans la digestion ou dans le métabolisme énergétique chez l'huître creuse *Crassostrea gigas* afin de mettre en évidence des effets sélectifs et/ou des relations entre polymorphisme et fonctions physiologiques. Un nucléotide polymorphe (SNP) est observé en moyenne toutes les 40 paires de bases et 36 % des polymorphismes exoniques sont non synonymes, le polymorphisme le plus fort ayant été observé sur la séquence hexokinase. Les analyses en populations naturelles ont révélé une différenciation génétique significative entre sites présentant des niveaux de pollution contrastés, suggérant l'intérêt de marqueurs tels que la phosphoglucosyltransferase pour caractériser les pressions de sélection dues aux polluants. De plus, la contre-sélection de certains allèles des gènes de la glutamine synthétase ou la delta-9 désaturase a été montrée au sein de deux générations successives de familles sélectionnées pour leur sensibilité à la mortalité estivale. Enfin, une relation entre le polymorphisme de l'amylase et la croissance a été clairement établie, au sein de familles hétérozygotes aux marqueurs amylase, suggérant l'utilité des marqueurs amylase dans des programmes de sélection pour améliorer la croissance de l'huître. L'ensemble de nos résultats souligne l'intérêt, à la fois en terme de génétique des populations et d'application aquacole, de ce type d'étude chez une espèce présentant de nombreux cas de polymorphisme non neutres.

bivalve/ croissance/ métabolisme énergétique/ pollutant/ polymorphisme/ SNP

1. INTRODUCTION

Chez les bivalves marins, la connaissance des stratégies d'acquisition de l'énergie par l'alimentation (dépendantes des capacités digestives de l'animal) et celles de gestion et d'utilisation de cette énergie (dépendantes des conditions environnementales) est d'un intérêt majeur du fait de l'importance économique et écologique de ces espèces. En effet, de l'allocation de l'énergie vont dépendre des fonctions physiologiques majeures telles que la croissance, la reproduction et la survie. Plusieurs études de génétique quantitative ont montré qu'il existe une variabilité génétique importante pour certains de ces caractères chez l'huître creuse *Crassostrea gigas* [19], [9], [2]. L'analyse récente du phénomène complexe des mortalités estivales de *C. gigas* [29], caractère présentant une forte héritabilité [8]), a notamment mis en évidence le rôle du métabolisme énergétique et, plus précisément des glucides, dans ces mortalités à déterminisme multifactoriel [28]. C'est pourquoi l'étude du polymorphisme de gènes codant pour des enzymes impliquées dans les voies énergétiques et leurs conséquences sur la fonction revêt un intérêt essentiel pour une meilleure compréhension de la variabilité de ces fonctions physiologiques et, à terme, l'amélioration de caractères d'intérêt en aquaculture comme la survie estivale du naissain.

De nombreux exemples de non neutralité du polymorphisme de gènes de la digestion et du métabolisme énergétique existent dans le règne animal. Des relations ont par exemple été observées entre des génotypes amylase et la préférence nutritionnelle chez des crustacés [14] ou encore la conversion nutritionnelle et la croissance chez le poulet [15]. De même, des travaux récents ont montré, chez des espèces comme la drosophile, que le polymorphisme de gènes codant pour de nombreuses enzymes impliquées dans le métabolisme des sucres (*e.g.* glycogène synthase, hexokinase, phosphoglucomutase) pouvait affecter les performances physiologiques de l'animal dont son métabolisme énergétique et sa respiration [23].

Ce travail sur l'huître creuse *Crassostrea gigas* avait pour objectif de caractériser le polymorphisme de gènes « candidats » codant pour des enzymes impliquées dans des voies énergétiques. Les Single Nucleotide Polymorphism (SNP) constituent la forme de variation la plus fréquente au sein des génomes [27]. Leur abondance et leur stabilité supérieures aux microsatellites en font des marqueurs co-dominants de choix. Les SNP sont une forme de polymorphisme engendré par une mutation ponctuelle qui crée différents allèles à un même locus. Seuls quelques SNPs étaient jusqu'à présent caractérisés chez les bivalves marins, chez l'huître *C. gigas* [5] et chez la coquille Saint Jacques *Patinopecten caurinus* [11]. Les recherches *in silico* ont montré le fort potentiel de cette approche avec par exemple quelques

milliers de SNPs découverts chez l'huître américaine *C. virginica* [26]. Dans notre étude, la recherche de polymorphisme a été réalisée au sein de familles issues d'une génération de sélection pour la survie estivale [29], puis au sein de populations naturelles échantillonnées dans des sites contrastés pour leur niveau de pollution afin d'étudier l'effet de facteurs environnementaux potentiellement sélectifs sur la composition génétique des populations. En complément, le troisième objectif de ce travail était la recherche et l'identification de polymorphismes fonctionnels, c'est-à-dire qui se traduisent directement par une modification quantifiable d'un phénotype. Pour cet objectif, l'effort a été essentiellement porté sur le polymorphisme des gènes amylase en relation avec la croissance et du gène phosphoglucomutase en relation avec la survie estivale.

2. MATÉRIEL ET MÉTHODES

2.1. Recherche et caractérisation de polymorphisme

La recherche de polymorphisme pour les gènes A et B de l'amylase (*AMYA*, *AMYB*; [32]), la glucose 6 phosphatase (*G6P*), la glycogène synthase et phosphorylase (*GYS* et *GPH*; [1]), le sodium glucose transporteur (*SGLT*; [18]) et l'hexokinase (*HK*) a été réalisée par séquençage direct de produits de PCR. 24 huîtres partiellement apparentées (familles de demi-frères) issues du programme MOREST [29] ont été utilisées car cette recherche s'inscrivant en parallèle dans la mise en place d'une cartographie génétique impliquant ces individus. Cet apparentement nous a aidé à distinguer parmi les singletons les erreurs de séquençage du « vrai » polymorphisme partagé entre 24 individus apparentés issus de 12 géniteurs prélevés en population naturelle [30]. Les amplifications par PCR puis séquençage ont été réalisés suivant le protocole décrit par Sauvage *et al.*, [30] à partir de 100 ng d'ADN génomique extrait de tissus branchiaux. Les 24 séquences pour chaque gène ou EST ont été alignées en utilisant le logiciel Seqscapev 2.1 (Applied Biosystems) en utilisant l'algorithme KB. Les séquences ont été soumises à une recherche d'homologie (<http://www.ncbi.nlm.nih.gov/BLAST>) afin de vérifier si la séquence correspondait à celle attendue. Dans le cas contraire, les séquences n'ont pas été utilisées.

Pour 3 autres gènes - phosphoglucomutase (*PGM*, [35], glutamine synthétase (*GS*, [34]) et delta-9 désaturase (*D9*, [4]) - la recherche de polymorphisme a été réalisée par PCR-SSCP sur des zones exoniques dans 3 générations successives de familles produites dans le cadre du programme MOREST [29]. L'ADN génomique de chaque individu (100 ng) a été amplifié par PCR puis les fragments obtenus ont été analysés sur gel de polyacrylamide et séquencés suivant le protocole détaillé dans David *et al.*

[7]. Dans le cas particulier de la PGM, une analyse par technique allozymique avait permis de mettre en évidence que les parents de familles sélectionnées « sensibles à la mortalité estivale » étaient tous porteurs des allèles 95 ou 100 (à l'état homozygote ou hétérozygote). Le séquençage des haplotypes, identifiés par PCR-SSCP, a permis de détecter les mutations diagnostiques entre ces allèles 95 et 100 au niveau de l'exon 6.

2.2. Analyse du polymorphisme au sein de familles sélectionnées

Suite à une expérience de sélection divergente pour la résistance aux mortalités estivales au stade naissain [29], des familles des générations 1, 2 et 3 caractérisées comme sensibles ont été étudiées aux locus *PGM*, *GS* et *D9*. La première génération (G1) est constituée de 43 familles bi-parentales, structurée en 17 familles de demi-frères [8]. Les générations suivantes ont été produites par croisements multi-parentaux entre individus G1, ne permettant pas d'estimer les fréquences attendues au sein de ces « lots ». Nous avons donc comparé les fréquences génotypiques (1) attendues et observées (familles bi-parentales de première génération) ; (2) avant et après mortalité ; ou (3) entre sites d'élevage affecté ou non par la mortalité estivale. Le schéma de croisement des familles étudiées est présenté (fig.1).

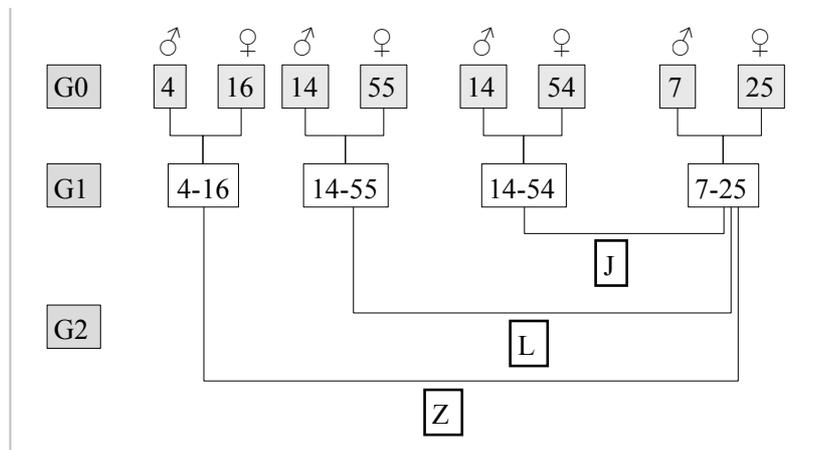


Figure 1 : Familles MOREST (sélectionnées Résistantes et Sensibles à la mortalité estivale ; Samain & McCombie, 2007[29]) utilisées dans cette étude, Les parents (G0) sont identifiés par des nombres, Les familles G1 sont identifiées par « x-y » où x est le mâle et y la femelle utilisés pour le croisement, Les lots G2 sont identifiés par les lettres J, L et Z,

2.3. Analyse du polymorphisme en populations naturelles

Le polymorphisme des gènes *GS*, *D9*, *PGM* et *HK* a été analysé pour 7 populations naturelles échantillonnées en janvier 2005 dans les estuaires de

la Vilaine, de la Gironde, de la Loire ainsi que dans l'estuaire du Belon, considéré comme un estuaire de référence car sous faible influence anthropique (fig. 2).

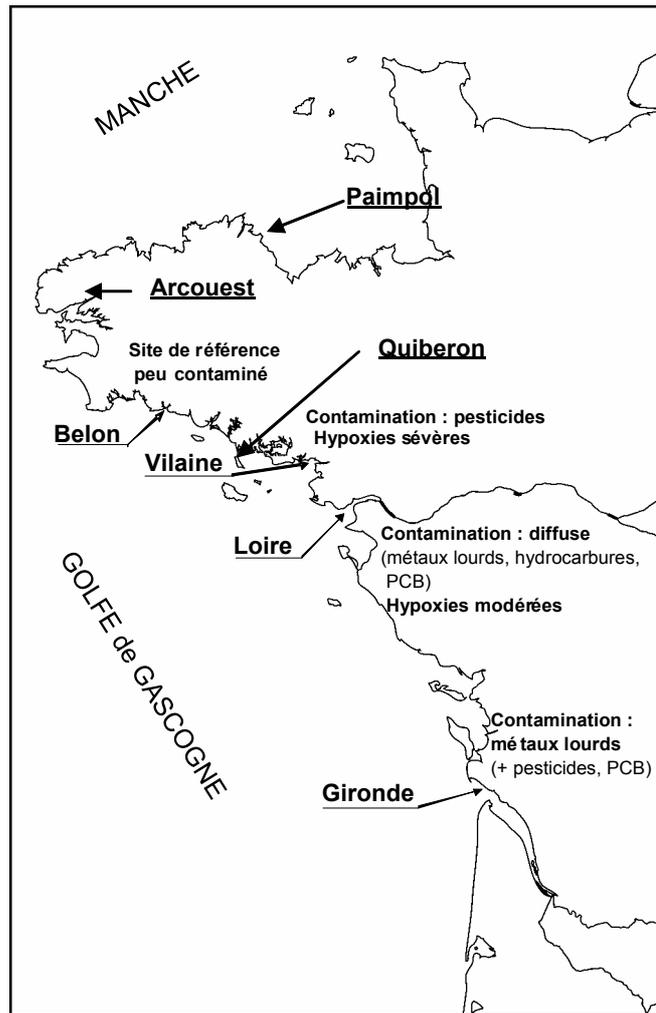


Figure 2 : Localisation des différents sites de prélèvement sur la façade Atlantique, avec leurs caractéristiques majeures en terme de contamination chimique et des risques hypoxiques.

L'estuaire de la Vilaine est caractérisé par des teneurs élevées en pesticides, celui de la Gironde par une pollution de type métallique, et celui de la Loire par une contamination diffuse et complexe de micropolluants chimiques [22]. A ces populations s'ajoutent 3 autres populations localisées dans des sites

présentant une typologie de type océanique, peu pollués, et localisés sur un gradient Nord-Sud de la Bretagne : Arcouest (Paimpol, Côtes d'Armor), Squiffiec (Rade de Brest, Finistère) et Quiberon (Morbihan). L'analyse du polymorphisme a été réalisée par la technique de PCR-SSCP sur les mêmes exons de chacun des gènes retenus pour l'analyse des familles sélectionnées. Les migrations des produits d'amplification ont été effectuées sur des gels d'acrylamide en condition non dénaturante. Par la suite, seuls les allèles présentant des variations significatives de fréquences entre les populations ont été clonés et séquencés afin de localiser et d'identifier les mutations.

2.4. Relation entre polymorphisme aux gènes de l'amylase et caractères physiologiques

Cinq familles bi-parentales, produites pour être hétérozygotes aux marqueurs PCR-RFLP des gènes de l'amylase, ont été analysées pour leur croissance et leur survie durant une année en condition standard d'élevage dans deux sites (Auray en Bretagne Sud et Baie des Veys en Normandie) suivant le protocole décrit par Prudence *et al.* [25]. 150 huîtres par famille ont été échantillonnées, pesées (poids total et de chair humide) et génotypées par PCR-RFLP suivant Sellos *et al.* [32] au début et à la fin de la période d'étude.

Pour la famille 2 pour laquelle les différents génotypes ont montré les plus fortes différences de poids, 800 individus ont ensuite été génotypés afin de disposer de 200 huîtres par génotype (G1 : *AMYA1/1, AMYB1/2* ; G2 : *AMYA1/2, AMYB2/2* ; G3 : *AMYA1/5, AMYB1/6* ; G4 : *AMYA2/5, AMYB2/6*). Ces huîtres ont été mises en conditionnement à la station Ifremer d'Argenton sous 2 régimes alimentaires : 80 % *Chaetoceros gracilis* / 20 % *Rhodomonas salina*, et 80 % *R. salina* / 20 % *C. gracilis* sachant que *R. salina* contient 3,4 % (équivalent poids sec) d'amidon, le substrat de l'amylase et que *C. gracilis* en est dépourvue.

Pour chaque génotype et condition, les paramètres écophysologiques ont été mesurés [21] : l'ingestion ($\text{cellule.h}^{-1}.\text{g}^{-1}$), l'efficacité d'absorption (exprimé en %) et la respiration ($\text{mg O}_2.\text{h}^{-1}.\text{g}^{-1}$) (paramètres définis [31]) pour obtenir 10 jours de données continues par génotype. Ces valeurs ont été converties en valeurs énergétiques, en utilisant les coefficients de conversion 0,32 μJ par cellule de *C. gracilis*, 1,58 μJ par cellule de *R. salina* et 13,8 J/mg O_2 , afin d'obtenir une estimation de l'Espérance de Croissance ou Scope For Growth (SFG exprimé en $\text{J. h}^{-1}.\text{g}^{-1}$ [2]).

A la fin de l'expérimentation, 30 huîtres par génotype ont été échantillonnées, pesées (poids total et de chair humide) et leur glande digestive disséquée, pesée et broyée à froid pour analyses enzymatique et moléculaire. L'activité amylasique et la constante apparente de Michaelis-

Menten (K_M) ont été mesurées [20]. L'extraction des ARN totaux, le traitement DNase, la rétro-transcription en ADNc et la mesure de l'expression des deux gènes de l'amylase ont été faits selon le protocole décrit [18] et normalisé par le niveau d'ARNm constant ($P = 0,134$, $CV = 2,7\%$) du facteur d'élongation I (amorces [13]).

Toutes les analyses statistiques, dont les tests de normalité, ont été réalisées avec le logiciel Statgraphics par analyse de variance suivie de comparaison multiple avec le test Tukey HSD. Les valeurs de survie ont été traitées après transformation angulaire. Les comparaisons des fréquences génotypiques ont été faites par test χ^2 .

3. RÉSULTATS

3.1. Recherche et caractérisation du polymorphisme

Pour la recherche de SNPs, un total de 2 kilobases a été séquencé pour 7 gènes dont 550 pb dans des zones non codantes et 1370 paires de bases dans des zones codantes. 80 SNP ont été caractérisés chez les 24 individus étudiés (descendants de 12 individus sauvages), dont 61 % sont des transitions et les 39 % restants des transversions. Ils s'intègrent dans une caractérisation plus large de SNPs publiés par Sauvage *et al.* [30] à l'exception de ceux identifiés sur *PGM*, *GS* et *D9* (Tanguy *et al.*, en préparation). Ces SNPs n'ont pas été retrouvés dans d'autres populations car aucun autre groupe de séquences n'a été soumis pour ces gènes dans GenBank. Sur ces 80 SNP, 58 ont été observés dans des régions codantes parmi lesquels 21 sont apparus être des marqueurs non synonymes (36 %) (tabl. I). 22 SNP ont été observés en zones non codantes. Les intervalles sont apparus très variables en fonction du gène étudié : 1 SNP est observé sur 185 pb analysées de la séquence *GPH* alors que 25 SNPs l'ont été sur 404 pb de la séquence *GYS* et jusqu'à 11 SNP sur 127 pb de la séquence exonique 7 de *HK*.

Pour 3 autres gènes (*PGM*, *D9*, *GS*), la recherche de polymorphisme a d'abord été effectuée par PCR-SSCP sur des familles sélectionnées sensibles aux mortalités estivales. Le séquençage des différents variants a permis de mettre en évidence 3 allèles pour l'exon 6 de la *GS* (allèle A : C en position 1014, C en position 1017, C en position 1027 et A en position 1035 ; allèle B : C en position 1014, C en position 1017, T en position 1027 et C en position 1035 et allèle C : T en position 1014, G en position 1017, T en position 1027 et C en position 1035), 2 allèles pour l'exon 6 de la *PGM* avec 3 SNP (allèle 95 : T en position 167, G en position 168 et C en position 173 et allèle 100 : C en position 167, A en position 168 et T en position 173) et

pour la *D9*, 2 allèles ont été observés (allèle A : C en position 151 et A en position 157 et allèle B : T en position 151 et T en position 157).

Tableau I : Caractérisation moléculaire des SNPs pour 10 gènes d'huître.

Gènes (abréviation)	N°Accession	Position du SNP (allèle le plus commun en premier)	Codant/ Non codant	Type	Synonyme/ Non synonyme
Amylase Gène B (AMYB)	AJ496603	170: T/C	C	Transition	S
		179: G/T	C	Transversion	NS
Glucose 6 Phosphatase (G6P)	AM076951	47: T/C	C	Transition	S
		64: G/C	NC	Transversion	-
		91: T/C	NC	Transition	-
		120: A/G	NC	Transition	-
		127: A/G	NC	Transition	-
Amylase Gène A (AMYA)	AJ496597	27: A/G	C	Transition	S
		45: A/G	C	Transition	S
		48: A/C	C	Transversion	S
		85: A/G	C	Transition	NS
		130: C/T	C	Transition	NS
		134: A/C	C	Transversion	NS
		135 T/C	C	Transition	S
		151: T/A	C	Transversion	NS
		207 A/T	C	Transversion	NS
Glycogène Synthase (GS)	AY496064	234: T/A	C	Transversion	S
		8: G/A	NC	Transition	-
		28: G/C	NC	Transversion	-
		78: T/A	NC	Transversion	-
		94: C/T	NC	Transition	-
		96: A/G	NC	Transition	-
		110: G/T	C	Transversion	NS
		118: A/G	C	Transition	NS
		142: T/C	C	Transition	NS
		147: A/T	C	Transversion	NS
		178: G/A	C	Transition	NS
		184: A/G	C	Transition	NS
		201: A/G	C	Transition	S
		254: T/C	NC	Transition	-
		255: A/G	NC	Transition	-
		257: A/G	NC	Transition	-
		258: A/G	NC	Transition	-
		261: T/G	NC	Transversion	-
		285: T/C	NC	Transition	-
		291: A/G	NC	Transition	-
302: G/C	NC	Transversion	-		
312: G/T	NC	Transversion	-		
336: C/G/T	NC	Transversion	-		
354: A/G	NC	Transition	-		
362: T/A	NC	Transversion	-		
372: T/G/C	NC	Transversion	-		
Sodium Glucose Cotransporteur (SGT)	AY551098	15 : G/C	C	Transversion	S
		34 : G/C	C	Transversion	NS
		35 : A/C/T	C	Transversion	NS
		36 : G/A	C	Transition	NS
		44: T/C	C	Transition	S
		92: T/C	C	Transition	S
95 : G/C	C	Transversion	S		
Glycogène Phosphorylase (GPH)	AY496065	86: A/T	C	Transversion	NS

Glutamine synthetase (GS)	CB617403	1014 : C/T	C	Transition	S
		1017 : C/G	C	Transversion	S
		1027 : C/T	C	Transition	NS
		1035 : A/C	C	Transversion	S
Delta-9 desaturase (D9)	CX069227	151 : C/T	C	Transition	S
		157 : A/T	C	Transversion	S
Phosphoglucomutase (PGM)	AJ512213	167: T/C	C	Transition	NS
		168: G/A	C	Transversion	NS
		173: C/T	C	Transition	NS
Hexokinase (HK)	AM076954	9333: A/G	C (Exon 7)	Transition	S
		9336: A/G	C	Transition	S
		9342: A/G	C	Transition	S
		9345: A/G	C	Transition	S
		9848: G/T	C	Transversion	S
		9357: C/T	C	Transition	S
		9366: A/G	C	Transition	NS
		9378: A/G	C	Transition	S
		9384: A/G	C	Transition	S
		9405: A/G	C	Transition	S
		9414: C/T	C	Transition	S
		9761 : A/G	C (exon 8)	Transition	S
		9773 : A/G	C	Transition	S
		9779 : A/G	C	Transition	S
		9787 : A/G	C	Transition	S
		9791 : A/T	C	Transversion	S
		9800 : C/T	C	Transition	S
		9815 : A/G	C	Transition	S
		9822 : C/T	C	Transition	S
9824 : A/C	C	Transversion	S		
9836 : C/G	C	Transversion	S		

3.2. Analyse du polymorphisme en populations naturelles et dans les familles sensibles aux mortalités estivales

Les distributions de fréquences génotypiques aux marqueurs *GS* et *D9* avant et après mortalité sont présentées dans David *et al.* [7]). Aucune différence n'a été observée en première génération entre fréquences génotypiques attendues et observées après mortalité sauf pour *D9* dans la famille F7-25 sur le site d'Auray, où les hétérozygotes AB sont moins fréquents qu'attendu. En seconde génération, une différence significative a été observée pour le lot Z avec une diminution significative des fréquences au marqueur *GS* des génotypes porteurs de l'allèle C dans les 2 sites où les mortalités ont été fortes (Auray, Ronce-les-Bains). Pour le gène *D9*, les fréquences observées sont apparues différentes entre Auray et Baie-des-Veys pour les lots Z et J. Pour le lot Z, une diminution de fréquence des génotypes porteurs de l'allèle B a été relevée avec absence d'homozygote BB. Enfin pour la *PGM*, un déficit significatif des hétérozygotes 95/100 dans la famille G1 F7-25 au site d'Auray (forte mortalité : 56 %) ainsi que des déficits significatifs des homozygotes 95/95 et des hétérozygotes 95/100 pour les lots G2 J et Z ont été observés (tabl. II).

Tableau II : Fréquences observées des génotypes (porteurs des allèles 95 et/ou 100) identifiés pour le locus phosphoglucomutase [34] au cours des deux générations G1 et G2 avant et après mise sur site. * : écart significatif par rapport à la fréquence attendue. M= mortalité en %.

G1 (n=60) Juillet 2001 avant mortalité	100/100: 41 % 95/100: 59 %									
G1 (n=48) Octobre 2001 après mortalité	Génotype	Ronces Les Bains (M=57,6)			Auray (M=56) *			Baie des Veys (M=19,6)		
	100/100	33 %			54 %			44 %		
	95/100	67 %			46 %			56 %		
G2 (n=48) Octobre 2002 après mortalité	Génotype	L* M=30,4	J* M=41,2	Z* M=40,2	L* M=53,7	J* M=33,3	Z* M=64,2	L* M=15,2	J* M=39,1	Z* M=6,3
	100/100	61 %	73 %	80 %	68 %	84 %	76 %	53 %	55 %	67 %
	95/95	17 %	0 %	0 %	8 %	0 %	0 %	13 %	2 %	4 %
	95/100	22 %	27 %	20 %	24 %	16 %	24 %	34 %	43 %	29 %

En élargissant cette analyse aux populations naturelles, 6, 4 et 11 allèles supplémentaires ont été caractérisés pour les marqueurs *D9*, *GS* et *PGM*, respectivement. Pour le gène *GS*, l'allèle C est plus fréquent (de 8 à 19 %) dans les estuaires de la Loire et de la Vilaine que dans les autres populations (tabl. III). Pour le gène *D9*, l'allèle B est apparu moins fréquent (de 8,3 %) dans l'estuaire de la Vilaine que dans l'estuaire du Belon. Enfin pour *PGM*, on observe une infériorité (de 7 à 16 %) des fréquences de l'allèle 95 sur les estuaires contaminés par rapport à l'estuaire de référence, mais aussi entre l'estuaire de référence et 2 sites océaniques (Squiffiec et Quiberon). Pour *HK*, les fréquences des 3 allèles de l'intron 9 apparaissent très différentes dans les populations, avec la présence très majoritaire de l'allèle A (exemple 81 % à Belon, tableau III) ou B (exemple 72 % à Quiberon).

Tableau III : Fréquences alléliques dans les populations de *Crassostrea gigas* échantillonnées sur 4 estuaires (Gironde, Loire, Vilaine et Belon) et 3 sites océaniques (Arcouest, Squiffiec et Quiberon) pour 3 loci : *GS* glutamine synthétase, *D9* delta-9 désaturase et *PGM* Phosphoglucomutase. (N : nombre d'individus, Ho : hétérozygotie observée, He : hétérozygotie attendue, *Fis* par population). En rouge sont représentées les valeurs *Fis* significatives. / non mesuré.

Locus	Allèle	Gironde	Loire	Vilaine	Belon	Arcouest	Squiffiec	Quiberon	
GS	A	0,056	0,014	0,028	0,069	0,090	0,064	0,083	
	B	0,500	0,319	0,389	0,472	0,399	0,401	0,443	
	C	0,431	0,639	0,542	0,458	0,468	0,471	0,417	
	D	0,000	0,014	0,042	0,000	0,011	0,052	0,021	
	E	0,000	0,000	0,000	0,000	0,016	0,000	0,021	
	F	0,014	0,014	0,000	0,000	0,016	0,012	0,010	
	G	0,000	0,000	0,000	0,000	0,000	0,000	0,005	
	He	0,561	0,489	0,553	0,562	0,613	0,610	0,623	
	Ho	0,333	0,472	0,417	0,583	0,638	0,628	0,604	
	<i>Fis</i>	0,418	0,049	0,26	-0,024	-0,036	-0,023	0,035	
D9	A	0,889	0,819	0,944	0,861	0,860	0,863	0,854	
	B	0,097	0,153	0,056	0,139	0,129	0,053	0,104	
	C	0,014	0,014	0,000	0,000	0,000	0,000	0,000	
	D	0,000	0,014	0,000	0,000	0,000	0,016	0,016	
	E	0,000	0,000	0,000	0,000	0,011	0,011	0,016	
	F	0,000	0,000	0,000	0,000	0,000	0,011	0,010	
	G	0,000	0,000	0,000	0,000	0,000	0,042	0,000	
	H	0,000	0,000	0,000	0,000	0,000	0,005	0,000	
		He	0,200	0,305	0,105	0,239	0,243	0,250	0,259
		Ho	0,222	0,361	0,111	0,278	0,237	0,232	0,292
	<i>Fis</i>	-0,096	-0,171	-0,045	-0,148	0,033	0,079	-0,121	
PGM	A	0,000	0,014	0,014	0,042	0,068	0,016	0,042	
	B	0,000	0,028	0,000	0,000	0,000	0,000	0,000	
	95	0,139	0,042	0,069	0,208	0,080	0,053	0,026	
	100	0,847	0,917	0,917	0,750	0,835	0,842	0,865	
	C	0,014	0,000	0,000	0,000	0,000	0,000	0,005	
	D	0,000	0,000	0,000	0,000	0,006	0,005	0,000	
	E	0,000	0,000	0,000	0,000	0,000	0,005	0,010	
	F	0,000	0,000	0,000	0,000	0,011	0,058	0,042	
	I	0,000	0,000	0,000	0,000	0,000	0,000	0,010	
	K	0,000	0,000	0,000	0,000	0,000	0,016	0,000	
	J	0,000	0,000	0,000	0,000	0,000	0,005	0,000	
		He	0,263	0,157	0,155	0,392	0,291	0,284	0,248
		Ho	0,250	0,111	0,111	0,250	0,057	0,221	0,167
	<i>Fis</i>	0,062	0,305	0,295	0,375	0,807	0,227	0,333	
HK	A	0,7500	0,3478	0,4375	0,8125	/	0,4333	0,2500	
	B	0,1875	0,5870	0,5417	0,1875	/	0,5500	0,7167	
	A'	0,0625	0,0652	0,0208	0	/	0,0167	0,0333	
	He	0,4069	0,540	0,5257	0,3112	0,4647	0,5181	0,4299	
	Ho	1,125	0,0435	0,1667	0,2917	0,3810	0,4330	0,4330	
	<i>Fis</i>	0,697	0,921	0,688	0,064	0,182	0,166	-0,008	
TOTAL									
	N	36	36	36	36	93	95	96	
	Ho	0,269	0,315	0,213	0,370	0,311	0,360	0,354	
	He	0,341	0,317	0,271	0,398	0,383	0,382	0,377	
	<i>Fis</i> total	0,227	0,021	0,223	0,083	0,193	0,061	0,065	

Les populations des 3 estuaires contaminés ont présenté des taux d'hétérozygotie multi-locus inférieurs ($0,213 < H_o < 0,315$) à ceux présentés par l'estuaire de référence du Belon, Squiffiec ou Quiberon ($0,354 < H_o < 0,370$), particulièrement pour le gène *HK* (tabl. III). Les déficits en hétérozygotes au niveau multi-locus observés pour les populations de Gironde et la Vilaine sont notamment générés par le gène *GS*.

Tableau IV : Indice de différenciation génétique (F_{st}) calculé par locus et par paire de populations, Les F_{st} significatif au seuil $\alpha=0,05$ sont indiqués en rouge, *GS* glutamine synthétase, *D9* delta-9 desaturase, *PGM* Phosphoglucomutase et *HK* Hexokinase.

<i>GS</i>	Loire	Vilaine	Belon	Arcouest	Squiffiec	Quiberon
Gironde	0,053	0,005	-0,015	0,001	0,001	-0,007
Loire		-0,002	0,040	0,026	0,023	0,047
Vilaine			0,000	-0,001	-0,005	0,008
Belon				-0,004	-0,003	-0,007
Arcouest					-0,003	-0,001
Squiffiec						0,000
<i>D9</i>	Loire	Vilaine	Belon	Arcouest	Squiffiec	Quiberon
Gironde	0,004	0,003	-0,006	-0,005	0,001	-0,004
Loire		0,047	-0,008	-0,004	0,017	-0,001
Vilaine			0,027	0,020	0,011	0,017
Belon				-0,009	0,010	-0,005
Arcouest					0,010	-0,003
Squiffiec						0,004
<i>PGM</i>	Loire	Vilaine	Belon	Arcouest	Squiffiec	Quiberon
Gironde	0,020	0,008	0,007	-0,001	0,009	0,021
Loire		-0,013	0,077	0,007	0,008	0,002
Vilaine			0,063	0,003	0,007	0,004
Belon				0,021	0,044	0,066
Arcouest					0,002	0,002
Squiffiec						-0,002
<i>HK</i>	Loire	Vilaine	Belon	Arcouest	Squiffiec	Quiberon
Gironde	0,227	0,168	-0,019	0,0265	0,181	0,371
Loire		-0,028	0,2896	0,1071	-0,0189	0,0005
Vilaine			0,2218	0,052	-0,0269	0,0424
Belon				0,048	0,230	0,4358
Arcouest					0,062	0,230
Squiffiec						0,1816

Le F_{st} multi-locus révèle une différenciation génétique significative entre certaines paires de populations (tabl. IV). Pour le locus *GS*, une différenciation génétique est observée entre les populations de la Loire et toutes les autres, exception faite de celle de la Vilaine. Cette différenciation correspond à une fréquence élevée des homozygotes CC pour la Loire, Vilaine et Gironde alors qu'elle est faible pour le Belon et les 3 populations océaniques. Pour *PGM*, les populations de la Loire, de la Vilaine, de Squiffiec et de Quiberon se différencient de celle du Belon. Pour *HK*, la différenciation génétique est apparue importante entre les différentes

populations, surtout entre la population Belon et les autres (sauf Gironde et Arcouest). En considérant la totalité des populations et tous les estuaires confondus, les *Fst* multi-locus montrent une différenciation génétique importante entre les populations de la Loire et de l'ensemble des autres populations, exception faite de celle de la Vilaine.

Enfin, précisons que les séquences des gènes *HK* et *PGM* [35] ont été caractérisées au cours de ce travail. La séquence partielle *HK* de 10961 pb (du site ATG à la queue polyA) est composée de 11 exons (**EU342886**) et code une protéine de 478 AA identique à la séquence ADNc (**AM076954**). La séquence du gène *PGM* fait 7071 pb. Elle est composée de 8 exons et correspond à une protéine déduite de 555 acides aminés (référence Genbank : **AJ512213**).

3.3. Relation entre polymorphisme aux gènes de l'amylase et caractères physiologiques

L'analyse des fréquences génotypiques au début de l'expérimentation au sein des 5 familles étudiées est en accord avec l'hypothèse d'une liaison des 2 gènes de l'amylase avec une distance estimée à 1,7 cM (voir [25]).

Au niveau intra-familial, des différences significatives de poids total et de chair ont été observées à T1 entre génotypes dans le site d'Auray en Bretagne Sud pour 2 des 5 familles (familles 2 et 3) [25] alors qu'aucune différence de mortalité n'a été observée entre génotypes quelle que soit la famille considérée. La différence maximale de poids est apparue d'environ 35 % entre les deux génotypes extrêmes (*A1/2, B2/2* vs *A1/5, B1/6*) de la famille 2 (fig. 3B, soit 4,9 g et 0,5 g pour le poids total et le poids de chair, respectivement) soit un rendement journalier moyen estimé à 3 % pour *A1/5, B1/6*, à 3,7 % pour *A1/2, B2/2*.

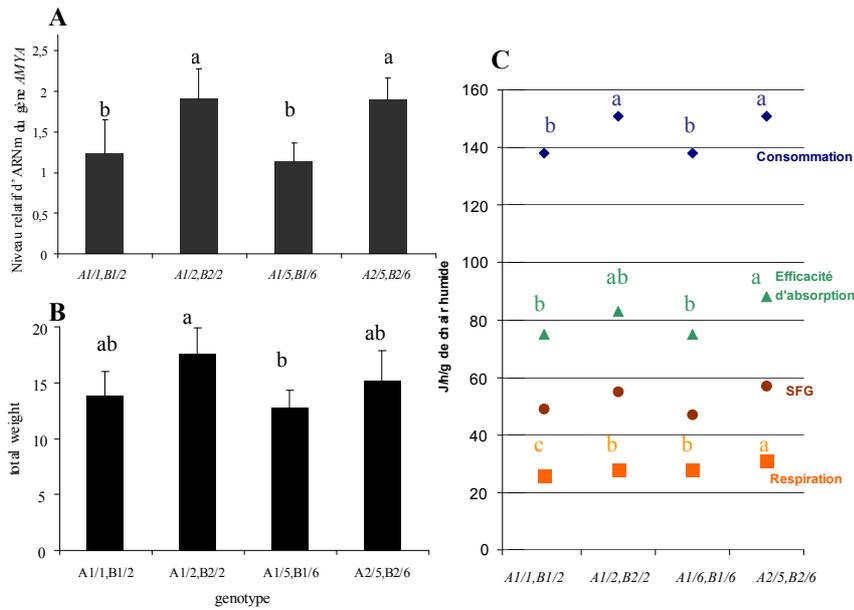


Figure 3 : Mesures phénotypiques pour 4 génotypes amylase d'une même famille d'huître (famille 2 de Prudence et al., 2006 [23]). Les lettres correspondent aux groupes homogènes à 5% obtenus par le test Tukey's HSD à partir de 30 mesures individuelles par génotype.

3A : Niveau relatif (unités arbitraires) d'ARNm du gène amylase *AMYA* qui est égale à la somme des niveaux relatifs d'ARNm des gènes amylase *AMYA* et *AMYB* (puisque la valeur obtenue pour *AMYB* est 400 fois inférieure à celle obtenue pour *AMYA*).

3B : Poids total (en gramme).

3C : Paramètres écophysiologiques : respiration, consommation algale (= ingestion), efficacité d'absorption et espérance de croissance (SFG) en joule/heure/gramme de chair humide.

Les lettres correspondent aux groupes homogènes à 5% obtenus par le test Tukey's HSD à partir de 30 mesures individuelles par génotype.

Au sein de la famille 2, des différences significatives sont observées entre génotypes pour les paramètres écophysiologiques (fig. 3C) et les niveaux d'ARNm des gènes amylase (fig. 3A). Les 2 génotypes à plus forte croissance (A1/2, B2/2 et A2/5, B2/6) ont également des valeurs d'ingestion, d'efficacité d'absorption, de consommation d'oxygène et de niveaux d'ARNm amylase, significativement plus élevées que celles des 2 génotypes à plus faible croissance (A1/1, B1/2 et A1/5, B1/6) (fig. 3C). Les valeurs maximales d'ingestion ou d'efficacité d'absorption varient d'environ 10 % entre génotypes extrêmes (exemple 4,72 cellules d'algues ingérées h⁻¹.g⁻¹ pour A1/2,

B2/2 et 4,32 cellules h⁻¹,g⁻¹ pour *A1/5*, *B1/6*), Deux groupes statistiques sont relevés pour les niveaux d'ARNm amylase : les ARNm amylase de *A1/2*, *B2/2* et *A2/5*, *B2/6* sont 2 fois plus présents que pour les 2 autres génotypes (fig.3A), Par contre, aucune différence significative entre génotypes n'a été observée pour les paramètres enzymatiques : activité et Km de l'amylase (données non présentées).

4. DISCUSSION

4.1. Polymorphisme nucléotidique élevé

Dans ce travail, nous avons recherché des polymorphismes sur un total de dix gènes codant pour des enzymes impliquées soit dans la digestion (les gènes A et B de l'amylase *AMYA*, *AMYB*) soit dans le métabolisme énergétique essentiellement glucidique (la glucose-6 phosphatase *G6P*, la glycogène synthase *GYS* et phosphorylase *GPH*, le Sodium Glucose Transporteur *SGT*, l'hexokinase *HK*, la phosphoglucomutase *PGM*, la glutamine synthétase *GS* dans le cycle de l'azote, et la delta-9 désaturase *D9* intervenant dans le métabolisme lipidique). Un total de 80 SNP a été caractérisé, ce qui correspond à un polymorphisme moyen de 1 SNP toutes les 40 paires de bases (pb). Les régions codantes réunissent 72 % des marqueurs. Parmi ces cSNP (SNP situés dans des régions codantes), 64 % sont des marqueurs synonymes alors que les autres 36 % sont des marqueurs non synonymes, c'est-à-dire modifiant la composition de la protéine en acide aminé et donc potentiellement sa conformation spatiale ou son activité dans le cas d'une enzyme. Enfin, dans les zones non codantes séquencées, le niveau de polymorphisme atteint 1 SNP toutes les 25 pb. Cependant, ce résultat est à relativiser car les séquences non codantes étudiées parmi ces 10 gènes sont relativement courtes et peu nombreuses et ne reflètent pas le niveau de polymorphisme moyen observé chez l'espèce. Seuls trois gènes (*GYS*, *AMYB*, *G6P*) ont révélé une zone intronique insérée dans la partie amplifiée par PCR et ont donc été séquencés. Le niveau de polymorphisme apparaît donc être de 1/60pb en partie codante et 1/40pb en partie non codante, quant on analyse davantage de séquences [30], ce qui corrobore l'intervalle d'environ 40 pb précédemment estimé au niveau des régions non codantes du génome de *C. gigas* [5]. Il est intéressant de noter que ce taux est très variable selon les gènes étudiés avec par exemple sur nos 7 gènes, 1 SNP sur 185 pb pour *GPH*, 1 SNP pour 16 pb pour *GYS*, la valeur maximale atteinte étant 1 SNP pour 14 pb pour *HK* en partie exonique. Ce niveau de polymorphisme de type SNP détecté chez *C. gigas*, est parmi les plus hauts décrits dans la bibliographie. Pour information, le plus haut niveau de polymorphisme a été mis en évidence chez le nématode

Caenorhabditis remanei [6]) et l'ascidie *Ciona savignyi* [33] avec en moyenne 1 SNP toutes les 20 pb.

4.2. Effets sélectifs suggérés dans les familles sélectionnées comme sensibles aux mortalités estivales et dans les populations naturelles

Après un suivi au cours d'un été dans trois sites d'élevage français, l'analyse des générations d'huîtres produites dans le cadre du programme MOREST (via un plan de sélection divergente sur le caractère survie estivale [9]) a révélé (1) des déviations de fréquences alléliques aux marqueurs *GS* et *PGM* en première génération par rapport à l'attendu après les mortalités subies sur sites, et (2) des différences de fréquences génotypiques entre sites ayant présenté des taux de mortalité contrastés. Parmi les différentes hypothèses [7], la sélection apparaît comme la plus probable, la mortalité subie au sein des familles aurait affectée préférentiellement certains génotypes. A Auray et Ronce-les-Bains où les mortalités ont été fortes, la fréquence des hétérozygotes porteurs de l'allèle *GS-C* est proche de 0 ce qui suggère de forts effets sélectifs contre cet allèle, ici à l'état hétérozygote. L'allèle *GS-C* est porteur de mutations synonymes et sa contre sélection pourrait donc refléter des effets indirects impliquant d'autres régions liées du génome [7]. De plus, les résultats obtenus au marqueur *D9* suggèrent également l'existence d'un effet délétère de l'allèle *D9-B* à l'état homozygote. A noter que ces deux gènes n'apparaissent pas liés [7]. Au marqueur *PGM*, les résultats obtenus ont à nouveau permis de suggérer un phénomène de sélection à l'encontre de l'allèle *PGM-95*.

La superposition de l'ensemble des données avec les taux de mortalité fait apparaître que les plus forts déficits en hétérozygotes sont associés aux taux de mortalité les plus élevés, ce qui est en contradiction avec une fitness supérieure des hétérozygotes souvent observée, et notamment pour la gestion des coûts énergétiques et leurs conséquences avantageuses sur la tolérance aux stress environnementaux (voir [7]). Quoiqu'il en soit, il est connu que des paramètres environnementaux stressants, tels que l'exposition aux polluants, peuvent réduire l'hétérozygotie aux marqueurs allozymiques [10].

L'analyse des populations naturelles a révélé une différenciation génétique importante aux marqueurs analysés (*Fst* moyen de 0,01, soit 1,7 fois plus élevé que le *Fst* moyen observé pour des marqueurs microsatellites entre populations de la côte atlantique française, [16]). Ceci est intéressant à noter car le fort flux génique entre populations françaises de *C. gigas* due notamment à l'activité ostréicole est supposé important. Pour certains marqueurs, une différenciation génétique locale plus importante peut être

observée, comme par exemple pour la population de la Loire qui apparaît significativement différenciée des autres populations (exception faite de celle de la Vilaine) au marqueur Glutamine synthétase. Pour *PGM*, les populations de la Loire, de la Vilaine, de Squiffiec et de Quiberon se différencient de celle du Belon. Certains génotypes reflètent cette différence, avec une contre sélection des allèles 100 et 95 dans la Loire, une sélection de l'allèle 100 à l'état homozygote dans les populations de la Vilaine (86 % des génotypes), de Squiffiec et de Quiberon (71 et 74 %) contre 63 % dans le Belon. En considérant la totalité des populations et tous les estuaires confondus, il apparaît une différenciation génétique importante entre les populations de la Loire et de l'ensemble des autres populations (sauf avec la Vilaine). Ces résultats sont à rapprocher de la typologie des sites échantillonnés. Ceci nous permet de proposer une sélection spécifique dans les populations estuariennes contaminées lié à une contamination diffuse et complexe (Loire) et/ou de très fortes teneurs en pesticides (Vilaine). Les populations estuariennes contaminées contrastent avec les autres populations, océaniques et estuariennes de référence, pour leurs forts déficits en hétérozygotes, notamment au locus hexokinase. Ces très forts déficits en hétérozygotes observés dans les populations estuariennes contaminées suggèrent, à l'instar des résultats obtenus sur les familles sélectionnées sensibles, que l'exposition aux polluants peut agir et contre sélectionner les hétérozygotes. Ce résultat reste à confirmer du fait du faible nombre d'allèles caractérisés, Des travaux complémentaires seront donc nécessaires afin d'apprécier la « valeur adaptative » de certains de ces polymorphismes en relation avec de fortes contaminations environnementales.

4.3. La non neutralité du polymorphisme aux gènes de l'amylase

Les différences significatives de poids total et de chair observées entre génotypes amylase de deux familles au site d'Auray suggèrent que le polymorphisme analysé est non neutre et potentiellement sous effet sélectif en relation avec le rôle de ce gène dans la digestion. Ces différences sont apparues fortes atteignant plus de 30 % entre les deux génotypes extrêmes de la famille 2 [25] ce qui souligne la valeur potentielle de ces marqueurs amylase pour une éventuelle sélection pour améliorer la croissance de l'huître.

Pour mieux comprendre cette corrélation croissance-polymorphisme aux gènes amylase, nous avons entrepris de mesurer les paramètres écophysologiques et l'expression des gènes amylase (au niveau moléculaire et enzymatique) sur les quatre génotypes de la famille 2 montrant les plus fortes différences de croissance. Deux conditions nutritives, variables pour l'apport en amidon (le substrat de l'amylase), ont été appliquées mais aucune différence significative n'est apparue entre régimes nutritifs

(données non présentées). Par contre, il apparaît clairement des différences significatives entre ces quatre génotypes, notamment pour l'ingestion et l'efficacité d'absorption, à l'origine de variations d'espérance de croissance entre génotypes, même si elles sont quelque peu modérées par les valeurs de consommation d'oxygène. Les génotypes à forte croissance présentent les plus fortes valeurs écophysiologicals, l'inverse étant également vrai : les génotypes à faible croissance présentant les valeurs écophysiologicals les plus faibles. Ceci suggère l'existence d'une association entre ce polymorphisme PCR-RFLP amylosique et le comportement alimentaire de l'huître. Toutefois, en appliquant les valeurs extrêmes d'ingestion ou d'efficacité d'absorption dans le modèle dynamique de budget énergétique (DEB) appliquée à l'huître *C. gigas* [24], il apparaît que seules les valeurs d'ingestion obtenues deviennent explicatives. En effet, en appliquant dans ce modèle l'ingestion pour les génotypes *A1/2*, *B2/2* ($4,72 \text{ cellules h}^{-1} \text{g}^{-1}$) et *A1/5*, *B1/6* ($4,32 \text{ cellules h}^{-1} \text{g}^{-1}$), nous obtenons une croissance modélisée totalement identique à celle observée après une année d'élevage [25].

Outre la prise alimentaire [3], l'efficacité d'absorption est habituellement considérée comme l'une des principales causes des variations de croissance observées chez *C. gigas* [17]. Un modèle, du type Michaelis-Menten, reliant l'équipement digestif (enzymes amylase et laminarinase) et les performances d'absorption, a été validé chez l'huître. Ainsi, il apparaissait très intéressant de caractériser les modalités d'expression et les propriétés catalytiques des enzymes digestives au sein de ces quatre génotypes. Les résultats montrent clairement que le niveau d'ARNm des gènes amylase est dépendant du polymorphisme analysé avec des différences très significatives entre génotypes. Le classement hiérarchique des génotypes pour les niveaux d'ARNm amylase, massivement représenté par le niveau d'ARNm du gène *AMYA*, apparaît identique à celui obtenu pour les efficacités d'absorption, l'ingestion ou la croissance suggérant des corrélations positives entre ces différents phénotypes. Par contre, aucune différence enzymatique significative n'a été observée entre génotypes (activité spécifique et constante d'affinité de l'amylase), corroborant le rôle moins important de l'absorption vu par la modélisation.

5. CONCLUSION

Le très fort niveau de polymorphisme nucléotidique observé chez l'huître permet de disposer d'un grand nombre de séquences codantes polymorphes pour des gènes d'intérêt impliqués dans le fonctionnement énergétique. Sachant notamment que 36 % de ce polymorphisme est non synonyme, ces résultats ouvrent de très nombreuses perspectives de

recherche de relations entre génotypes et phénotypes, d'effets sélectifs dans le milieu naturel ou d'élevage, et de co-localisation de gènes candidats et de zones QTLs pour des caractères dans lesquels ces gènes seraient susceptibles d'intervenir. Les perspectives d'applications de ce travail sont donc nombreuses à la fois en terme de suivi de la ressource génétique, et notamment face aux pressions anthropiques croissantes (pollution, changement global...), et aussi en terme de gestion de cette ressource essentiellement pour l'aquaculture.

Certains polymorphismes étudiés dans ce projet se sont révélés non neutres. En effet, la contre sélection de certains allèles des gènes de la glutamine synthétase ou la delta-9 desaturase a été montrée au sein de deux générations successives de familles sélectionnées pour leur sensibilité à la mortalité estivale, soulignant de nouvelles pistes de travail pour élucider les mécanismes complexes de la résistance aux mortalités estivales. Ajoutons que ce travail s'inscrit dans une démarche en cours qui vise à rechercher et caractériser des QTLs pour la résistance aux mortalités estivales (Sauvage *et al.*, en préparation). Pour ce qui est de la croissance, la production de familles polymorphes aux marqueurs PCR-RFLP des gènes de l'amylase a permis de montrer l'existence de relation entre ce polymorphisme et la croissance, suggérant l'utilité de ces marqueurs dans des programmes de sélection pour améliorer la croissance de l'huître. D'autres caractères, essentiellement l'ingestion, se sont révélés corrélés à cette relation ce qui supposerait que les génotypes amylase les plus performants en terme de croissance le sont de part une ingestion accrue. Ces données sont d'autant plus intéressantes qu'il s'agit d'une espèce dont l'élevage dépend directement de la productivité primaire naturelle des sites de production.

Enfin, les analyses en populations naturelles ont révélé une différenciation génétique relativement élevée entre sites présentant des niveaux de pollution contrastés, suggérant l'existence d'effets sélectifs à certains marqueurs comme la phosphoglucomutase. Ainsi, ce travail montre l'intérêt de marqueurs fonctionnels pour caractériser les pressions de sélection d'origine anthropiques, de plus en plus fortes dans le milieu marin.

REMERCIEMENTS

Sincères remerciements à Stéphane POUVREAU pour le travail en écophysiologie, à Jean-Yves DANIEL, Virgile QUILLIEN, Serge HEURTEBISE et à toute l'équipe de la station expérimentale Ifremer d'Argenton (spécialement à Pierrick LE SOUCHU et Isabelle QUÉAU) pour leur aide technique, Merci également à Isabelle BOUTET, Marie PRUDENCE, Hélène BACCA, Gilles LE MOULLAC et Marie-Thérèse THÉBAULT pour leur

concours précieux à ce travail. Ce travail a été co-financé par le Bureau des Ressources Génétiques (contrat « Polygigas » n°05/5210460/YF).

RÉFÉRENCES BIBLIOGRAPHIQUES

(* indique les publications résultant au moins en partie de ce travail),

- [1] Bacca H., Huvet A., Fabioux C., Daniel J.Y., Delaporte M., Pouvreau S., Van Wormhoudt A., Moal J., 2005. - Molecular cloning and seasonal expression of the oyster glycogen phosphorylase and glycogen synthase genes. *Comparative Biochemistry and Physiology*, 140B: 635-646.
- [2] Bayne B.L., Newell R.C., 1983. - Physiological energetics of marine molluscs. In: *The Mollusca*, 4 (KM Wilburg and ASM Saleuddin) ed, Academic Press, London, pp 407-515.
- [3] Bayne B.L., Hedgcock D., McGoldrick D., Rees R., 1999. - Feeding behaviour and metabolic efficiency contribute to growth heterosis in Pacific oysters *Crassostrea gigas* (Thunberg). *Journal of Experimental Marine Biology and Ecology*, 233: 115-130.
- [4] Boutet I., Tanguy A., Moraga D., 2004. - Response of the Pacific oyster *Crassostrea gigas* to hydrocarbon contamination under experimental conditions. *Gene*, 329: 147-157.
- [5] Curole J.P., Hedgcock D., 2005. - High Frequency of SNPs in the Pacific Oyster genome. http://intl-pag.org/13/abstracts/PAG13_W026.html.
- [6] Cutter A.D., Baird S.E., Charlesworth D., 2006. - High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics*, 174: 901-913.
- [7] *David E., Boudry P., Dégremont L., Tanguy A., Quéré N., Samain J.F., Moraga D., 2007a. - Study of genetic polymorphism of glutamine synthetase and delta-9 desaturase, in families of Pacific oyster *Crassostrea gigas*, during summer mortalities events. *Journal of Experimental Marine Biology and Ecology*, 349: 272-283.
- [8] Dégremont L., 2003. Etude des bases génétiques de la mortalité estivale et des relations avec la croissance chez les juvéniles de l'huître creuse *Crassostrea gigas*. Thèse de l'Université de Caen-Basse Normandie, 333p.
- [9] Dégremont L., Ernande B., Bédier E., Boudry P., 2007. Summer mortality of hatchery produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. *Aquaculture*, 262: 41-53.
- [10] De Nicola M., Gambardella C., Guarino S.M., 1992. Interactive effects of cadmium and zinc pollution on PGI and PGM polymorphisms in *Idotea baltica*. *Mar. Poll. Bull.*, 24: 619-621.
- [11] Elfstrom C.M., Gaffney P.M., Smith C.T., Seeb J.E., 2005. - Characterization of 12 single nucleotide polymorphisms in weathervane scallop. *Molecular Ecology Notes*, 5: 406-409.
- [12] Ernande B., Boudry P., Clobert J., Haure J., 2004. - Plasticity in resource allocation based life history traits in the Pacific oyster, *Crassostrea gigas*. I. Spatial variation in food abundance. *Journal of Evolutionary Biology*, 17: 342-356

- [13] Fabioux C., Huvet A., Lelong C., Robert R., Pouvreau S., Daniel J.Y., Minguant C., Le Pennec M., 2004. - Oyster vasa-like gene as a marker of the germline cell development in *Crassostrea gigas*. *Biochemical and Biophysical Research Communications*, 320: 592-598.
- [14] Guarna M.M., Borowsky R.L., 1993. - Genetically controlled food preference: biochemical mechanisms. *Proc. Natl. Acad. Sci. USA*, 90: 5257-5261.
- [15] Hughes B.L., Suniga R.G., Yardley D.G., 1994. - Influence of amylase genotypes on growth rate and feed conversion of chickens. *Poultry Sciences*, 73: 953-957.
- [16] Huvet A., Lapègue S., Magoulas A., Boudry P., 2000. - Mitochondrial and nuclear DNA phylogeography of *Crassostrea angulata*, the Portuguese oyster endangered in Europe. *Conservation Genetics*, 1: 251-262.
- [17] Huvet A., Dubois S., Daniel J.Y., Quéré C., Prudence M., Van Wormhoudt A., Sellos D., Samain J.F., Moal J., 2003. - Tissue expression of two amylase genes in the Pacific oyster *Crassostrea gigas*. Effects of two different food rations. *Aquaculture*, 228: 321-333.
- [18] Huvet A., Herpin A., Dégremont L., Labreuche Y., Samain J.F., Cunningham C., 2004. - The identification of genes from the oyster *Crassostrea gigas* that are differentially expressed in progenies exhibiting opposed susceptibility to summer mortality. *Gene*, 343: 211-220.
- [19] Langdon C., Ford E., Jacobson D., Blouin M., 2003. - Yields of cultured Pacific oysters *Crassostrea gigas* Thunberg improved after one generation of selection. *Aquaculture*, 220: 227-244.
- [20] Le Moine S., Sellos D., Moal J., Daniel J.Y., Serrano F.S.J., Samain J.F., Van Wormhoudt A., 1997. - Amylase in *Pecten maximus* (Mollusca, bivalves): protein and cDNA characterization, quantification of the expression in the digestive gland. *Mol. Mar. Biol. Biotech.*, 6: 228-237.
- [21] Le Moullac G., Bacca H., Huvet A., Moal J., Pouvreau S., Van Wormhoudt A., 2007. - Transcriptional regulation of pyruvate kinase and phosphoenolpyruvate carboxykinase in the adductor muscle of the oyster *Crassostrea gigas* during prolonged hypoxia. *Journal of Experimental Zoology*, 307A: 371-382.
- [22] Marchand J., Tanguy A., Laroche J., Quiniou L., Moraga D., 2003. - Responses of European flounder (*Platichthys flesus*) population to contamination in different estuaries from the East Atlantic Coast. *Marine Ecology Progress Series*, 260: 273-284.
- [23] Montooth K.L., Marden J.H., Clark A.G., 2003. - Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. *Genetics*, 165: 623-625.
- [24] Pouvreau S., Bourles Y., Lefebvre S., Gangnery A., Alunno-Bruscia M., 2006. - Application of a dynamic energy budget model to the Pacific oyster, *Crassostrea gigas*, reared under various environmental conditions. *Journal of Sea Research*, 56: 156-167.
- [25] *Prudence M., Moal J., Boudry P., Daniel J.Y., Quéré C., Jeffroy F., Minguant C., Ropert M., Bédier E., Van Wormhoudt A., Samain J.F., Huvet A., 2006. - An amylase gene polymorphism is associated with growth differences in the Pacific cupped oyster *Crassostrea gigas*. *Animal Genetics*, 37: 348-351.
- [26] Quilang J, Wang S, Li P., Abernathy J., Peatman E., Wang Y., Wang L., Shi Y., Wallace R., Guo X., Liu Z., 2007. - Generation and analysis of ESTs from the

- eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. *BMC Genomics*, 8: 157.
- [27] Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J., Willey D.L. et al., 2001. - A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409: 928-933.
- [28] Samain J.F., Dégremont L., Soletchnik P., Haure J., Bédier E., Ropert M., Moal J., Huvet A., Bacca H., Van Wormhoudt A., Delaporte M., Costil K., Pouvreau S., Lambert C., Boulo V., Soudant P., Nicolas J.L., Le Roux F., Renault T., Gagnaire B., Gêret F., Boutet I., Burgeot T., Boudry P., 2007. - Genetically based resistance to summer mortality in the Pacific oyster (*Crassostrea gigas*) and its relationship with physiological, immunological characteristics and infection process. *Aquaculture*, 268: 227-243.
- [29] Samain J.F., McCombie H., 2007. - Summer Mortality of Pacific Oyster *Crassostrea gigas*. Ed. Ifremer/Quae, 332 p.
- [30] *Sauvage C., Bierne N., Lapègue S., Boudry P., 2007. - Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene*, 406, 13-22.
- [31] Savina M., Pouvreau S., 2004. - A comparative ecophysiological study of two infaunal filter-feeding bivalves: *Paphia rhomboides* and *Glycymeris glycymeris*. *Aquaculture*, 239: 286-306.
- [32] Sellos D., Moal J., Dégremont L., Huvet A., Daniel J.Y., Nicoulaud S., Boudry P., Samain J.F., Van Wormhoudt A., 2003. - Structure of amylase genes in populations of Pacific cupped oyster (*Crassostrea gigas*): tissue expression and allelic polymorphism. *Mar. Biotech.*, 5: 360-372.
- [33] Small K.S., Brudno M., Hill M., Sidow A., 2007. - Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences of USA*, 104: 5698-5703.
- [34] Tanguy A., Boutet I., Moraga D., 2005. - Molecular characterization of the glutamine synthetase gene in the Pacific oyster *Crassostrea gigas*: expression study in response to xenobiotic exposure and developmental stage. *Biochimica and Biophysica Acta*, 1681: 116-125.
- [35] *Tanguy A., Boutet I., Boudry P., Dégremont L., Laroche J., Moraga D., 2006. - Molecular identification and expression of the phosphoglucomutase (PGM) gene from the Pacific oyster *Crassostrea gigas*. *Gene*, 382: 20-27.

Spécialisation de la tique des oiseaux marins et diversité des bactéries du complexe *Borrelia burgdorferi sensu lato*, agents de la maladie de Lyme: effets en cascade dans les systèmes à vecteur

Karen D. MCCOY^{(1)*}, David DUNEAU^(1,2), Thierry BOULINIER⁽³⁾

⁽¹⁾Génétique et Evolution des Maladies Infectieuses, UMR 2724 CNRS-IRD, Centre IRD, 911 Avenue Agropolis, BP 64501, 34394 Montpellier

⁽²⁾Zoologisches Institut, Evolutionsbiologie, Universität Basel, Vesalgasse 1, CH-4051 Basel, Switzerland

⁽³⁾Centre d'Ecologie Fonctionnelle et Evolutive, CNRS - UMR 5175, 1919 Route de Mende, 34293 Montpellier

Abstract: Specialisation of the seabird tick and diversity of Lyme borreliosis bacteria: cascading host effects in vector-borne systems. The role of vectors in micropathogen evolution is often neglected. However, vector specialisation for a given host type can result in the isolation of pathogen populations (*via* cascading effects) and can subsequently affect pathogen genetic variation with major consequences for the co-evolutionary process and the epidemiology of associated disease. We investigated the implications of vector specialisation by examining the diversity and population structure of bacteria of the *Borrelia burgdorferi sensu lato* complex found in the marine cycle of Lyme borreliosis, involving different seabird species and their specialist tick *Ixodes uriae* as vector. The analysis of a conserved bacterial gene (flagellin gene *FlaB*) revealed the presence of three species of the complex in the tick vector: *B. garinii*, *B. lusitaniae* and *B. burgdorferi sensu stricto*. This is the first record of *B. lusitaniae* and *B. burgdorferi s.s.* in the marine system. The bacterium was present in all colonies examined in relatively high prevalence (population average $26.0\% \pm 3.9\%$) confirming that this microparasite is endemic to the marine system. In agreement with predictions, our results show an isolation of *Borrelia* isolates by tick host race. Nevertheless, these patterns differ among seabird colonies with only three of the four examined colonies showing clear patterns of divergence. The combination of variable degrees of isolation due to specialisation of the tick vector and the effect of distance among seabird colonies seems to create a complex landscape for the evolution of these bacteria. More detailed molecular studies involving both the pathogen and the vector should help us evaluate different hypotheses that might explain these patterns. These data can also be used to infer the potential interaction between marine and terrestrial endemic cycles of borreliosis.

* Correspondance et tirés à part : mccoym@mpl.ird.fr

animal population/ biological adaptation/ genetic differentiation/ genetic epidemiology / disease vectors

Résumé : Le rôle des vecteurs dans l'évolution des micropathogènes est souvent négligé. Néanmoins, la spécialisation du vecteur pour un type d'hôte peut entraîner un isolement entre populations du microparasite (effet en cascade) et pourrait donc affecter la variation génétique de celui-ci avec des conséquences majeures sur les processus de coévolution et l'épidémiologie des maladies associées. Nous avons testé cette hypothèse en examinant la diversité et la structure des bactéries dans le cycle marin de la borréliose de Lyme, impliquant des oiseaux et leurs tiques spécialistes, *Ixodes uriae*. L'analyse d'un gène conservé codant pour une protéine de la flagelline (*FlaB*) a révélé la présence de trois espèces du complexe *Borrelia burgdorferi sensu lato* chez *I. uriae* : *B. garinii*, *B. lusitaniae* et *B. burgdorferi sensu stricto*. Une forte prévalence globale a aussi été trouvée (26 %), suggérant que ces microparasites sont endémiques au système. Conformément aux prédictions, nos résultats sont cohérents avec un isolement relatif de la borréliose en fonction des races d'hôte de tiques. Néanmoins, le degré de divergence diffère selon les colonies d'oiseaux. La combinaison de l'isolement variable dû aux vecteurs et de la distance génère donc un contexte évolutif complexe pour ce système. Des études moléculaires plus détaillées, à la fois sur l'organisme pathogène et sur le vecteur, nous aideront à tester différentes hypothèses qui pourraient expliquer ces patterns.

population animale/ adaptation biologique/ différenciation génétique/ génétique épidémiologique/ vecteurs de maladie**1. INTRODUCTION**

L'évolution de la diversité parasitaire est une cause majeure de dégâts économiques et médicaux chez les humains et les êtres vivants domestiqués dont ils dépendent. Une compréhension de la répartition de cette variabilité et des facteurs qui l'affectent est essentielle pour le contrôle de ces organismes pathogènes (par exemple, l'identification de variants pour la vaccination, etc.). Dans les systèmes à vecteur il est parfois avancé que seule l'interaction du microparasite avec son hôte vertébré détermine la direction de l'évolution et la pathogénicité du microparasite (*e.g.*, [25], [3]). Néanmoins, une telle supposition n'est jamais vérifiée et s'avère même particulièrement inadéquate dans les cas où le vecteur impose des contraintes importantes sur la dynamique du micropathogène (*e.g.*, [1]). En effet, la spécialisation du vecteur pour un type d'hôte peut entraîner un isolement entre populations du microparasite (effet en cascade) et pourrait donc affecter la variation génétique de cet organisme avec des conséquences majeures sur les possibilités d'adaptation du microparasite à ses hôtes et réciproquement [9], [28].

Afin d'étudier le rôle de la spécialisation d'un vecteur sur l'évolution d'un microparasite, nous avons examiné un système impliquant des races spécia-

lisées de la tique d'oiseaux marins, *Ixodes uriae*, et un groupe de microparasites transmis par cette tique, les bactéries du complexe *Borrelia burgdorferi sensu lato* (Bbsl), agents étiologiques de la maladie de Lyme chez l'homme. Ces bactéries sont relativement bien caractérisées dans leur cycle terrestre (impliquant surtout des mammifères et les tiques du complexe *Ixodes ricinus*), mais les informations sur le cycle marin sont quasi-inexistantes [20]. Nos travaux précédents sur *I. uriae* ont révélé une spécialisation d'*I. uriae* pour différentes espèces d'oiseaux [15], [17]. Cette étude a donc visé à caractériser la diversité des borrelies au sein du cycle marin et à examiner comment cette diversité est affectée par la présence de races d'hôtes de la tique. Si la spécialisation des tiques vectrices est une contrainte importante pour la transmission du microparasite, on s'attend à une structuration de la prévalence et de la variation génétique du microparasite entre races de vecteurs et à différentes échelles spatiales.

2. MÉTHODES

2.1. Espèces concernées

La tique *Ixodes uriae* exploite des oiseaux marins coloniaux des régions circumpolaires. Dans chaque hémisphère, les populations de ces oiseaux sont fortement subdivisées à plusieurs échelles spatiales et forment souvent des groupes hétérospécifiques dans les zones de reproduction; il y a donc la possibilité de plusieurs types d'interactions avec des parasites, au sein et entre espèces hôtes. Les principaux hôtes oiseaux infestés par *I. uriae* partagent plusieurs similarités car ils sont tous coloniaux, longévifs et pélagiques (passent l'hiver en mer) et ont tendance à être fortement fidèles à leur site de reproduction. Cependant, le comportement varie entre ces espèces. Par exemple, elles se dispersent à différentes échelles spatiales, utilisent différents types de sites de nidification au sein des colonies, et ont des stratégies différentes de fourragement en pleine mer. Ces différences peuvent avoir un impact sur l'évolution des populations de tiques associées à ces hôtes. En effet, avec une approche de génétique des populations, nous avons mis en évidence une différenciation de cet ectoparasite entre espèces hôtes sympatriques de l'hémisphère Nord [15]. Nous avons également caractérisé l'effet de cette spécialisation présumée sur la structuration spatiale des populations de la tique [16]. Nous avons utilisé le «répliquat» naturel de ce système dans l'hémisphère Sud pour confirmer que la formation des races d'hôtes était bien un processus évolutif récurrent chez cette tique [17].

Il a été découvert en 1993 que les oiseaux marins hébergent les bactéries du complexe *Borrelia burgdorferi sensu lato* (Bbsl) et que la tique *I. uriae* en est le

vecteur [20]. Il existe à présent 13 espèces reconnues de ce complexe, dont au moins quatre sont pathogènes chez l'homme et responsables de la maladie de Lyme (*B. burgdorferi sensu stricto*, *B. garinii*, *B. spielmanii* et *B. afzelii*), qui est la maladie à vecteur la plus importante en zone tempérée [24]. Les effets pathogènes de ces bactéries diffèrent suivant l'espèce de *Borrelia*, mais aussi suivant l'individu hôte infecté; il y a donc une forte interaction entre le génotype du microparasite et celui de l'hôte [26]. Les conséquences de l'infection pour les oiseaux ne sont pas connues. À présent, *B. garinii* est la seule espèce décrite dans le cycle marin, mais le nombre d'études est limité pour l'instant. De la même manière, on n'a pas encore assez d'informations pour savoir s'il existe une liaison avec le cycle terrestre de la borréliose. Par ailleurs, les implications des vecteurs spécialistes pour la transmission n'ont jamais été évaluées.

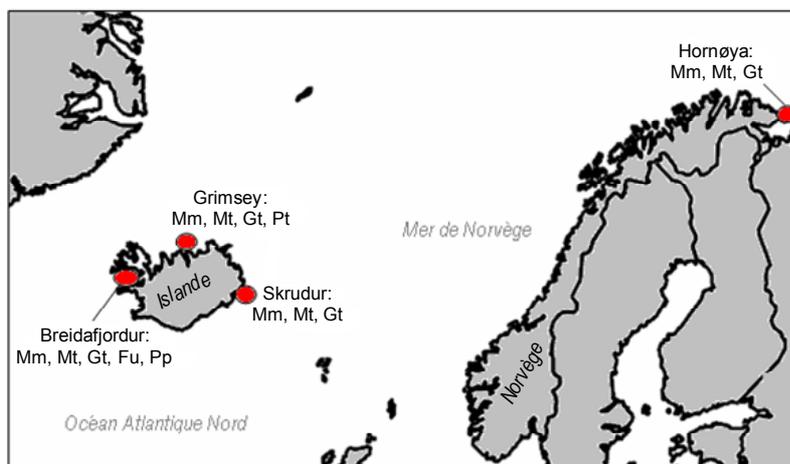


Figure 1. Localisation des colonies hétérospecifics d'oiseaux marins où la tique *Ixodes uriae* a été échantillonnée. Les espèces d'oiseaux sur lesquelles les tiques ont été récoltées sont indiquées (Mm: macareux moine, Mt: mouette tridactyle, Gt: guillemot de Troil, Fu: fulmar, Pt: pingouin torda)

Dans le cadre de notre étude, nous avons sélectionné quatre sites où nous avons pu prélever des tiques sur au moins trois espèces hôtes oiseaux nichant en sympatrie (fig. 1). L'échantillonnage a été fait sur une colonie dans le nord de la Norvège (île d'Hornøya) et sur trois colonies d'Islande (l'île de Skrudur, l'île de Grimsey et à Breidafjörður). Les oiseaux échantillonnés appartiennent à cinq espèces : la mouette tridactyle (*Rissa tridactyla*), le macareux moine (*Fratercula arctica*), le pingouin Torda (*Alca torda*), le guillemot de Troil (*Uria aalge*) et le fulmar (*Fulmarus glacialis*). Nous avons essayé de prélever environ 30 tiques pour chaque espèce d'oiseau par

site lorsque l'espèce considérée était infestée. Dans la mesure du possible, les tiques ont été collectées sur des individus différents ; il s'agit majoritairement d'adultes femelles, plus rarement de nymphes.

2.2. Analyses génétiques

L'extraction de l'ADN total de la tique et des micropathogènes associés a été effectuée à l'aide du kit DNeasy Tissue (Qiagen, Valencia, CA) à partir d'un broyat de tique contenant le tube digestif et les glandes salivaires. Pour déterminer si une tique est infectée par une borrelie ou non, nous avons employé un protocole de PCR nichée pour amplifier le gène de la flagelline (*FlaB*), qui est un gène de ménage fortement conservé au sein du complexe Bbsl. [14], [5]. Ce gène code pour une protéine du flagelle de la bactérie. Les régions marginales de ce gène sont très conservées par rapport aux régions plus médianes qui sont plus polymorphes. Cela en fait un marqueur moléculaire idéal pour la détection de la présence/absence de la bactérie et pour l'évaluation de la diversité génétique. La PCR nichée (avec deux étapes d'amplification et deux séries d'amorces, les deuxièmes étant nichées dans la séquence de la première amplification) nous a permis d'amplifier un fragment de 390 pb de la région polymorphe du gène considéré [5].

Le mélange PCR utilisé était: 2,5µl de tampon (10x), 2µl de MgCl₂ (25mM), 2µl de dNTP (2,5mM), 0,5µl d'amorces (20µM), 1 unité de Taq polymérase (Promega), 20 ng d'ADN, eau qsp 25µl. Le programme PCR était : dénaturation initiale à 95 °C 1 minute, puis 35 cycles de 94 °C 30 secondes, 52 °C 40 secondes, et 72 °C 1 minute, et une étape d'élongation finale de 72 °C 5 minutes. Le mélange PCR a été le même pour les deux PCRs successives, sauf pour la deuxième dans laquelle 0,5µl du produit de la première PCR a été utilisé comme matrice d'amplification et où des conditions PCR plus sévères ont été utilisées (température d'annealing à 55 °C). Pour chaque PCR, nous avons utilisé un contrôle négatif (eau) et un contrôle positif, une borrelie provenant de culture et fournie par le Centre de Référence des Spirochètes de l'Institut Pasteur de Paris (la souche POTib2 de *B. lusitaniae* et la souche 20047 de *B. garinii*). L'infection a été révélée par la présence d'une bande de 300 à 400 pb sur un gel d'agarose (2 %). Ces produits PCR ont été ensuite séquencés (Genome Express, Meylan).

Nous avons ensuite effectué une analyse phylogénétique des souches de borrelie de notre échantillonnage avec des espèces reconnues de LB. Les séquences de référence, correspondantes aux différentes espèces du complexe *Borrelia burgdorferi* s.l. reconnues, ont été obtenues à partir de Genbank. Les séquences ont été alignées à l'aide du logiciel SEAVIEW [10]. Un arbre phylogénétique du gène *FlaB* a été ensuite construit par la méthode de

maximum de vraisemblance (PAUP 4.0b10: [29]) après sélection d'un modèle de substitution des nucléotides par le programme MODELTEST 3.6 [22]. La robustesse des nœuds a été testée par une analyse de bootstrap (1000 répétitions) [29] et l'arbre final a été dessiné par le programme TREEDYN [4].

2.3. Analyses statistiques

Afin de tester la prédiction que la spécialisation des tiques pour différentes espèces d'oiseaux conduirait à des prévalences différentes, des tests exacts de Fisher ont été effectués au sein de chaque colonie mixte, en utilisant le programme STRUC [23]. Ce test nous a permis de voir si certaines races de tiques sont plus infectées que d'autres au sein de chaque site. La procédure de Fisher a été employée pour combiner des tests indépendants effectués dans chaque colonie [23]. De même, nos colonies échantillonnées étant séparées par des distances géographiques plus ou moins grandes, nous avons voulu tester si la prévalence de la borrélie variait selon les colonies étudiées, en fonction de l'éloignement. Nous avons donc fait également des tests exacts de Fisher pour savoir si, pour chaque race de tique, la prévalence était différente d'un site à l'autre. Certaines espèces n'ayant pas pu être échantillonnées dans tous les sites (fig. 1), cette dernière analyse a été faite uniquement pour les tiques de la mouette tridactyle, du macareux moine et du guillemot de Troil. Nous avons aussi effectué un test global non paramétrique (test de Kruskal-Wallis), sur la variabilité de la prévalence entre colonies incluant les données des trois races hôtes.

La comparaison des prévalences n'est pas le test le plus puissant pour tester une divergence potentielle des borréliés entre populations hôtes, car le taux d'infection des tiques peut se révéler similaire même si la composition génétique des borréliés est variable entre populations de tiques. Afin de déterminer si la divergence entre séquences de borréliés est liée à un isolement des populations, des analyses de variance moléculaire (AMOVA) ont été effectuées à l'aide du logiciel ARLEQUIN 3.01 [27]. L'analyse a été faite dans un premier temps à partir de populations établies sur la base des races d'hôtes de tique pour chaque site. Ceci nous permet d'estimer la part de variance moléculaire expliquée par l'isolement lié à la spécialisation des tiques. Dans un deuxième temps, l'analyse a été faite pour chaque race échantillonnée dans toutes les colonies pour tester le rôle potentiel de l'espace dans la divergence des borréliés. Ces analyses ont été faites uniquement sur des séquences de *B. garinii*, car il n'y avait pas un nombre de séquences suffisant pour les autres espèces pour effectuer ce genre d'analyse. Nous ne pouvons pas inclure toutes les espèces dans une même AMOVA car l'inclusion d'espèces différentes au sein d'une même population peut masquer les divergences entre population.

3. RÉSULTATS

Parmi les 402 tiques testées pour la borrélie, 109 ont été positives en *Borrelia* par PCR (soit une prévalence globale moyenne de 26,0% ± 3,9). Une étude de répétabilité de détection de la borrélie par cette méthode a suggéré que cette valeur sous-estime la vraie prévalence d'infection chez ces tiques [18], notre estimation de prévalence est donc conservatrice. Parmi les positifs, nous avons obtenu les séquences de 97 isolats, avec 31 séquences différentes et onze cas de co-infections par des borrélioses (tabl.I). Seules les séquences contenant une seule souche ont pu être utilisées pour les analyses de structuration (soit 86 séquences).

Tableau I : Prévalence et répartition des isolats de *Borrelia burgdorferi* s.l (Bl = *B. lusitaniae*; Bg = *B. garinii*). Le nombre de tiques infectées avec plusieurs souches de *Borrelia* spp. est indiqué dans la colonne 'Co-inf'. Voir la figure 1 pour la localisation des colonies.

Colonie	Hôte (No. Échantillonnés)	No. Tiques (A=adultes, N=nymphe)	Prévalence de <i>Borrelia</i>		Espèces**		
			No. tiques positives (No. hôtes)*	Moyen	Bl	Bg	Co-inf
Hornøya (Norvège)	Mouette (21)	29(26A, 3N)	2 (2)			1	
	Macareux (27)	30(9A, 25N)	7(7)			5	
	Guillemot (29)	30(28A, 2N)	7(7)			4	
			Total: 16	18,0 %			
Skrudur (Islande)	Mouette (17)	28(25A, 3N)	5(5)		2	2	1
	Macareux (19)	31(26A, 5N)	7(4)			7	
	Guillemot (27)	32(32A)	6(6)		2	4	
			Total: 18	19,8 %			
Grimsey (Islande)	Mouette (24)	30(16A, 14N)	10(10)		2	5	3
	Macareux (28)	39(29A, 10N)	13(11)		2	11	
	Guillemot (23)	30(25A, 5N)	10(10)		3	5	2
	Pingouin (12)	17(1A, 16N)	0				
			Total: 33	28,4 %			
Breidafjörður (Islande)	Mouette (11)	20(2A, 18N)	9(8)		2	6	1
	Macareux (25)	30(14A, 16N)	19(16)			13	2
	Guillemot (7)	12(12A)	3(3)			1	1
	Pingouin (9)	14(12A, 2N)	2(2)			2	
	Fulmar (16)	30(11A, 19N)	9(9)			6	1
			Total: 42	39,6 %			

* No. hôtes = nombre d'individus hôtes échantillonnés avec au moins une tique infectée.

** des séquences complètes n'ont pas été obtenues pour 12 tiques positives. Un isolat de *B. burgdorferi sensu stricto* a été trouvé chez une tique de guillemot sur Hornøya

Après alignement, les séquences de 311 pb ont été utilisées pour l'analyse phylogénétique. Dans cette région, il y avait 63 sites variables, dont 21 sites informatifs. Pour éviter l'attribution de trop de poids pour un événement de mutation, les séquences trouvées plusieurs fois n'ont été incluses qu'une seule fois dans l'arbre (fig. 2). La comparaison des séquences obtenues aux séquences de références des borrélioses connues a suggéré la présence d'au moins trois espèces de *Borrelia burgdorferi* s.l. au sein du système marin : *B.*

garinii (84 % des positifs), *B. lusitaniae* (15 % des positifs), et *B. burgdorferi sensu stricto* (1 % des positifs) [7].

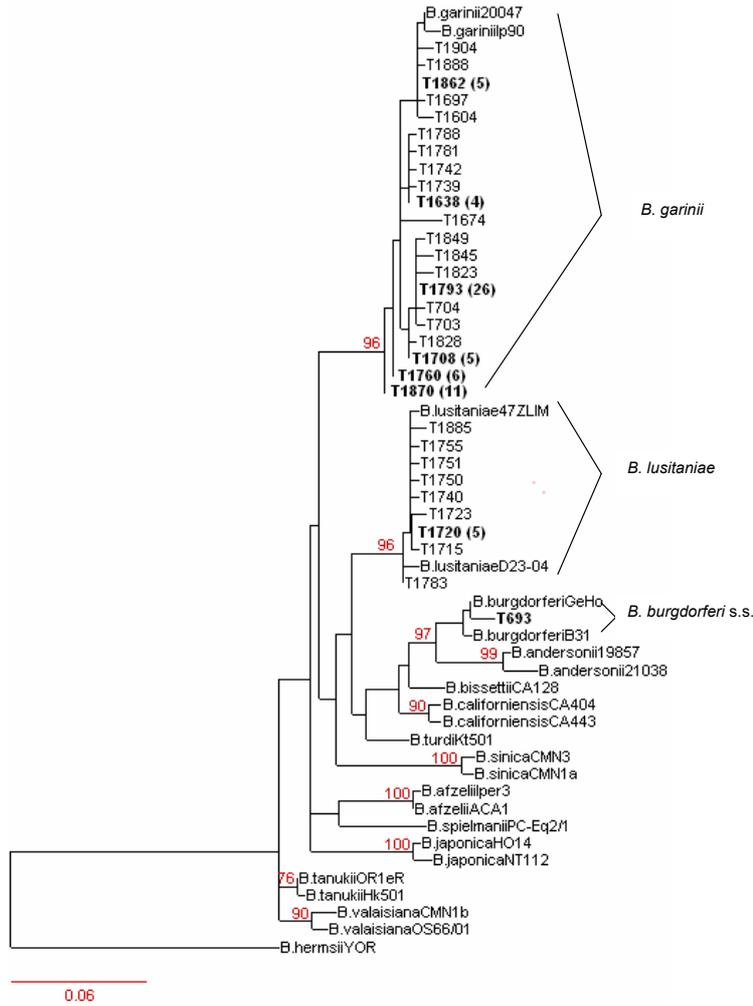


Figure 2 : Arbre phylogénétique du gène de la flagelline des borrelies obtenu par la méthode de maximum de vraisemblance. Les séquences de référence ont été obtenues à partir de Genbank et correspondent aux différentes espèces du complexe *Borrelia burgdorferi s.l.* actuellement reconnues (voir Duneau *et al.* sous presse). Plusieurs séquences identiques ont été trouvées dans les populations testées, leur nombre est indiqué entre parenthèses dans l'arbre. La robustesse des noeuds est testée par une analyse de bootstrap (1 000 répétitions) et seules les valeurs supérieures à 75% sont représentées. Figure d'après Duneau *et al.* (sous presse).

3.1. Prévalence des borrélioses

Pour détecter une différence de prévalence entre les races de tiques sympatriques, nous avons effectué des tests individuels pour chaque colonie. Sur Hornøya, malgré une tendance pour les tiques de mouettes à être moins infectées, il n'y avait pas de différence significative de prévalence en borréliose entre races de tiques (Test exact de Fisher, $p = 0,18$, tabl. I). Il n'y avait également pas de différence significative entre les trois systèmes d'hôtes sur Skrudur (Test exact de Fisher, $p=0,895$). Pour Breidafjordur, une différence de prévalence significative entre des tiques infectées chez les guillemots, mouettes, macareux, fulmars et pingouins a été trouvée (Test exact de Fisher, $p = 0,011$). Une différence de prévalence entre les tiques des quatre races d'hôtes a été aussi trouvée sur Grimsey (Test exact de Fisher, $p = 0,019$). Quand les résultats de ces différents tests sont combinés, la prévalence globale varie significativement entre races au sein des colonies (Procédure de Fisher, $p = 0,0081$). Ces résultats pourraient être largement expliqués par un faible taux d'infection des tiques de pingouins par rapport aux autres races (tabl. I). En effet, si on retire les échantillons de cette race de tique, il n'y a plus de différences de prévalence au sein des colonies. Il semble donc que la prévalence de la borréliose soit relativement homogène entre races de tique pour les quatre sites étudiés.

Pour tester l'influence de la distance géographique, nous avons comparé la prévalence entre sites pour les tiques de guillemots, mouettes et macareux. Globalement, la prévalence varie significativement entre colonies (test de Kruskal-Wallis, $p = 0,001$). Néanmoins, les résultats sont variables pour chaque race (tabl. I). Pour les tiques de guillemots, il n'y avait pas de différence de prévalence de la bactérie entre les sites (Test de Fisher, $p = 0,63$). En revanche, une différence est observée pour les hôtes mouettes et macareux (Test de Fisher, respectivement, $p = 0,0084$; $p = 0,0025$).

3.2. Structuration des borrélioses

Comme le résume le tableau II, on observe un fort effet de l'isolement des souches de *B. garinii* par race de tiques dans toutes les colonies sauf Skrudur. Les espèces d'oiseaux sont donc associées avec une part significative de la variance moléculaire des souches de borréliose (Procédure de Fisher, $p = 0,0003$).

Tableau II : Description de la structuration de *B. garinii* entre races d'*Ixodes uriae* au sein de chaque site hétérosécificique obtenue à partir d'une analyse de variance moléculaire (AMOVA). La différenciation entre populations est estimée par Φ_{ST} . Les chiffres en gras sont significatifs.

Colonie	N	Est. Φ_{ST}	Valeur de P
Skrudur	13	0,01	0,440
Hornøya	10	0,32	0,037
Breidafjordur	28	0,14	0,030
Grimsey	21	0,27	<0,001

Dans un deuxième temps, cette même analyse a été faite à partir des sites pour trois races de tiques. Au vu des résultats des AMOVA (tabl. III), il semblerait qu'il y ait un effet du site uniquement pour les borrelies provenant des tiques collectées sur deux des trois espèces considérées, les tiques de macareux et de guillemots. Pour les borrelies des tiques de mouettes, le site n'a pas d'effet sur leur divergence, avec une répartition homogène des souches présentes entre sites. Néanmoins, une augmentation du nombre de sites d'échantillonnage et du nombre de séquences par site serait utile pour confirmer la généralité de tels patterns.

Tableau III : Description de la structuration de *B. garinii* entre sites pour trois races d'*Ixodes uriae* obtenue à partir d'une analyse de variance moléculaire (AMOVA). La différenciation entre populations est estimée par Φ_{ST} . Les chiffres en gras sont significatifs.

Hôte	N	Est. Φ_{ST}	Valeur de P
Guillemot de Troil	14	0,28	0,014
Macareux moine	38	0,17	0,002
Mouette tridactyle	14	0	0,52

4. DISCUSSION

Nous avons évalué la diversité et la structure des borrelies isolées dans un cycle tout à fait original de la borreliose de Lyme: le cycle marin. Ce cycle fait intervenir différentes espèces d'oiseaux de mer coloniaux et la tique *Ixodes uriae*, un ectoparasite qui se spécialise sur ces différentes espèces d'oiseaux. Notre étude visait à évaluer l'influence de la spécialisation des tiques vectrices dans la diversité des borrelies. Si la spécialisation des tiques vectrices est une contrainte importante pour la transmission du

microparasite, on s'attendait à une structuration de la prévalence et de la variation génétique du microparasite entre races de tiques et à différentes échelles spatiales en fonction de ces races. Nos résultats sont cohérents avec ces prédictions.

Globalement, la prévalence était forte dans le système marin avec 26 % des tiques infectées. Néanmoins, l'infection variait significativement en fonction des races d'hôtes de tiques et des sites géographiques. Par exemple, la prévalence était variable entre races de tiques dans seulement deux des quatre colonies et seules les tiques récoltées sur les pingouins torda montraient un taux d'infection très faible. La faible prévalence chez les tiques de pingouin pourrait être due à une faible compétence de ces vecteurs et/ou de ces oiseaux pour la borrelie, cependant cet oiseau a été déjà décrit comme hôte compétent pour la borrelie [20]. Plus généralement, ces différences pourraient être expliquées par la nature dynamique de la divergence des tiques en fonction de l'hôte oiseau. En effet, des études sur ce système suggèrent que la formation des races de tiques est en cours et pourrait avoir lieu en parallèle dans différentes zones géographiques [17], (Kempf et McCoy, données non publiées). Dans certaines colonies, les tiques sont potentiellement moins spécifiques de l'hôte que dans d'autres colonies, ce qui pourrait résulter en une prévalence plus homogène dans ces colonies. Des études, actuellement en cours, sur la structuration détaillée des tiques au sein des colonies considérées ici nous fourniront les données pour tester cette hypothèse plus explicitement. Il faut souligner que les échantillons considérés dans cette étude ont été largement collectés dans des zones limitées au sein des colonies et que la prévalence de la borrelie pourrait varier à cette échelle spatiale [11]. Nous avons également récolté des tiques sur des années différentes en Islande (2003) et en Norvège (1998). Si l'infection par les borrelies varie significativement dans l'espace et dans le temps, il serait important d'étudier la variabilité de ces prévalences au sein des colonies et entre années de récolte pour mieux comprendre le fonctionnement de chaque système hôte-tique. Un tel travail devra aussi considérer plus spécifiquement l'effet du stade de développement de la tique (nymphe ou adulte femelle).

Si l'isolement par races de tiques ou par zones géographiques peut avoir un effet sur la prévalence des borrelies, il peut également avoir un effet sur la divergence moléculaire de celles-ci. Nos résultats soutiennent cette prédiction puisque les estimations de divergence (Φ_{ST}) entre borrelies des tiques collectées sur différents hôtes pour trois des quatre sites sont fortes ($\Phi_{ST} > 0,14$). De plus, un effet du site a été trouvé pour les tiques de macareux et de guillemot. Ceci n'était pas le cas pour les tiques de mouette. Ce résultat suggère donc que l'isolement de la borrelie diffère selon les différents systèmes de races de tiques. Ici, nous avons utilisé un gène

conservé pour examiner la diversité et la structure des borrelies. Il est donc possible que la structure et la diversité selon la race de vecteur soient plus fortes à d'autres marqueurs. Une étude par MLSA (multi-locus sequence analysis; [24]) sera nécessaire pour mieux évaluer la nature de la diversité présente dans ce système.

Nos résultats sont cohérents avec l'isolement en fonction de races de tiques mais ceci pourrait aussi être expliqué par l'adaptation des borrelies pour les oiseaux eux-mêmes. Cette hypothèse semble peu probable car aucune association entre isolat de borrelie et espèce oiseau n'a été suggérée par des analyses du gène de la flagelline (voir [7]). Néanmoins, l'analyse de la variabilité au niveau de gènes soumis à sélection tel que l'*ospC* (outer surface protein C), impliqué dans l'infection de l'oiseau par la borrelie [2], ou l'*ospA* (outer surface protein A), impliqué dans la capacité pour la borrelie à résider dans la tique [21], pourrait nous renseigner sur l'importance relative des hôtes oiseaux et des vecteurs tiques dans la divergence des borrelies.

D'un point de vue épidémiologique et écologique, l'estimation de la prévalence d'un pathogène est cruciale afin de connaître la pression potentielle que le parasite peut exercer sur l'écosystème. Dans la littérature, la prévalence de la borrelie chez les tiques du cycle terrestre peut être très forte, par exemple, une prévalence de 33 % chez les nymphes de *I. ricinus* a été trouvée en Alsace, France [8]. Nos données indiquent que la prévalence peut également être forte dans le système marin. Ceci confirme que la présence de la borrelie dans le cycle marin est endémique et que, en fonction de la liaison entre cycle marin et terrestre, ce cycle pourrait servir de réservoir pour les agents de la borreliose [19]. Clairement, l'existence d'associations indépendantes pour chaque système hôte oiseau-tique pourrait avoir un effet important sur la distribution des isolats car chaque oiseau a ses propres tendances de dispersion [16].

Dans le cadre de cette étude nous avons détecté la présence de trois espèces de *B. burgdorferi s.l.* dans le cycle marin: *B. garinii* déjà connue dans ce système [20], *B. lusitaniae*, une espèce typiquement associée aux pays chauds [30] et soupçonnée d'être aussi pathogène [6], et *B. burgdorferi sensu stricto*, une espèce considérée comme généraliste [13]. Cette étude représente les premières détections des espèces *B. lusitaniae* et *B. burgdorferi s.s.* au nord de l'Europe et chez les oiseaux de mer. Néanmoins, des analyses génétiques plus détaillées seront nécessaires pour savoir si ces isolats correspondent réellement à *B. lusitaniae* et *B. burgdorferi s.s.*, ou bien à des espèces apparentées (voir ci-dessus). Ces analyses pourraient aussi nous aider à identifier la liaison potentielle entre les cycles marin et terrestre de la borrelie. Est-ce que les oiseaux de mer sont vraiment responsables de la dissémination globale de cette bactérie ou est-ce que le transfert des bactéries entre cycles est rare ?

Peu d'études à ce jour prennent en compte l'importance de l'hôte vecteur dans l'évolution des micropathogènes. Ici, nous avons pu mettre en évidence que l'évolution des races de tiques peut avoir des conséquences en cascade sur la distribution et la diversité génétique des souches de bactéries. Plus généralement, on sait maintenant que la structure et la compétence des vecteurs peuvent varier selon l'espèce et les populations considérées dans de nombreux systèmes à vecteur (*e.g.* plasmodium – moustiques; mouches tsé-tsé – trypanosomes ; [12]). Le rôle du vecteur dans l'écologie et l'évolution des pathogènes devraient donc à l'avenir être pris en compte plus largement afin de mieux comprendre ces systèmes dont l'épidémiologie est cruciale en terme de santé publique.

REMERCIEMENTS

Nous voudrions remercier D. Postic et M. Garnier (Centre de Référence des Spirochètes, Institut Pasteur de Paris), A. Petersen, T. Tverra, R.T. Barrett, E. Gomez-Diaz, C. Barnabé, V. Staszewski et les membres du groupe «Tiques et Maladies à Tiques» du REID pour leurs avis et assistances. En plus d'un soutien du BRG, nous reconnaissons des financements de recherche attribués par l'Agence Nationale de Recherche et de l'Institut Polaire Français – IPEV (programme 333).

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Azad A.F., Beard C.B., 1998. - Rickettsial pathogens and their arthropod vectors. *Emerg. Infect. Dis.* 4:179-186.
- [2] Brisson D., Dykhuizen D.E., 2004. - *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* 168:713-722.
- [3] Bunikis J., Tsao J., Luke C.J., Luna M.G., Fish D., Barbour A.G., 2004. - *Borrelia burgdorferi* infection in a natural population of *Peromyscus leucopus* mice: a longitudinal study in an area where Lyme borreliosis is highly endemic. *J. Infect. Dis.*, 189:1515-1523.
- [4] Chevenet F., Brun C., Banuls A.-L., Jacq B., Christen R., 2006. - TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, 7: 439.
- [5] Clark K., 2004. - *Borrelia* species in host-seeking ticks and small mammals in northern Florida. *J. Clin. Microbiol.*, 42:5076-5086.
- [6] Collares-Pereira M., Couceiro S., Franca I., Kurtenbach K., Schafer S.M., Vitorino L., Goncalves L., Baptista S., Vieira M.L., Cunha C., 2004. - First isolation of *Borrelia lusitaniae* from a human patient. *J. Clin. Microbiol.*, 42:1316-1318.
- [7] *Duneau D., Boulinier T., Gomez-Diaz E., Petersen A., Tveraa T., Barrett R.T., McCoy K.D., 2008 - Prevalence and diversity of Lyme borreliosis bacteria in marine birds. *Infect. Genet. Evol.*, 8:352-359.

- [8] Ferquel E., Garnier M., Marie J., Bernede-Bauduin C., Baranton G., Perez-Eid C., Postic D., 2006. - Prevalence of *Borrelia burgdorferi sensu lato* and Anaplasmataceae members in *Ixodes ricinus* ticks in Alsace, a focus of Lyme borreliosis endemicity in France. *Appl. Environ. Microbiol.*, 72:3074-3078.
- [9] Gandon S., Michalakis Y., 2002. Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time. *J. Evol. Biol.*, 15:451-462.
- [10] Galtier, N., Gouy, M., Gautier, C., 1996. - SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.*, 12: 543-548.
- [11] Gasparini J., McCoy K.D., Haussy C., Tveraa T., Boulinier T., 2001. - Induced maternal response to the Lyme disease spirochaete *Borrelia burgdorferi sensu lato* in a colonial seabird, the kittiwake *Rissa tridactyla*. *Proc. R. Soc. Lond. B*, 268:647-650.
- [12] Gooding R.H., 1996. - Genetic variation in arthropod vectors of disease-causing organisms: obstacles and opportunities. *Clin. Microbiol Rev.* 9:301.
- [13] Hanincova, K., Kurtenbach, K., Diuk-Wasser, M., Brei, B., Fish, D., 2006. - Epidemic spread of Lyme borreliosis, northeastern United States. *Emerg. Infect. Dis.*, 12: 604-611.
- [14] Johnson B., Happ C., Mayer L., Piesman J., 1992. - Detection of *Borrelia burgdorferi* in ticks by species-specific amplification of the flagellin gene. *Am. J. Trop. Med. Hyg.*, 47:730-741.
- [15] McCoy K.D., Boulinier T., Tirard C., Michalakis Y., 2001. - Host specificity of a generalist parasite: genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. *J. Evol. Biol.*, 14:395-405.
- [16] McCoy K.D., Boulinier T., Tirard C., Michalakis Y., 2003. - Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. *Evolution*, 57:288-296.
- [17] McCoy K.D., Chapis E., Tirard C., Boulinier T., Michalakis Y., Le Bohec C., Le Maho Y., Gauthier-Clerc M., 2005. - Recurrent evolution of host-specialized races in a globally distributed parasite. *Proc. R. Soc. Lond. B.* 272:2389-2395.
- [18] *McCoy K.D., Duneau D., Doherty P., Soumis. - Infected or not? Dealing with detection when estimating prevalence.
- [19] Olsen B., Duffy D.C., Jaenson T.G.T., Gylfe A., Bonnedahl J., Bergström S., 1995. - Transhemispheric exchange of Lyme disease spirochetes by seabirds. *J. Clin. Microbiol.*, 33:3270-3274.
- [20] Olsen B., Jaenson T.G.T., Noppa L., Bunikis J., Bergström S., 1993. - A Lyme borreliosis cycle in seabirds and *Ixodes uriae* ticks. *Nature*, 362:340-342.
- [21] Pal U., Li X., Wang T., Montgomery R.R., Ramamoorthi N., deSilva A.M., Bao F.K., Yang X.F., Pypaert M., Pradhan D., Kantor F.S., Telford S., Anderson J.F., Fikrig E., 2004. - TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. *Cell*, 119:457-468.
- [22] Posada, D., Crandall, K. A., 1998. - MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14: 817-818.
- [23] Raymond M., Rousset F., 1995. - GENEPOP(version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.*, 86:248-249.

- [24] Richter D., Postic D., Sertour N., Livey I., Matuschka F.R., Baranton G., 2006. - Delineation of *Borrelia burgdorferi sensu lato* species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp nov. Int. J. System. Evol. Microbiol., 56:873-881.
- [25] Ricklefs R.E., Fallon S.M., 2002. - Diversification and host switching in avian malaria parasites. Proc. R. Soc. Lond. B, 269:885-892.
- [26] Rosa P.A., Tilly K., Stewart P.E., 2005. - The burgeoning molecular genetics of the Lyme disease spirochaete. Nature Rev. Microbiol., 3:129-143.
- [27] Schneider S., Roessli D., Excoffier L., 2000. - Arlequin ver. 2.0: A software for population genetics analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- [28] Stireman J.O., Nason J.D., Heard S.B., Seehawer J.M., 2006. - Cascading host-associated genetic differentiation in parasitoids of phytophagous insects. Proc. R. Soc. Lond. B, 273:523-530.
- [29] Swofford, D.L. 2002., - PAUP Version 4.10 Phylogenetic Analysis Using Parsimony (and other methods). Illinois Natural History Survey, Champaign, IL.
- [30] Younsi H., Sarih M., Jouda F., Godfroid E., Gern L., Bouattour A., Baranton G., Postic D., 2005. Characterization of *Borrelia lusitaniae* isolates collected in Tunisia and Morocco. J. Clin. Microbiol., 43:1587-1593.

Tragédie des anti-communaux et gestion collective dans les biotechnologies

Yann MÉNIÈRE⁽¹⁾, Michel TROMMETTER^{(2)*}, Henry FEYT⁽³⁾,
Catherine POTVIN⁽³⁾

⁽¹⁾CERNA, Ecole des mines de Paris, 60 bd Saint Michel, 75272 Paris Cedex 6, France

⁽²⁾UMR GAEL INRA /UPMF, BP 47, 38040 Grenoble Cedex 9, France

⁽³⁾CIRAD, Avenue Agropolis - TA 173/04 34398 Montpellier Cedex 5, France

Abstract : The tragedy of anticommons and collective mangement in plant biotechnology. When Heller and Eisenberg (1998) studied the implementation of intellectual property rights on the diverse inputs which make it possible to carry out cumulative innovations in biotechnologies, they highlighted the existence of multitude rights leading potentially to an underdevelopment of these inputs and, with a deceleration of the innovation. To prevent this tragedy known as “the tragedy of anti-commons”, Heller and Eisenberg (1998) plead for a better collective management of the upstream research results (to recreate regulated common goods). This tragedy exists because of coexistence of too many rights and, too many accesses and uses conditions which are, themselves, diverse. Within this framework, we confirm that the controversy on the blocking of research by an excess of intellectual property does exist. We show that this controversy finds its origin in the fact that the concept of blocking of research is undoubtedly excessive compared to other less radical consequences but also penalizing which are the delays and/or the modifications in the research programs. We support our result, by interviews, we could identify that the characteristics which lead to these difficulties of access and diffusions are more diverse than the intellectual property right as such: statute of the material, owner of the material, sanitary responsibility, recourse to the secrecy as well on genetic resources as on information, ...

intellectual property right/ patent/ biotechnology/ plant

Résumé : Lorsque Heller et Eisenberg (1998) ont étudié la mise en œuvre de droits de propriété privée sur les différents moyens qui permettent de réaliser des innovations cumulatives dans les biotechnologies, ils ont mis en évidence l'existence d'une multitude de droits conduisant potentiellement à une sous-exploitation de ces instruments et à un ralentissement de l'innovation. Pour prévenir cette tragédie dite des « anti-communaux », les auteurs ont plaidé pour une meilleure gestion collective des résultats amont de la recherche (recréer des biens communs régulés). Cette tragédie existe du fait de la coexistence de droits trop nombreux et de conditions d'accès et d'usages qui sont, elles-mêmes, variables. Dans ce cadre, nous confirmons

* Correspondance et tires à part : trommetter@grenoble.inra.fr

que la controverse sur le blocage des recherches par un excès de propriété intellectuelle existe. Nous montrons que cette controverse trouve son origine dans le fait que la notion de blocage de la recherche est sans doute excessive par rapport à d'autres conséquences moins radicales mais tout aussi pénalisantes que sont les retards et/ou les modifications dans les programmes de recherche. Nous appuyons notre résultat sur le fait que, par enquête, nous avons pu identifier que les caractéristiques qui conduisent à ces difficultés d'accès et de diffusion sont de nature plus diverses que les droits de propriété intellectuelle en tant que tel : statut du matériel, propriétaire du matériel, responsabilité sanitaire, recours au secret tant sur des ressources génétiques que sur les informations les concernant, ...

propriété intellectuelle/ brevet/ biotechnologie/ plante

1. INTRODUCTION

Dans «La tragédie des communaux» publiée en 1968, le biologiste Garret Hardin part de l'exemple d'un village d'éleveurs, où chacun peut faire paître ses animaux dans un pré n'appartenant à personne en particulier. L'usage du pré étant gratuit et sans contrainte, et l'éleveur tirant son revenu de son bétail, l'intérêt de chaque éleveur est de conduire ses animaux au pré le plus souvent possible, le plus tôt possible et le plus longtemps possible. Inévitablement, le pré se transforme en champ de boue. Tout le monde a perdu. L'auteur propose alors deux solutions. Soit le pré reste le bien commun du village, mais une autorité disposant d'un pouvoir de sanction est chargée de gérer la ressource (le pré) et d'en répartir l'utilisation entre chaque éleveur. Soit chaque éleveur bénéficie d'un droit de propriété sur une parcelle du pré, et donc se chargera d'en gérer la ressource. C'est la seconde option, celles des «enclosures», des droits de propriété privée, qui a inspiré la plupart des textes internationaux concernant la gestion des ressources communes de l'humanité, telles que la biodiversité. Lorsque Heller et Eisenberg [13] ont étudié la mise en œuvre de droits de propriété privée sur les différents moyens qui permettent de réaliser des innovations cumulatives dans les biotechnologies, ils ont mis en évidence l'existence d'une multitude de droits conduisant potentiellement à une sous-exploitation de ces instruments et, partant, à un ralentissement de l'innovation. Pour prévenir cette tragédie dite des «anti-communaux», Heller et Eisenberg [13] plaident pour une meilleure gestion collective des résultats amont de la recherche (recréer des biens communs régulés).

Cette tragédie n'est pas uniquement liée à l'existence de multiples brevets enchevêtrés (buissons de brevets)[†]. Elle existe du fait de la coexistence de

[†] Heller & Eisenberg (1998) identifient les facteurs suivants: *The existence of numerous right holders with claims on the inputs into the discovery process or on elements of a given product;*

droits trop nombreux et de conditions d'accès et d'usages qui sont elles-mêmes variables : conditions d'accès aux licences de brevets sur les innovations biotechnologiques et aux licences sur les certificats d'obtentions végétales pour la protection des variétés végétales ; conditions d'accès aux collections de ressources génétiques et de mutants ; conditions d'accès aux bases de données et aux logiciels, ... Dans la pratique, les travaux sur la mise en évidence de la tragédie des anti-communaux ont principalement porté sur les « buissons de brevets » et les controverses qui ont suivi (section 1). Nous présenterons dans la section 2 les contraintes qui ont pu être mises en évidence dans les conditions d'accès aux collections de matériels, aux bases de données et aux logiciels. Dans une troisième section nous étudierons les modèles de gestion collective (création de biens communs) tant au niveau de la propriété intellectuelle (centre d'échange de brevets) qu'au niveau de l'organisation de la recherche (consortium et open science). Enfin, dans une quatrième section nous présenterons les premiers résultats d'une enquête réalisée auprès d'une trentaine de laboratoires français.

2. ANTI-COMMUNAUX ET DROIT DE PROPRIÉTÉ INDUSTRIELLE (DPI)

A la fin des années 1990, en plus des impacts liés à la coexistence du brevet et du Certificat d'Obtention Végétale (COV) pour protéger les innovations variétales et de la multiplication des demandes de droits déposées, les scientifiques, tant publics que privés, ont identifié un changement profond du paradigme scientifique : un gène pouvait intervenir dans plusieurs fonctions et parallèlement une fonction pouvait dépendre de l'interaction entre plusieurs gènes [15]. Dans ce contexte d'interactions complexes entre gènes et fonctions, les économistes ont mis en évidence des situations de « buissons de brevets » qui sont potentiellement à l'origine de blocages.

2.1. « Buissons de brevets » et enjeux pour la recherche

Shapiro [32] définit les « buissons de brevets » comme des situations où une innovation dépend d'un très grand nombre de brevets antérieurs. Pour continuer à innover, l'innovateur doit avoir accès à l'ensemble des licences de ces brevets. Cette situation engendre différentes formes d'inefficacité lorsque les ayants-droit (*claimants*) sont trop nombreux. Tout d'abord la recherche des ayants-droit, puis la négociation et le suivi de nombreux

The presence of different kinds of institutions holding those rights that are heterogene in goals, norms, and managerial practices and experiences; The uncertainty over the value of rights, which is acute for upstream discoveries and research tools.

contrats de licences parallèles entraîne des coûts de transaction élevés. De plus l'incertitude sur l'existence d'ayants-droit non identifiés en amont crée un risque juridique de « hold-up » susceptible de décourager les innovateurs en aval. Enfin chaque ayant-droit, s'il accorde une licence indépendamment des autres, aura tendance à demander un prix trop élevé, de sorte que l'accumulation excessive de royalties risque en fin de compte de décourager à nouveau l'innovation.

Ainsi, le GoldenRice utilise des technologies qui sont elles-mêmes protégées par des brevets. La production commerciale du GoldenRice nécessite une négociation de licences avec les titulaires (plus d'une dizaine) de plus de 70 brevets relevant de 5 champs technologiques : les méthodes de transformation, les marqueurs de sélection, les promoteurs, les peptides de transit et les gènes de biosynthèse du carotène [17]. Ces enchevêtrements de droits existent également du fait que l'on peut protéger les variétés végétales par COV et/ou par brevet selon les pays (exemple, le cas des USA, encadré 1). D'ailleurs, Trommetter [38] explique que : « Sans aller jusqu'à recourir à la licence obligatoire de l'Etat pour des motifs d'intérêt général, il existe des flexibilités effectivement mises en œuvre pour favoriser la co-existence entre COV et brevet : les licences croisées obligatoires pour **régler les différends privés** dans la directive européenne 98/44[‡] ; la non protection du contenant d'une innovation brevetée (transposition de la directive 98/44) qui vise à garantir un accès à la diversité génétique y compris des variétés OGM qui intègrent un (plusieurs) gène(s) breveté(s) (art L 613 5-3 de la loi du 8 décembre 2004). Ces licences garantissent une protection de l'innovation avec un « libre accès rémunéré » aux ressources génétiques, retrouvant ainsi les résultats de Scotchmer [28], [29]. » Scotchmer cherche en effet les caractéristiques qui vont rendre une licence de brevet optimale pour la société, lorsque les innovations sont cumulatives. Le contrat de licence doit assurer à la fois des incitations pour l'innovateur amont à réaliser des innovations, et pour l'innovateur aval à assurer des prolongements. Cela

‡ Ces licences garantissent une protection de l'innovation avec un « libre accès rémunéré » aux ressources génétiques. Ainsi, lorsqu'un innovateur ne peut pas réaliser une innovation sans porter atteinte à un brevet antérieur, il peut demander une licence de dépendance pour l'exploitation non exclusive de l'invention protégée par ce brevet, moyennant une redevance appropriée et/ou une licence croisée entre les deux innovations. Si le détenteur d'un brevet refuse de céder une licence de dépendance alors la directive européenne 98/44 prévoit de la rendre obligatoire sur décision publique (le Tribunal de Grande Instance en France). Même si l'instauration d'un système de licences de dépendances obligatoires entre variétés protégées par COV et OGM protégés par brevet limite les risques de blocages des recherches ultérieures, les semenciers passent d'un système de libre accès gratuit à un système de libre accès payant. L'accès à la diversité génétique n'est plus automatique, ni gratuit, mais il reste une priorité en Europe.

paraît théoriquement simple, mais plus délicat à mettre en œuvre dans la société principalement du fait d'asymétrie d'information entre les acteurs.

Encadré 1 : Protection des variétés végétales aux USA : la co-existence de plusieurs systèmes.

Trois types de protection existent aux USA pour protéger les variétés végétales :

Le PPA (Plant Patent Act, 1937), géré par l'US Patent Office, protège exclusivement les variétés d'espèces se reproduisant par multiplication asexuée, essentiellement les plantes ornementales et les arbres fruitiers puisque les espèces à tubercules (pomme de terre et topinambour) sont exclues. Cette protection relève bien du brevet, en particulier, elle n'autorise pas l'accès libre à la variabilité génétique de la variété protégée en vue d'un nouveau cycle de sélection.

Le PVPA (Plant Variety Protection Act) qui délivre un Certificat d'Obtention Végétale (COV) selon la Convention UPOV (Acte de 1978) protégeant les variétés des espèces se reproduisant par graine, et qui prend donc en compte l'**exemption du sélectionneur**, à savoir l'accès libre par tout sélectionneur à la variabilité génétique de la variété protégée en vue d'un nouveau cycle de sélection.

Le Brevet industriel, qui est ouvert depuis 1998 (suite à une action de Pioneer) à la protection des variétés végétales à reproduction par semences, qui constitue une protection plus forte que le COV (pas d'accès libre à la variabilité génétique) et auxquels les sélectionneurs américains ont de plus en plus recours sur le territoire des USA.

Bien sûr, la cohabitation de ces trois systèmes est propre aux USA. Dans l'Union Européenne et la quasi-totalité des autres pays ou organisations internationales parties à la Convention UPOV, la protection des variétés végétales par COV exclut tout autre système de protection.

Le Certificat d'Obtention Végétale (COV de l'Union Professionnelle des Obtenteurs Végétaux, UPOV) est un système de protection des variétés de semences à usage agricole et agro-alimentaire, qui confère à son obtenteur un monopole d'exploitation sur une variété végétale mais garantit l'accès automatique et gratuit aux ressources génétiques qui la composent, et à la nouvelle ressource génétique qu'elle constitue, à des fins de recherche et de sélection variétale. L'objectif est de créer un pool génétique commun auquel contribue chaque nouvel obtenteur. Le COMEPRa recommande aux chercheurs de l'INRA de veiller à protéger leurs obtentions par COV.

Dans la version de 1991, on peut citer trois principales limitations à l'accès à l'invention "variété végétale" :

* le privilège des fermiers de réensemencer leurs récoltes est optionnel et peut conduire au paiement de compensations par les agriculteurs aux semenciers en cas de "triage à façon". Cette clause de compensation est limitée aux grandes exploitations dans la directive 98/44 de l'UE ;

* la réserve de l'obteneur est limitée par la clause de dépendance pour les variétés "essentiellement dérivées". Cette dépendance est basée sur la mesure de distances génétiques entre variétés. Cela revient à étendre la protection de l'innovation avec l'introduction de la notion de "différentiation minimale" des produits [13].

La levée de l'interdiction d'une double protection (brevet et COV) sur des espèces ou des genres végétaux

Source : Version adaptée de Teyssendier de la Serve et Trommetter [34]

Outre les caractéristiques propres au domaine technique concerné, ces enchevêtrements peuvent provenir de droits trop facilement accordés, de revendications trop facilement acceptées (extension du brevet) ou d'une combinaison des deux. Dans les biotechnologies agricoles, il s'agit sans aucun doute d'une combinaison des deux effets. Il existe une fragmentation des droits de plus en plus grande associée à des « buissons de brevets » (dépendance de plus en plus forte entre les recherches et risques de blocage lors de tentatives d'achat de licences) et à des exemptions pour la recherche de plus en plus limitées. Depuis la perte du procès par la Duke University face à Madey, qui fait aujourd'hui jurisprudence, l'exemption pour la recherche aux Etats-Unis est limitée à sa plus stricte expression. Au niveau théorique, Bessen [2] montre que ces situations de hold-up peuvent perdurer du fait qu'une entreprise peut refuser d'accorder des licences *ex-ante* lorsqu'elle ne dispose pas de suffisamment d'information sur les coûts de développement de l'innovation aval. Choi [5] montre les liens complexes entre licences croisées, probabilité de dépendance, probabilité de succès des brevets et enjeux de litiges dans un contexte de droit de la concurrence. Enfin, Shane *et al.* [31] proposent de durcir les règles en cas de litiges (pour qu'elles deviennent particulièrement coûteuses pour le perdant), car ils ont montré qu'il existait une relation négative entre litiges et valorisation des brevets dans les universités américaines. Mais Trommetter [39] a montré que pour que ce durcissement soit optimal, il faut la conjugaison de deux situations : « *Reinforce patent offices and competition institutions and reinforce sui generis rights in the plant varieties so that they continue to fulfill their function first "to share genetic diversity" without risk of private appropriation of specific characteristics ; Reinforce conditions of granting patents particularly on the gene sequences by limiting the claims to (with) the functions experimentally proven to avoid the intellectual property right overlapping.* » Dans ce contexte M.A. Hermitte et P.B. Joly dès le début des années 90 proposaient un système de licence de dépendance obligatoire pour limiter les effets pervers de brevets trop nombreux et mal identifiés (ce système a été repris en partie dans la directive européenne 98/44 présentée ci-avant). Dans ce domaine, la législation française est allée encore plus loin avec l'article L613-2-1 créé par Loi n°2004-800 du 6 août 2004 sur la bio-éthique, établissant la non dépendance d'une revendication ultérieure portant sur une séquence incluse dans un brevet si cette revendication est nouvelle§.

§ Article L613-2-1 créé par Loi n°2004-800 du 6 août 2004 - art. 17 (V)

Les laboratoires publics et privés se sont parallèlement rendu compte qu'un grand nombre de ces brevets sont sous licences exclusives avec de grands groupes qui peuvent de ce fait refuser au titulaire du brevet de céder toute autre licence, y compris dans des secteurs où le grand groupe n'intervient pas. Les licences exclusives sont présentées, y compris par la théorie économique, comme indispensables pour inciter les sociétés à consacrer des ressources importantes au développement d'innovations. Néanmoins, comme dans la mise en œuvre des brevets, il faut analyser les conditions de l'exclusivité. D'autant que comme le notent Weil *et al.* [46] : « cette position a largement évolué depuis, avec l'affirmation que la recherche, en particulier la recherche fondamentale, doit être considérée comme un bien public qui est crucial pour le développement économique à long terme. Il convient donc aussi de veiller à éviter les appropriations abusives, dans la mesure où elles peuvent constituer un frein à la bonne exploitation des résultats acquis ou à l'exploration de nouvelles voies de recherche. »

2.2. Les effets des « buissons de brevets » : des controverses

Pour analyser les effets nets des « buissons de brevets », nous avons retenu un certain nombre de recherches dont les objectifs peuvent se révéler très divers. Une des publications les plus citées est celle de John Walsh, *et al.* [43]. Ce travail repose sur des données américaines (USA), portant sur plus de 1300 laboratoires publics et plus de 300 laboratoires privés. Leur objectif est de chercher s'il existe un lien entre blocage dans l'accès aux moyens de la recherche et propriété intellectuelle. Ils montrent que les barrières à l'accès qui peuvent apparaître à la suite de brevets peuvent être évitées : *“by adopting “working solutions,” such as going offshore, inventing around the patent, licensing, using public databases and research tools, or simply using the invention without obtaining permission, i.e. infringing the patent”*. Une synthèse sur des études comparables dans d'autres organisations de recherche a été publiée par les mêmes auteurs en 2005 [45]. Ils retrouvent les mêmes résultats. Néanmoins, quelle est la

"La portée d'une revendication couvrant une séquence génique est limitée à la partie de cette séquence directement liée à la fonction spécifique concrètement exposée dans la description.

Les droits créés par la délivrance d'un brevet incluant une séquence génique ne peuvent être invoqués à l'encontre d'une revendication ultérieure portant sur la même séquence si cette revendication satisfait elle-même aux conditions de l'article L. 611-18 et qu'elle expose une autre application particulière de cette séquence."

portée de ces “*working solutions*”? Sont-elles de vraies solutions à la tragédie des anti-communiaux ou sont-elles potentiellement des tragédies en elles-mêmes ?

Certains notent que les cas d'étude proposés par Walsh *et al.* [43] ne sont pas convaincants. Ils argumentent que les différents outils de contournements sont eux-mêmes coûteux, ce qui limite le nombre de compétiteurs qui ont les capacités d'entrer sur le marché, donc de participer à une course à l'innovation. D'autres argumentent que ces coûts de contournement contraignent le type d'objectifs de recherche qui peut être poursuivi (principalement en terme de taille de marché). Certaines études ont d'ailleurs donné des résultats contradictoires à ceux de Walsh et Asher [44]. Ainsi, Hansen *et al.* [12] étudient une synthèse réalisée auprès de l'“*American Association for the Advancement of Science*” : “*They found that 40 percent of respondents reported difficulties in obtaining access to patented technologies, and over half of these said their research was delayed or changed course as a result.*” De même, Verbeure *et al.* [42] expliquent que : “*it is possible that overlapping patents might occur on the genetic data necessary to test for « hereditary non polyposis colorectal cancer*”. Merz *et al.* [21] montrent que “*30% of clinical laboratories reported not developing or abandoning testing for a gene associated with haemochromatosis once the patent issued. Another investigation of over 100 laboratories found that 25% of respondents discontinued clinical testing because of a patent or license; although the BRCA1/2 test was the most commonly identified, eleven other genetic tests ceased to be offered because of the existence of patent rights*”. Enfin, Maurer [20] donne l'exemple de la « *mutation database initiative* » où il existait une volonté de centraliser les données sur des mutations de gènes de l'humain qui appartiennent à une centaine de biologistes. Il s'agissait de regarder comment réorganiser les bases de données et les collections et comment financer cette réorganisation. Cette initiative ne s'est pas réalisée car les acteurs les plus petits ont refusé de participer au projet. Ils avaient peur de voir leur travail approprié par le secteur privé sans qu'ils soient rémunérés à la juste hauteur de leur travail. L'enjeu est donc fort quant à la définition des conditions de participation à la réalisation d'un bien collectif, voire d'un bien public. Il est d'autant plus intéressant de noter que la tragédie des anti-communiaux peut contraindre les concurrents pour poursuivre des recherches avales, mais également contraindre les détenteurs de droits enchevêtrés lorsque l'on ne parvient pas à mettre en œuvre une gestion collective de la propriété intellectuelle qui soit acceptée par tous (voir point 3.1. ci-dessous).

Dans un autre registre, Stern et Murray [33] étudient un ensemble de 169 “*patent-paper pairs*”. Pour tester l'existence d'une théorie des anti-communiaux, les auteurs analysent les taux de citations des papiers scientifiques avant et après le dépôt de brevets. Si la théorie est vérifiée, alors : “*the citation rate to a scientific publication should fall after formal IP rights*

*associated with that publication are granted.” Dans leurs analyses empiriques, ils montrent que : « there is a robust evidence for a quantitatively modest but statistically significant anticommons effect. ». D’ailleurs : “the citation rate after the patent grant declined by between 9 and 17%”, with the decline becoming “more pronounced with the number of years elapsed since the date of the patent grant.” Pour relativiser ces effets, citons Rodriguez *et al.* [23], [24] qui montrent qu’ils sont incapables de conclure sur le fait que “agreements signed by industry and government affect research agenda setting in academia”. Par ailleurs, il y a sans doute un effet "perte de la nouveauté" de la citation qui fait que la publication est de moins en moins citée.*

Tableau I: Motifs pour breveter : enquêtes auprès de l’Univ. Carnegie Mellon (1991-1993).

	Produits	Processus
Empêcher la copie	96 %	78 %
Bloquer les brevets	82 %	64 %
Empêcher les procès	59 %	47 %
Utiliser dans les négociations	48 %	37 %
Augmenter la réputation	48 %	34 %
Revenus de licence	28 %	23 %
Mesure de la performance	6 %	5 %

Il existe donc diverses controverses quant aux effets de la multiplication de brevets sur la recherche en termes de blocage. Tout d’abord est-il opportun de parler de blocage, à la lumière des travaux publiés ? Il apparaît plus opportun de parler de retards, de réorientation des recherches ou de contournements des brevets existants (voir tabl. I). Cette controverse est d’autant plus stratégique que certains papiers concluent sur le fait que le blocage de la recherche n’est pas avéré alors que d’autres affirment le contraire. S’il n’y a pas blocage, cela justifie le système actuel de protection des innovations. S’il y a blocage cela justifie d’en changer. Or, si l’on analyse par exemple, les retards qui peuvent apparaître, alors tout le monde sera d’accord sur leur existence et le fait que ces retards peuvent avoir des conséquences très négatives sur la recherche, tant sur la vitesse dans la course à l’innovation que sur les coûts de la recherche. Aujourd’hui, il semble bien que ces retards dans la recherche sont plus souvent le résultat de conditions d’accès à des moyens de la recherche, plutôt que de l’accès à des licences de propriété intellectuelle. Mais comme le rappelle Trommetter [37] : c’est l’évolution de la propriété intellectuelle (les brevets sur le vivant) qui a conduit les pays à revendiquer leur souveraineté nationale sur leurs

ressources génétiques dans la cadre de la Convention sur la Diversité Biologique (ces dernières étant jusque là « patrimoine commun de l'humanité »). En effet, l'offreur de ressources génétiques pouvait se voir refuser l'accès (comme demandeur) à des innovations contenant « ses » ressources. La propriété sur le vivant est donc l'un des éléments à l'origine des contraintes d'accès aux ressources génétiques et aux autres moyens de la recherche. Enfin, la multiplication des droits et de leurs revendications larges ont également conduit à la multiplication des litiges. Cohen [6] note que : *“Although there is a little systematic evidence that these concerns (anti commons, entry barriers, ...) have materialized in any substantial way, vigilance is nonetheless warranted.”* Bien souvent une des solutions est d'ignorer les brevets car comme le rappelle Cohen [6] : *“From scientists at drug firms who knew they were infringing research tool patents, we learn the risk of a lawsuit was slight since there is a six year statute of limitation of patent infringement cases. Typically, the infringement would only be detectable when a drug was actually commercialized, commonly a decade or more after the infringement occurred.”* Cela rejoint les travaux de Walsh *et al.* [43] qui proposaient de ne pas se soucier de la dépendance à un brevet antérieur. Ils partent du principe que les coûts de transactions deviennent très élevés pour d'une part, évaluer le niveau de dépendance à d'autres brevets et d'autre part, de négocier des licences sur les brevets identifiés comme potentiellement bloquants. Il peut alors être moins coûteux pour les entreprises de faire des provisions pour litige, sachant que même en cas de dépendance, le détenteur initial du brevet peut ne pas identifier la dépendance (la recherche d'innovation dépendante pouvant se révéler coûteuse pour lui en termes de coûts de transaction).

3. ANTI-COMMUNAUX ET AUTRES ACTIFS DE LA RECHERCHE

Les recherches sont de plus en plus complexes et nécessitent d'accéder à des moyens de plus en plus nombreux et de plus en plus complexes. L'accès à certaines technologies (génomique, protéomique, transcriptomique,...) nécessite d'accéder à des bases de données diverses ainsi qu'à des logiciels. Les biotechnologies nécessitent également l'accès à des ressources génétiques diverses qui vont de la ressource génétique de base à des collections de mutants, ESTs (Expressed Sequence Tags), banque BAC (Bacterial Artificial Chromosomes), etc. Walsh *et al.* [43] et [45] ont ainsi mis en évidence des difficultés croissantes dans le partage et l'accès à des matériaux et outils issus de la recherche. Selon ces auteurs, ces difficultés d'accès ne sont pas le résultat de droits de propriété intellectuelle mal définis. Ils montrent que c'est surtout le processus de construction des accords de transfert de matériel (ATM) qui va conditionner l'accès.

3.1. Conditions d'accès aux collections

Dans le cas de la gestion *ex situ* des collections de ressources génétiques, il existe plusieurs statuts pour les collections réalisées avant la mise en œuvre de la Convention sur la Diversité Biologique (ONU, [22]). Aux collections privées et aux collections publiques s'ajoutent en effet, dans le cas de ressources phylogénétiques pour l'agriculture et l'alimentation, des Collections Nationales (CN) qui s'inscrivent dans un système multilatéral d'échange facilité de ces ressources biologiques, proposé par la FAO. Le Traité FAO [8] ne fait aucune distinction entre matériel pré ou post 29/12/93. Tout matériel d'une collection nationale, donc du domaine public, appartenant à la liste d'espèces de l'Annexe 1 du Traité est soumis à ce texte ! Or, à ce jour, l'existence même de ces Collections Nationales est problématique puisqu'il n'existe pas encore de législation nationale et même européenne sur les ressources génétiques en application de la CDB et/ou du Traité international. Pour le moment en France, en toute rigueur, seul la CDB s'applique : accord bilatéral, consentement préalable et partage des avantages. L'offre d'accès à des collections ne tombant pas dans le cadre du traité de la FAO et étant composée de matériel acquis après 1992 est plus complexe encore, en dehors des accords de la FAO, puisque l'accès devrait implicitement nécessiter l'accord du « propriétaire de la ressource » sur le lieu de prospection de la ressource (cette clause devant être incluse dans le contrat de bioprospection). Dans ces contextes, la définition d'accord de transfert de matériel (ATM) devient essentielle.

Les ATM ont été élaborés au début des années 90 par les Centres Internationaux de Recherche Agricole (CIRA), pour l'accès à leurs collections de ressources génétiques végétales. Généralement bilatéraux, ils sont mis en œuvre soit dans des relations entre laboratoires publics, soit entre firmes et laboratoires publics, soit entre firmes. Ainsi, dans le cas des accords multilatéraux d'échanges de l'Engagement International de la FAO [8], l'accès aux collections de ressources génétiques n'implique pas que ce soit pour tous les usages. En particulier, l'autorisation de dépôt de brevet n'est pas automatique. La signature d'un ATM, pour accéder à des collections de ressources génétiques, permet ainsi de limiter le risque qu'un innovateur puisse bloquer (par un brevet) l'usage d'une ressource sans l'accord du détenteur initial de cette ressource (un laboratoire, un pays, ...). Ces contrats régissent les droits de propriété sur les innovations qui découleraient de la collection et le cas échéant le partage des redevances, en accord avec la Convention sur la diversité Biologique signée en 1992 (WIPO, [48] ; Henry *et al.* [14]).

Dans le cas de la gestion des autres collections, l'ATM forme le cadre juridique sous lequel transite actuellement la plupart des matériels échangés

entre laboratoires de recherche. Souvent considéré comme moins formel qu'un contrat de recherche ou de licence, l'ATM gère souvent des enjeux beaucoup plus importants et peut porter des clauses de « droits de suite » lourdes de conséquences. Il semble normal que l'ATM interdise au receveur de déposer un brevet sur le matériel qu'on lui fournit. Mais, par exemple, lorsque le matériel fourni est un mutant qui permettra de découvrir la fonction d'un gène, le sésame d'un brevet, il est tentant pour le fournisseur de revendiquer cette fois des droits sur le brevet en question, par exemple une licence d'exploitation. Dans l'ATM GARLIC, la société SYNGENTA revendique beaucoup plus (encadré 2). Or une majorité d'organismes publics ont accepté de signer cet accord pour que leurs chercheurs bénéficient d'un avantage compétitif.

Encadré 2 : Détenteur de mutants, contrats d'accès et revendications.

SYNGENTA lie la fourniture de mutants d'insertion d'*Arabidopsis* (collection GARLIC) à la signature d'un ATM à signer stipulant notamment :

- l'interdiction au chercheur receveur de travailler en collaboration avec les équipes d'autres organismes ou d'autres sociétés, sans l'accord de SYNGENTA ;
- l'obligation de transmettre, à titre gratuit et au moins une fois par an, à SYNGENTA tous ses résultats obtenus grâce au mutant sans que cette société soit obligée de préciser selon quelles modalités et avec qui elle les utilisera ;
- l'obligation de soumettre tous ses projets de publications à SYNGENTA, qui peut en faire supprimer toutes les informations confidentielles qu'ils contiendraient ;
- la possibilité pour SYNGENTA de breveter tout résultat que le receveur ne souhaiterait pas protéger lui-même ;
- l'obligation au receveur de consentir un droit de premier refus à SYNGENTA pour toute exploitation commerciale des résultats (en d'autres termes une priorité pour l'exploitation sous licence). Ceci interdit donc toute mise en concurrence de partenaires pour l'exploitation commerciale : le receveur sera obligé de négocier une licence au profit de SYNGENTA, exclusive si SYNGENTA le demande.

Source Teyssendier de la Serve et Trommetter [34].

Pour organiser et simplifier les échanges et promouvoir la circulation des ressources entre les laboratoires de recherche académique, le NIH a proposé un modèle d'accord (Uniform Biological Material Transfer Agreement, UBMTA), partie d'une convention que peuvent signer les instituts de recherche. Ce UBMTA interdit au receveur de breveter le matériel et d'en faire usage commercial sans l'accord de l'obteneur. Par contre il le laisse libre de déposer des brevets sur ses inventions réalisées à partir du matériel reçu. Pour mettre à disposition les mutants d'insertion de la collection de Versailles, l'INRA ajoute actuellement à un ATM de ce type une clause demandant de bénéficier d'une licence gratuite pour la recherche dans ses laboratoires sur les brevets passés par les destinataires des mutants sur des

inventions réalisées grâce à ces matériels. Peut-on en conclure que les ATM sont aussi néfastes que les brevets ou doit-on en conclure que les ATM sont une réponse à une politique de brevet trop laxiste dans les extensions accordées ? On peut conclure que l'ATM, au même titre que les accords de licences sur les brevets, serait un bon compromis à condition qu'il soit négocié gagnant/gagnant par les parties. Or, dans le cadre de la brevetabilité du vivant, la réglementation a évolué : les conditions de brevetabilité et d'extension sont plus restrictives aujourd'hui. Par contre l'accès aux autres moyens de la recherche repose sur les contrats de type ATM qui relèvent du droit privé, les clauses d'un contrat étant légales tant qu'elles ne sont pas contraires à la loi. C'est sans doute l'une des raisons pour lesquelles les contrats d'accès aux moyens sont aujourd'hui si restrictifs, c'est aussi pour cela qu'il faut revoir le rôle de l'information et de la formation pour les gestionnaires des moyens de la recherche.

3.2. Conditions d'accès aux bases de données et aux logiciels

Les Etats-Unis protègent les bases de données par droit d'auteur. L'Europe le fait aussi mais y a ajouté un droit *sui generis* (UE, [41]), qui protège le créateur contre "l'extraction et la réutilisation non autorisée de la totalité ou d'une partie substantielle du contenu d'une base de données". Maurer et Scotchmer [19] s'inquiètent de la portée réelle de ces nouveaux droits de propriété du fait de la multiplication des bases de données en particulier dans les biotechnologies, du croisement de résultats issus de bases de données multiples (recherche d'homologies) et des nouveaux types d'organisation de la recherche en bioinformatique (principalement les consortium). Les logiciels sont protégés par droit d'auteur et, de plus en plus, par le brevet. En effet les logiciels sont officiellement brevetables et couramment brevetés aux Etats-Unis. En Europe, des négociations sont en cours sur la brevetabilité des logiciels, mais de fait de nombreux logiciels sont déjà brevetés.

On a qualifié les bases de données scientifiques « d'accélérateurs non linéaires de la recherche » (Cerf 1999, *in* Wouters, [49]). Le principe de base qui s'impose comme référence internationale est que les données issues de la recherche publique sont un bien public produit pour l'intérêt public (OCDE 2003). Aux USA, la loi fédérale a étendu le "*Freedom Of Information Act*" au domaine de la recherche (circulaire A110, 1999 de l'*Omnibus Appropriation Bill*) ; cependant l'application de ces lois doit être conciliée avec les dispositions du *Bayh Dole Act* qui ouvre aux organismes publics le droit de déposer des demandes de brevet. En conséquence, les données issues de résultats de recherche publique doivent être mises à disposition de façon ouverte et le plus largement possible mais sous réserve de prévoir une période limitée d'exclusivité au profit des principaux obtenteurs, ainsi que

respecter confidentialité, informations privées, restrictions de sécurité nationale et règles de propriété intellectuelle (OCDE 2003). Ces principes étant déclinés, rien n'est réglé et un juste compromis reste donc affaire d'évaluation pour une politique de service public. Dans le cadre de l'accès aux bases de données détenues par des acteurs privés, les conditions d'accès peuvent être multiples (encadré 3) : avec un coût d'accès ; avec un tarif binôme (accès et royalties) ; avec des revendications en nature (octroi de licence sur les innovations avales, ...).

Encadré 3 : Conditions d'accès à des bases de données.

* Monsanto donne un accès gratuit à sa base de données mais sous contrainte de la signature d'un contrat type. Dans ce contrat, l'accès à la base de données est, article 4 sur les droits de propriété, relativement contraignant pour le signataire. Il s'engage à donner une licence non exclusive à Monsanto sur tout brevet qu'il déposerait à partir des données de la base, les royalties étant à négocier (voir article 4.4.). Cela garantit à Monsanto d'être toujours à la pointe de la recherche. On aboutit à un arbitrage entre faire de la recherche (ou tout du moins piloter la recherche) ou délocaliser la recherche (Ambec et Poitevin 2001), en sachant que dans ce dernier cas Monsanto s'est assuré d'être informé sur les innovations et de pouvoir accéder à ces innovations, par des accords de licences, ce qui ressemble fort à la mise en œuvre, par contrat, de licences obligatoires.

* La recherche en bioinformatique nécessite d'avoir accès à un grand nombre de bases de données, cet accès est devenu un enjeu majeur pour la réalisation rapide d'innovations, particulièrement dans les maladies génétiques. Dans le cadre des génomes microbiens, deux bases de données développées par Protéome : la base de données YPD qui est une source de référence pour le séquençage de *Saccharomyces cerevisiae* et la base wormPD, pour le séquençage de *Caenorhabditis elegans*, ont été récemment rachetées par Incyte pharmaceuticals. Pour accéder à ces deux bases, dont l'accès était initialement gratuit, Incyte pharmaceuticals exige paiement d'une licence (Aimé-Sempé, 2002). Une telle situation est à la fois source de coûts de recherche supplémentaires (à budget de recherche constant dans les laboratoires) et de coûts de transaction élevés si l'ensemble des propriétaires de bases imitent Incyte pharmaceuticals, en faisant payer l'accès à leur base.

Source : Henry *et al* [14]

Pour exemple, un délai maximum de six mois a été convenu dans le cadre du consortium Génoplante entre le moment où des données produites dans le cadre du consortium sont mises à disposition des membres de Génoplante et la date de leur publication effective dans le domaine public. Il demeure que la publication dans les bases de données (en particulier la publication de grands ensembles de résultats) ne doit pas se faire au détriment des programmes de recherche concernés ni de celui des brevets potentiels. L'accès aux principales bases de données de biologie moléculaire est très généralement public et libre pour la recherche académique fussent-elles produites par des privés ; il en est ainsi des données de polymorphisme

d'*Arabidopsis thaliana* obtenues par l'entreprise CEREON. Ces données peuvent également ne pas être nécessairement mises dans le domaine public, leur accès peut alors être soumis à conditions (encadré 2).

4. LIMITER LES ANTI-COMMUNAUX: METTRE AU POINT DES OUTILS DE GESTION COLLECTIVE

Comme le rappelle Scharff [27] dans le cas de la gestion des terres : *“before solutions to the anti-commons can be adequately explored, it is important to look below the surface to determine the true nature of any anticommons problem that may exist.”* De plus en plus d'opérations de gestion collective sont mises en œuvre tant dans la gestion collective de la propriété intellectuelle que dans la gestion collective de la recherche. Nous verrons que l'objectif de « limiter les risques liés aux situations d'anti-communales » n'est que l'une des raisons pour mettre en œuvre de telles opérations.

4.1. Gestion collective de la propriété intellectuelle

La gestion collective de la propriété intellectuelle doit être conforme aux règles édictées par les juridictions en charge de la mise en œuvre des réglementations anti-trust (par exemple, le conseil de la concurrence en France). Cette gestion collective doit aboutir à une amélioration du bien-être social et plus particulièrement du surplus des consommateurs. Donc, en l'absence d'une situation d'« abus de position dominante », les acteurs de la recherche peuvent envisager la création d'un club de gestion de leurs brevets, chacun restant individuellement propriétaire de ses brevets. Ces clubs visent à gérer l'accès et les usages d'un portefeuille collectif de brevets pour : identifier les complémentarités à l'intérieur du portefeuille pour construire des innovations basées sur des « grappes technologiques » (technologies modulaires identifiées *ex-post* aux dépôts des brevets concernés)** ; identifier les complémentarités avec d'autres portefeuilles de brevets publics ou privés ; faciliter l'accès aux licences de brevets complémentaires détenus par d'autres institutions ou entreprises. Participer au club augmente le pouvoir de négociation des acteurs. Certains de ces clubs sont aujourd'hui opérationnels, il s'agit des centres d'échanges des brevets (PIPRA, Atkinson *et al.* [1], EPIPAGRI, voire open source génomique) dans le domaine des biotechnologies végétales sont un bon exemple. Graff *et al.* [10] expliquent que la mise en œuvre d'un centre d'échange de brevet pour la gestion collective des brevets des universités

** Dans les innovations, identifiées comme des « technologies modulaires standards », les recherches sont construites *ex-ante* sur des modules brevetables et complémentaires.

américaines dans le domaine des biotechnologies végétales repose : sur la limitation des risques liés aux enchevêtrements de brevets, pour rechercher et proposer des licences de paniers de brevets attractives et/ou réaliser des innovations dans le secteur public (sur les variétés végétales orphelines ou en direction des PVD) ; sur un pouvoir de négociation plus important pour avoir accès aux licences de brevets détenus par les grandes multinationales de biotechnologies Verbeure *et al.* [42] expliquent que : « *If a patent thicket arises, an HNPPC patent pool, encompassing essential genomic patents, could help to eliminate the thicket and render proprietary genomic data more accessible for use.* »

Au niveau théorique, la gestion collective de la propriété intellectuelle pose la question de sa conformité avec les lois anti-trust, *i.e.* principalement comment limiter les risques de collusions ? Les économistes ont recherché et identifié les situations dans lesquelles la constitution d'un panier de brevets est économiquement et socialement efficace. Shapiro [32] puis Lerner et Tirole [18] montrent que les licences groupées ne doivent porter que sur des brevets complémentaires entre eux et indispensables à la technologie transférée. Ces derniers démontrent également qu'une clause autorisant chaque participant à fournir de manière indépendante une licence sur ses propres brevets permet de prévenir les licences abusives. Bessen [2] renforce cette idée de création de pool pour limiter les effets pervers de brevets passés, car il montre qu'il est souvent difficile de distinguer le bon du mauvais brevet ce qui a des effets dissuasifs.

4.2. Gestion collective de la recherche

Un élément favorable à la gestion collective de la recherche est que les recherches sont de plus en plus complexes et coûteuses avec le développement de toutes les technologies : génomique, protéomique, transcriptomique... La prise de conscience de la complexité de la recherche est liée au fait que des brevets sont tombés dans le domaine public avant même d'avoir pu être valorisés. Cela a conduit selon Wright et Pardey [50] à un double mouvement : concentrations dans les biotechnologies agricoles ; mise dans le domaine public de grandes masses de données génétiques. Pour contourner le risque d'être bloqué -retardé- dans la recherche, les organismes de recherche tant publics que privés, proposent la constitution d'un nouveau mode organisationnel de la recherche basé sur de nouvelles relations collectives et coopératives dans les phases précompétitives de la recherche. Au niveau théorique, la coopération en recherche est basée sur la théorie de la valeur de l'information et en théorie de l'organisation industrielle sur les *spillovers*. L'objectif est de définir les règles de fonctionnement des consortiums de recherche pour qu'ils soient économiquement efficaces et qu'ils ne tombent pas sous le coup des règles du droit de la concurrence : accès privilégié des membres du réseau à

certaines informations ; retard dans la diffusion de certaines informations à l'intérieur du réseau ; peu ou pas de copropriété de brevets. Des auteurs comme Roijackers et Hagedoorn [25] et Gay et Dousset [9] ont observé une augmentation dans la mise en œuvre de partenariats entre acteurs publics et privés. Cela est d'autant plus compliqué à mettre en évidence que Rosenkranz et Schmitz [26] ont démontré les difficultés à diffuser les connaissances dans les accords de coopérations entre universités et secteur privé. Au niveau théorique on trouve les papiers sur les « *joint patent* » qui sont de plus en plus nombreux (Hagedoorn, [11]), mais également les travaux de Cohendet *et al.* [7] qui voient le brevet comme un outil institutionnel de coordination en plus d'être un simple outil d'incitation (cela nécessitant d'avoir plusieurs types de propriété intellectuelle selon le niveau d'avancement de la recherche). Bittner [3] insiste d'ailleurs sur la nécessité de partager la propriété intellectuelle lorsqu'il y a une réelle contribution de chacun des partenaires. Il s'agit de partager les coûts mais également les résultats (cela rejoint bien des analyses qui ont été faites sur des cas comme Génoplante en France).

A titre d'exemple, l'objectif principal de la création de consortiums sur la génomique structurale est d'accélérer la production de nouvelles structures de protéines résultant des croisements de différentes séquences partielles de gènes que l'on trouve dans les bases de données publiques [47]. Dans le cadre du consortium sur les *Single Nucleotide Polymorphism (The SNP Consortium, TSC)* plus de 300 000 SNP humaines ont été mises dans le domaine public (Williamson [47])^{††}. Ces recherches sont financées sur des fonds privés et sont réalisées par des instituts de recherche publics et privés, l'enjeu étant grâce à ces données publiques de découvrir plus rapidement des gènes intéressants pour la mise au point de médicaments (d'innovations brevetables), en croisant différentes bases de données sans avoir à se soucier de l'existence ou non d'un titulaire de droits de propriété intellectuelle sur les SNP utilisées (laisser un libre accès aux séquences pour toutes fonctions). Lorsque des ressources sont produites par des consortiums publics/privés, la question des « droits de suite » est un point particulièrement délicat (encadré 4). Généralement lorsqu'une ressource stratégique comme des

^{††} Les membres de ce consortium sont : Wellcome Trust ; APBitech ; AstraZeneca PLC ; Aventis ; Bayer AG ; Bristol-Meyers Squibb Company ; F. Hoffmann-LaRoche ; Glaxo Wellcome PLC ; IBM ; Motorola Novartis ; Pfizer Inc ; Searle ; SmithKline Beecham PLC. Les travaux financés par le consortium sont réalisés par 4 instituts de recherche : Stanford Human Genome Center ; Washington University ; Wellcome Trust's Sanger Centre et Whitehead Institute. La participation financière de chaque membre est de US\$ 3 Millions ; tout nouveau participant au consortium doit verser cette somme comme droit d'entrée, le Wellcome trust ayant apporté US\$ 14 Millions (Williamson, [47]).

mutants d'insertion est produite, un délai d'exclusivité est prévu au-delà duquel la ressource est rendue publique (passage d'un bien de club à un bien public, voir précédemment). Walsh *et al.* [43] insistent sur le fait que : « *For example, with substantial public, private and foundation support, public and quasi public databases have been created, making genomic information widely available. The NIH has also negotiated with owners of foundational technologies, such as stem cells or genetically altered mice, to ease publicly funded researcher's access to these important research tools. Scientific journals have pushed authors to deposit sequences in publicly available databases as a condition of publication.* » On peut noter que ce sont des initiatives individuelles pour la réalisation d'un collectif. Est-on en mesure institutionnellement et politiquement d'aller plus loin ? De même, dans le contrat « Génoplante » (consortium français de recherche entre des laboratoires publics et des entreprises privées) il est prévu que les brevets plus fondamentaux seront déposés par les instituts de recherche alors que les brevets plus en aval seront déposés par les entreprises privées. Par ailleurs, il est prévu d'avoir des licences obligatoires gratuites entre les membres du consortium. Gay et Dousset [9] posent la question de la constitution des réseaux, des alliances et des questions de partage de la propriété. Enfin, de nouveaux modes organisationnels se mettent en place, par exemple, le modèle « *open source biotechnology* » (en Australie) qui assure l'accès aux séquences génétiques par le biais de contrats (de licences), chaque laboratoire s'engageant par contrat à ne pas bloquer l'accès aux séquences dans les innovations, sans pour autant garantir la gratuité de l'accès à l'innovation. Ce modèle est à rapprocher de la loi française déjà citée plus haut, qui est bien plus ouverte puisqu'elle ne prévoit aucune dépendance si la nouvelle application, bien que portant sur la même séquence et fonction, se révèle indépendante des revendications initiales.

Donc, à l'intérieur des consortiums et des activités de recherche en réseau, la définition des règles est un préalable à chaque recherche car, comme le montrent Cassier et Foray [4], une absence de règle au départ conduit automatiquement à des comportements opportunistes et à la fin des réseaux de recherche. L'introduction de règles est donc nécessaire pour limiter l'aléa moral -une asymétrie d'information sur l'action réelle des agents-, les actions des partenaires ne pouvant pas être toutes observées (encadré 4). Il y a en effet une double incertitude sur l'activité réelle d'évaluation génétique et sur la diffusion de l'information aux membres du réseau. Les règles sur les conditions de participation voire d'exclusion du consortium sont donc centrales. Dans ce système de coordination de la recherche, il n'y a pas de dépôt de brevets sur les résultats des phases pré-compétitives de la recherche, donc un accès libre mais contractualisé à la diversité génétique alors que le modèle COV implique que l'accès aux variétés protégées en vue de la création variétale est, pour tout sélectionneur (personne morale ou physique, publique ou privée !), totalement libre, donc

non contractuel et sans coûts de transactions. Le brevet sur le vivant et surtout ses excès conduisent à une contractualisation (économiquement et socialement nécessaire) du libre accès, seule garante de la non appropriation privative des ressources génétiques et d'un retour à un modèle organisationnel proche du COV.

Encadré 4 : Participations à des consortiums et revendications sur les résultats de la recherche.

Mutants d'insertion d'*Arabidopsis*, ZIGIA : les mutants d'insertion du Max Planck Institute issus du programme européen ZIGIA ne peuvent être obtenus par des équipes de la recherche publique que sous réserve de la signature d'un contrat d'option de licence mondiale non exclusive au profit des partenaires privés du programme. De tels contrats sont normalement consentis aux privés ayant contribué directement et lourdement au financement d'un programme conduit dans le laboratoire public, ce qui n'est généralement pas le cas des contrats européens.

Facteurs de transcription, consortium REGIA : le programme REGIA avait pour objet la recherche systématique de facteurs de transcription chez *Arabidopsis thaliana* ; les premières versions de l'accord de consortium prévoyaient d'accorder aux partenaires privés du programme une option de licence exclusive sur les résultats. L'opposition de diverses institutions publiques dont l'INRA a fait échouer cette tentative d'appropriation.

Source : Teysseidier de la Serve et Trommetter [34]

5. LE PROTOCOLE D'ENQUÊTE ET LES PREMIERS RÉSULTATS

Les résultats présentés ici ne peuvent être que préliminaires, puisque notre recherche doit aboutir en Juillet 2008. Nous présentons donc ici le protocole de recherche et les premiers résultats que nous avons pu identifier d'une première analyse des questionnaires.

5.1. Protocole de recherche

Dans un premier temps (Janvier 2006), nous avons présenté ce projet devant le groupe « propriété intellectuelle » de l'Inra. A la suite de cette présentation, nous avons décidé de limiter le champ de notre questionnaire aux ressources biologiques animales ou végétales et aux micro-organismes qui leur sont liés (parasites, pathogènes, etc.). Sont donc exclues *a priori* de cette enquête les ressources biologiques humaines, les micro-organismes de biosynthèse industrielle et fermentaire pour l'agro-industrie, etc.

A la suite de cette présentation, nous avons construit un questionnaire sur les conditions d'accès et d'usage des différentes ressources mobilisées

par la recherche en biotechnologies (collection de matériel, bases de données, logiciels, ...). Ce questionnaire est destiné aux laboratoires publics de recherche tant dans leur rôle d'offreur que de demandeur de ces différentes ressources. Devant la complexité du questionnaire, nous avons dans l'idée d'identifier une soixantaine de laboratoires et d'organiser deux journées en Novembre 2006, une à Montpellier et une à Paris. L'objectif de ces journées était double. D'une part, présenter aux participants les enjeux de la mutualisation dans un contexte de droit de propriété évolutif dans le contexte de la mise en œuvre de la Convention de Rio [22] et du Traité international [8] (enjeux futurs associés à la certification d'origine et à la divulgation de l'origine ; évolution dans les directives européennes et leurs transpositions nationales, particulièrement au niveau de la transposition de la directive en droit français et de ses conséquences pour la recherche publique en terme d'exemption de la recherche et pour les acteurs privés en terme de périmètre de marché). D'autre part, les aider à remplir le questionnaire.

Une liste d'environ 60 laboratoires ou organismes de recherche a ensuite été établie sur les bases suivantes : représentation équilibrée des secteurs privé ou public, des domaines d'intérêt (végétal, animal, micro-organismes), des types d'activité (gestion de collection, génomique, outils bio-informatiques, etc).

Malgré le fait que les laboratoires nous aient été proposés et qu'ils n'étaient donc pas le résultat d'un choix au hasard de notre part, nous avons eu un faible taux de réponses et nous avons dû annuler les deux journées. Notre objectif a été de comprendre pourquoi un taux de réponses si faible et comment y remédier. Pour ce faire, nous avons réuni un groupe de chercheurs de l'Inra, du Cirad et de l'Ird qui eux avaient accepté de participer à nos journées.

Pour l'Inra, les non réponses sont liées à un problème de délai dans la diffusion du courrier et d'intérêt stratégique pour l'Inra. Au-delà de ce simple problème temporel, comme pour le Cirad et l'Ird, il existe des craintes de voir diffuser des comportements considérés comme stratégiques devant des collègues, mais également des craintes de voir des comportements « illégaux » mis à jour (mais non perçus comme tels aujourd'hui par le laboratoire) qui nécessiteraient alors une réorganisation coûteuse au sein du laboratoire. Cette crainte de réorganisation des comportements stratégiques des laboratoires est d'autant plus perceptible que la contrepartie à leur participation ne leur paraissait pas immédiate. Notre volonté d'information sur les enjeux à venir en termes de diffusion du matériel et de propriété intellectuelle n'a donc pas été comprise. Au final, l'étude a été réalisée au travers d'une entrevue directe auprès de chaque

laboratoire participant, ce qui a permis de collecter informations et avis sur des bases claires et homogènes.

5.2. Résultats préliminaires

Nous avons interrogé une trentaine de laboratoires qui correspondent à une gamme variée d'objectifs : des collections de plantes, des collections de graines, des collections de mutants, des collections d'ADN et de BAC, des bactéries du sol et des pathogènes. Nous avons également interrogé de manière plus spécifique des gestionnaires de bases de données en réseaux et de logiciels.

Plus de la moitié des laboratoires ont rencontré des difficultés pour accéder à ou diffuser des résultats de la recherche, y compris dans le cas de bases de données ou de logiciel. En revanche, très peu de laboratoires parlent de blocages (moins de 5 sur 30). Pour ceux qui parlent de blocage, on peut noter que dans deux cas il s'agit d'un blocage à l'accès de ressources génétiques dans les PVD (Inde et Ethiopie). Il y a donc bien des retards ou des modifications de programmes de recherche qui sont liés à des problèmes d'accès ou de diffusion des ressources biologiques. Ces retards sont peu liés à des droits de propriété intellectuelle puisque seuls 7/30 disent que leurs recherches dépendent de DPI et un seul dit qu'il ne dépend pas de DPI en ajoutant qu'« un jour on aura peut être des surprises ». Les difficultés d'accès ou de diffusion sont souvent liées à la question du statut du matériel ou de l'information (soit je ne sais pas à qui elle appartient donc je ne peux pas la diffuser, soit je ne sais pas à qui m'adresser pour y accéder directement). Cette question du statut du matériel et de l'information est primordiale pour la suite. Certains diffusent indistinctement selon les mêmes conditions des matériels acquis avant ou après 1993, d'autres diffusent en prenant le soin d'en informer le fournisseur initial, d'autres renvoient la demande à ce fournisseur initial, d'autres encore s'interdisent toute diffusion... Il y a donc un réel besoin d'harmonisation et de statut clair pour les ressources. Des lourdeurs administratives sont également mises en avant en particulier sur la responsabilité en cas de problèmes sanitaires qui seraient liées au matériel, ainsi que la mise en œuvre du protocole de Carthagène.

Dire que les DPI n'ont que peu d'effets sur les retards de la recherche n'est sans doute pas vrai. En effet, le risque de voir une ressource génétique appropriée par d'autres (brevets sur les séquences de gènes) a sans doute eu des conséquences sur les conditions d'accès aux collections quelles qu'elles soient et aux informations sur ces collections. D'ailleurs des exemples de ATM revendiquant des royalties sur des collections de mutants ont été mentionnés. On peut noter que l'accès aux bases de données est par exemple assez complexe : bases à usage local, bases à usage d'un centre,

d'une institution, biens publics, ... tous les cas ont été observés. Dans certains cas on peut se demander si l'absence de coordination dans la constitution de la base n'est pas volontaire de manière à garder le « secret ». D'ailleurs les initiatives de mise en place des bases de données harmonisées et coordonnées sont citées par plusieurs laboratoires mais leur mise en œuvre semble plus problématique.

On peut noter dans ce cadre que les ATM sont divers, ils dépendent du type de collection et des usages qui peuvent en être fait. Une première analyse nous permet de dire que les ATM sont relativement simples et peu harmonisés. Par contre, il semble qu'un recours au secret avec des contrats exclusifs soit de plus en plus fréquent. Cela signifie qu'aujourd'hui soit le matériel est facilement accessible, quasiment avec le statut d'un bien public (le coût d'accès est égal aux coûts liés aux frais pour envoyer l'échantillon), soit il est tenu secret. Il existe en outre du matériel issu de consortiums de recherche et qui reste en « bien de club » avant de devenir un jour un bien public (l'exclusivité d'accès au niveau du club pouvant durer plusieurs années). Ce type d'analyse est également valable pour l'accès aux informations contenues dans les bases de données. Cela signifie que, à l'avenir, la perception de difficultés d'accès devrait en théorie être limitée puisque l'on ignore l'existence de ce à quoi l'on n'a pas l'accès. Par contre ce recours au secret ne peut être que préjudiciable en terme de bien-être social. Cela signifie que le recours à la propriété intellectuelle est considéré comme trop coûteux (y compris pour la faire respecter) et que l'on prend le risque de recourir au secret. Certains laboratoires interrogés ont reconnu avoir recours au secret tout en ayant en réserve une publication toute prête au cas où un concurrent approcherait de trop près les résultats attendus.

Il y a donc une prise en compte du fait que les DPI étaient accordés trop en amont avec un recours au secret et des publications prêtes comme une assurance contre les innovations des concurrents.

6. CONCLUSIONS

La controverse sur le blocage des recherches par un excès de propriété intellectuelle existe. Cette controverse trouve son origine dans le fait que la notion de blocage de la recherche est sans doute excessive par rapport à d'autres conséquences moins radicales mais tout aussi pénalisantes que sont les retards et/ou les modifications dans les programmes de recherche. Donc des freins existent du fait de la mise en œuvre d'une propriété intellectuelle sur le vivant dans les années 80 et 90. Ces freins sont aujourd'hui atténués par des réformes dans le système des brevets du fait de changements dans la réglementation.

Nous avons montré que les caractéristiques qui conduisent à ces difficultés d'accès et de diffusions sont de nature plus diverses que les DPI en tant que tels, elles relèvent pour certaines du contrat et des ATM : statut du matériel, propriétaire du matériel, responsabilité sanitaire, recours au secret tant sur des ressources génétiques que sur les informations les concernant, ... Or certains de ces contrats de type ATM n'existent qu'en réponse à la possibilité de déposer des brevets sur des séquences de gènes. Etant dans le domaine du contrat privé, il est alors plus difficile de revoir les conditions d'accès, elles reposent sur la bonne volonté de chacun, donc sur des avancées en termes d'information et de formation.

Nous avons également montré que la gestion collective de la propriété intellectuelle et de la recherche avaient pour origine de limiter les risques liés aux brevets. Dans le premier cas de limiter les effets pervers de brevets passés et dans le second d'éviter de pouvoir dépendre de brevets qui auraient été déposés par d'autres dans le futur. La gestion collective de la recherche a également pour objectif de limiter les risques de ne pouvoir accéder à certains moyens de la recherche même s'ils ne sont pas protégés par une propriété intellectuelle : collection de ressources génétiques ; collections de mutants, ...

Enfin dans des perspectives, il serait sans doute intéressant de faire un questionnaire simplifié au niveau de l'ensemble des laboratoires de recherche pour voir d'une part, s'ils acceptent ou non de céder des licences sur leurs brevets et quelles sont leurs motivations, et d'autre part, s'ils ont déjà essuyé des refus pour accéder à des licences de brevets.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Atkinson R.C. *et al.* (2003)- Public sector collaboration for agricultural Intellectual Property management- *Science*, vol. 301, pp. 174-175.
- [2] Bessen J.E. (2006).- Hold up and licensing in cumulative innovations with private information, MIT Sloan School of Management, 8 pages.
- [3] Bittner S. (2004).- Scientific collaboration between government and industry, *The journal of innovation, The public sector innovation Journal*, vol 9, n°2, 19 pages.
- [4] Cassier, M., Foray, D. (1999). "La régulation de la propriété intellectuelle dans les consortiums de recherche: les types de solutions élaborées par les chercheurs." *Economie Appliquée* LII(2): 155-182.
- [5] Choi J.P. (1993), Cooperative R&D with product market competition, *International Journal of Industrial Organization.*, 11 (1993) 553-571.
- [6] Cohen W. M. (2005).- Patents and appropriation : concerns and evidence. *Journal of technology transfer*, 30, pp. 57-71.
- [7] Cohendet P. *et al.*, (2006).- Entre incitation et coordination, repenser le rôle économique du brevet dans une économie fondée sur la connaissance. Colloque en route vers Lisbonne, Novembre, 32 Pages.

- [8] FAO (2001).- Traité international sur les ressources phytogénétiques pour l'agriculture et l'alimentation, FAO, Rome, Italie, 27 pages.
- [9] Gay B., Dousset B. (2006).- Innovation and network structural dynamics, study of the alliance network of a major sector of the biotechnology industry, *Research Policy*, 34, pp. 1457-74
- [10] Graff G., *et al.* (2001). "An intellectual property clearinghouse for agricultural biotechnology." *Nature Biotechnology* Vol.19 (December). 1179-1180 (<http://biotech.nature.com>).
- [11] Hagedoorn J. (2003).- Growth, patterns in R&D partnership : an exploratory statistical study, *International journal industrial organization*, 21, 517-31.
- [12] Hansen A.B., J. Asher, (2005) "Intellectual Property in the AAAS Scientific Community: A Descriptive Analysis of the Results of a Pilot Survey on the Effects of Patenting on Science" (2005).
- [13] Heller M., Eisenberg R. (1998), "Can Patents Deter Innovation? The Anticommons in biomedical Research", *Science* 280 (may), pp. 698-701.
- [14] Henry C., Trommetter, M., Tubiana, L. (2003). "Innovation et droits de propriété intellectuelle : quels enjeux pour les biotechnologies ?". In: Tirole, J. ; Henry C., Trommetter M., Tubiana L., Caillaud B., (éd.), Propriété intellectuelle. Rapport du "Conseil d'Analyse Economique". n°41. Paris: La Documentation Française. pp.49-112.
- [15] Joly P.B., Hervieu B. (2003). "La marchandisation du vivant - Pour une mutualisation des recherches en génomique." *Futuribles* 292: 5-29.
- [16] Joly P.B., Hermitte M.A., (1993). - «Plant biotechnology and patents in Europe : an economic analysis of alternative models of intellectual property rights» - *Biotechnology Review* , n° 1 , pp 76-94 ,
- [17] Kryder R.D., Kowalski S.P., Krattiger A.F. (2000). "The intellectual and technical property components of pro vitamin-A rice, (Golden Rice™) a preliminary freedom-to-operate review". ISAAA Brief N0. 20. Ithaca: ISAAA
- [18] Lerner J., Tirole J. (2004). "Efficient Patent Pools". *The American Economic Review*. Vol. 94. n°3. pp. 691-711.
- [19] Maurer, S. M. et Scotchmer, S. (1999). Database protection: is it broken and should we fix it? *Science*, 284, 1129-1130.
- [20] Maurer S.M. (2006).- Inside the anti-commons : academic scientist' struggle to build a commercially self-supporting human mutations database 1999-2001, *Research Policy*, 35, pp. 839-853.
- [21] Merz J.P., *et al.*, (2002) "Industry opposes genomic legislation" (2002) 20:7 *Nature Biotechnology* 657.
- [22] Organisation des Nations Unies, ONU (1992).- *Convention sur la diversité biologique*, Programme des Nations Unies pour l'environnement, PNUE, 85 pages.
- [23] Rodriguez *et al.*, (2007a) "Do material transfer agreements affect the choice of research agendas? The case of biotechnology in Belgium" 71:2 *Scientometrics* 239.
- [24] Rodriguez *et al.*, (2007b) "Material transfer agreements and collaborative publication activity: the case of a biotechnology network" 16:2 *Research Evaluation* 123
- [25] Roijakkers N., Hagedoorn J., (2006).- Inter firm R&D partnering in pharmaceutical biotechnology since 1975, trends, pattern and networks, *Research Policy*, 35, pp. 431-46.

- [26] Rosenkranz S., Schmitz P.W. (2003), Optimal allocation of ownership rights in dynamic R&D alliances. *Game and economic behavior*, 43, pp.153-173.
- [27] Scharff R.L. (2007). "A Common Tragedy: Condemnation and the Anticommons," *Natural Resources Journal*, vol.47, n°1.
- [28] Scotchmer S., (1991). "Standing on the shoulders of giants: cumulative research and the patent law." *Journal of economic perspectives*. 5(1). pp. 29-41.
- [29] Scotchmer S., (1999). "On the optimality of the patent renewal system". *Rand Journal of Economics* (30). pp181-196.
- [30] Scotchmer S., (2005), *Innovation and Incentives*, Cambridge: MIT press.
- [31] Shane S.A. et al. (2007).- The effects of patent litigation on university licensing efforts, *JEBO*, vol.63, pp.739-55.
- [32] Shapiro C. (2000). "Navigating the patent thicket : cross licenses, patent pools, and standard-setting". in A.B. Joffe and J. Lerner eds. *Innovation Policy and the economy*. MIT Press. Cambridge (Mass) : 119-150.
- [33] Stern S., Murray F., (2005) "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis" NBER Working Paper No. W11465
- [34] Teyssendier de la Serve B., Trommetter M., (2004).- *Protection et diffusion des résultats de génomique et biotechnologies végétales : quels enjeux pour la recherche publique ?* in J.F. Briat et J.F. Morot-Gaudry (eds), *La génomique végétale*, INRA éditions, Paris, 515-34
- [35] Tirole J., (2003). "Protection de la propriété intellectuelle : Une Introduction et Quelques Pistes de Réflexion" . In :Tirole, J. ; Henry, C. ; Trommetter, M. ; Tubiana, L. ; Caillaud B., (éd.) *Propriété intellectuelle*. Paris. Documentation Française. pp. 5-48.
- [36] Trilateral project B3b, 2000. Comparative study on biotechnology patent practices. Theme : Patentability of DNA fragments. June, www.jpo.go.jp/saikine/tws/sr-3.htm
- [37] Trommetter M., (2005). "Biodiversity and international stakes : a question of access ?" *Ecological Economics*. vol. 53. n° 4. pp. 573-583.
- [38] Trommetter M. (2006).- La propriété intellectuelle dans les biotechnologies agricoles : quels enjeux pour quelles perspectives ? *Reflets et perspectives de la vie économique*, pp. 37-48.
- [39] Trommetter M., (2007). - «Flexibility in the implementation of intellectual property rights in agricultural biotechnology». - Working Paper GAEL ; 2007-13., 33 pages.
- [40] Trommetter M., (2005), *Evolutions de la R&D dans les biotechnologies végétales et de la propriété intellectuelle*, in M. Frison Roche ed. «Propriété intellectuelle : visions d'économistes et de juristes », Paris: LGDJ-Montchrestien.
- [41] Union Européenne, UE (1996).- Directive du 11 mars 1996 sur la protection des bases de données, Journal officiel n° L 077 du 27/03/1996, Bruxelles, pp. 20 – 28.
- [42] Verbeure B. et al. (2006).- Patent pools and diagnostic testing, *Trends in biotechnology*, 24, 115-20.
- [43] Walsh J.P. et al., (2003) "Science and the Law. Working Through the Patent Problem" 299 *Science* 1021.

- [44] Walsh J.P. *et al.*, (2005a) “Science and Law. View from the Bench: Patents and Material Transfers” 309 *Science* 2002;
- [45] Walsh J.P., *et al.*, (2005b) “*Patents, material transfers and the access to research input in biomedical research*”, Final report to the National academy of Sciences, Washington, USA, 60 pages.
- [46] Weil A., (coordonateur) *et al.* (2004), *Proposition d’un centre public d’échange et de ressources en biotechnologie*, Rapport du groupe de travail aux Ministres de l’Agriculture et de la Recherche, Paris.
- [47] Williamson A.R., (2000). “Creating a structural genomics consortium”. *Nature structural genomics. Supplement*. pp.953.
- [48] WIPO, World Intellectual Property Organization, (2001). Matters concerning intellectual property and genetic resources, traditional knowledge and folklore – an overview. Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge And Folklore. First Session Geneva, April 30 to May 3, 2001.
- [49] Wouters P. (2002), Policies on Digital Research Data: An International Survey; In: Book series: Paul Wouters and Peter Schröder (eds.), *The Public Domain of Digital Research Data*, NIWI-KNAW, Amsterdam, ISBN 90 6472 183 1, 72 pp.
- [50] Wright B., Pardey (2006).- Changing intellectual property regimes: implications for developing country agriculture. *International Journal of Technology and Globalisation*, v.2, no.1-2, pp.93-114.

Analyse de séquences multi-loci (MLSA) dans le genre *Bradyrhizobium*

Angèle N'ZOUÉ^(1,2,3,4,5), Gisèle LAGUERRE^(1,2,3,4,5,6,7),
Farida BOULILA^(1,2,3,4,5,6,7,8), Chinnaswamy APPUNU^(6,7,8,9,10),
Nourredine SEBBANE^(1,2,3,4,5,8), Raoul RIVAS⁽¹¹⁾, Anne WILLEMS⁽¹¹⁾,
Lionel MOULIN^(1,2,3,4,5), Philippe de LAJUDIE^{(1,2,3,4,5)**}

⁽¹⁾IRD, UMR 113 Symbioses Tropicales et Méditerranéennes
34398 Montpellier, France

⁽²⁾INRA, UMR 113 Symbioses Tropicales et Méditerranéennes
34398 Montpellier, France

⁽³⁾CIRAD, UMR 113 Symbioses Tropicales et Méditerranéennes
34398 Montpellier, France

⁽⁴⁾SupAgro, UMR 113 Symbioses Tropicales et Méditerranéennes
34398 Montpellier, France

⁽⁵⁾Université de Montpellier II, UMR 113 Symbioses Tropicales et Méditerranéennes
34398 Montpellier, France

⁽⁶⁾INRA, UMR 1229 Microbiologie et Géochimie des Sols, 21065 Dijon, France

⁽⁷⁾Université de Bourgogne, UMR 1229 Microbiologie et Géochimie des Sols,
CMSE, BP 86510, 21065 Dijon, France

⁽⁸⁾Laboratoire de Microbiologie Appliquée,
Faculté des Sciences de la Nature et de la Vie, Université A. Mira, Béjaïa, Algérie

⁽⁹⁾Institute of Agricultural Sciences, Banaras Hindu University,
Microbial Genetics Laboratory, Varanasi-221005, Inde

⁽¹⁰⁾M. S. Swaminathan Research Foundation, Microbiology Division,
Chennai-600113, Inde

⁽¹¹⁾Laboratorium voor Microbiologie, Faculteit Wetenschappen, Universiteit Gent,
K.L. Ledeganckstraat 35, B-9000 Gent, Belgique

*Laboratoire des symbioses tropicales et méditerranéennes,
UMR 113 IRD/Agro-M/CIRAD/UM II, USC INRA 1242,
Campus International de Baillarguet,
TA A-82/J, 34398 Montpellier Cedex 5, France

Abstract: Multilocus sequence analysis in *Bradyrhizobium*. To clarify the taxonomic and phylogenetic relationships in the *Bradyrhizobium* genus, we performed Multi Locus Sequence Analysis (MLSA) on several bradyrhizobial strain collections. We first studied 2-3 representative strains for each of the 12 described *Bradyrhizobium* genospecies, including the named species *B. japonicum*, *B. elkanii*, *B. yuanmingense*, *B. canariense*, *B. liaoningense*, *B. betae*. We also studied a collection of 38 *Bradyrhizobium* spp. (*Aeschynomene*) strains from Senegal, representative for both their symbiotic and taxonomic

* Correspondance et tirés à part : P-De.Lajudie@mpl.ird.fr

diversity (three cross inoculation groups, photosynthesis and genomic species previously evidenced by polyphasic taxonomy). Nine housekeeping genes were sequenced, *recA* (recombinase A), *dnaK* (Hsp70 molecular chaperone, heat shock protein *dnaK*), *rpoB* (DNA-dependent RNA polymerase β subunit), *gyrB* (DNA gyrase β subunit), *atpD* (ATP synthase β subunit), *glnII* (glutamine synthase II), *glnA* (glutamine synthase I), *gltA* (citrate synthase I), *thrC* (threonine synthase), and one symbiotic gene, the nitrogenase structural gene *nifH*. *recA*, *atpD*, *gyrB*, *rpoB* and *dnaK* could be amplified for all strains. Housekeeping gene sequences were analysed both individually and after concatenation, using different phylogenetic reconstruction algorithms, Neighbour-joining (NJ), Maximum Parsimony (MP), Maximum likelihood (ML), and Bayesian model (MB). The obtained phylogenetic representations and groupings are essentially consistent between them and with those obtained in the past from classical taxonomic techniques, ribosomal operon analysis, AFLP, DNA:DNA hybridisations. *atpD* exhibits the most aberrant phylogeny while the other genes show similar groupings although detailed clusterings are not always the same. Results are also discussed in view of the different strain features: photosynthesis, inoculation groups. Incongruence length difference tests, *glnII*, and to a less extent, *dnaK* and *recA* are the three genes that, individually, are most congruent with the representation obtained from the nine housekeeping gene concatenated sequences. Phylogeny of *Bradyrhizobium* spp. (*Aeschynomene*) *nifH* marker suggest cases of symbiotic gene lateral transfers. Then we characterised two new strain collections isolated from root nodules of (1) *Retama retam* et *Retama sphaerocarpa* in Algeria and (2) soybean and *Vigna* spp. (*V. unguiculata*, *V. radiata*, et *V. mungo*) in India. The strains were first grouped by PCR-RFLP of 16S-23S ITS ADN; *nifH* and *nodC*. Representative strains were then chosen for sequencing ITS, *glnII*, *dnaK*, *recA*, *nifH* and *nodC*. In each case MLSA results (*glnII*, *dnaK*, *recA*) were consistent with ITS, but increased phylogenetic stability. New *Bradyrhizobium* groups were evidenced and confirmed among isolates from India and from Eastern Algeria. Phylogenetic analysis of symbiotic genes *nifH* and *nodC* suggested diverse mechanisms for their transmissions depending on strains and groups, either vertically or horizontally.

***Bradyrhizobium* / MLSA / phylogeny / taxonomy**

Résumé : Dans le but de clarifier la taxonomie de *Bradyrhizobium*, nous avons appliqué la technique MLSA aux 12 espèces génomiques décrites dans ce genre et à une collection de 38 souches de *Bradyrhizobium* spp. (*Aeschynomene*) du Sénégal, représentatives des trois groupes d'inoculations croisées et des différents groupes génomiques décrits précédemment par d'autres techniques. Neuf loci de gènes de ménage (*recA*, *atpD*, *glnA*, *glnII*, *gyrB*, *rpoB*, *gltA*, *thrC*, *dnaK*) ont été séquencés. Les séquences ont été analysées soit individuellement, soit concaténées, par différents algorithmes de reconstruction phylogénétique (NJ, MP, ML et modèle Bayésien). Les représentations obtenues sont globalement congruentes, entre elles et avec celles observées dans le passé par d'autres techniques. Les marqueurs *glnII*, *dnaK* et *recA* sont les plus congruents avec les neuf marqueurs concaténés. Cette approche a ensuite été appliquée à l'étude de deux nouvelles collections de bradyrhizobia isolées en Algérie de *Retama* spp. et en Inde de soja et *Vigna*. L'étude MLSA a permis de préciser les relations phylogénétiques observées par analyse de l'opéron ribosomi-

que et de confirmer certains groupes comme nouveaux. En parallèle, certains gènes symbiotiques, *nifH* et *nodC*, ont été séquencés, mettant en évidence leur mode de transmission soit vertical soit horizontal selon les cas.

***Bradyrhizobium* / MLSA / phylogénie / taxonomie**

1. INTRODUCTION

Les bactéries fixatrices d'azote symbiotiques des légumineuses, souvent collectivement nommées rhizobia, ne constituent en fait une entité ni taxonomique ni phylogénétique. Au contraire les rhizobia forment un grand groupe fonctionnel hétérogène, polyphylétique dont on découvre régulièrement de nouveaux membres. Ils se répartissent dans 12 genres et une cinquantaine d'espèces, phylogénétiquement intercalés parmi de nombreux autres genres et espèces bactériennes à l'intérieur des α - et des β -Protéobactéries (<http://www.ccg.unam.mx/rhizobial-taxonomy/>).

La taxonomie des rhizobia suit les critères polyphasiques de la taxonomie bactérienne générale, avec comme socle central les valeurs d'hybridations ADN/ADN et l'analyse phylogénétique du gène codant pour l'ARN ribosomique 16S. Ces deux critères, tout en restant très utiles pour l'identification, présentent cependant des limites qui s'avèrent parfois importantes. En particulier chez les rhizobia, certains groupes possèdent des plasmides de grande taille qui représentent jusqu'à la moitié de l'information génétique de la cellule ; par ailleurs, le génome des rhizobia a souvent été décrit comme plastique et sujet à l'acquisition de matériel génétique par transfert horizontal. Les résultats des hybridations ADN/ADN s'en trouvent biaisés. L'approche par séquençage du gène codant pour l'ARN ribosomique 16S, quant à elle, se heurte aux problèmes liés à sa faible variabilité et aux recombinaisons homologues [2], [11]. Pour pallier ces limites, plusieurs auteurs ont préconisé l'analyse des séquences de plusieurs gènes dits « de ménage », codant des protéines, universellement présents et en copie unique (MLSA, *Multi Locus Sequence Analysis*), pour établir une phylogénie plus robuste nécessaire à la taxonomie bactérienne [8].

Le genre *Bradyrhizobium* (α -Protéobactéries) comprend 5 espèces nommées *B. elkanii*, *B. liaoningense*, *B. yuanmingense*, *B. canariense* et *B. betae*. La taxonomie du genre *Bradyrhizobium* pose cependant certains problèmes et reste confuse. L'homogénéité même du genre *Bradyrhizobium* pose question du fait que les différentes espèces sont phylogénétiquement séparées par des bactéries d'autres genres comme *Azippia*, *Rhodopseudomonas palustris* et *Nitrobacter*. L'espèce type, *B. japonicum* reste hétérogène et contient au moins deux groupes d'hybridation ADN/ADN, I et Ia. Au cours des dernières

années de nombreuses souches de *Bradyrhizobium* sp. ont été décrites, caractérisées par différentes méthodes phénotypiques et génotypiques, sans conclusion nomenclaturale faute de consensus, alors que certains groupes constituent clairement des espèces séparées [15]. Les études MLSA de gènes de ménage décrites chez *Bradyrhizobium* ont permis une meilleure résolution au niveau de l'espèce et ont mis en lumière des recombinaisons homologues fréquentes mais transmises verticalement [9], [10], [12], [13], [14]. L'étude parallèle de gènes symbiotiques (*nifH*, *nodC*, *nodZ*, *noeI*, *noeL*) a montré au contraire des transmissions latérales entre branches chez plusieurs espèces de *Bradyrhizobium* et ainsi deux biovariétés ont été proposées chez *B. japonicum* [9], [10], [12].

Ces dernières années, nous avons décrit de nombreuses souches de *Bradyrhizobium* spp. symbiotiques de plusieurs légumineuses arborées et herbacées du Sénégal, utilisant diverses approches moléculaires [15], [16], [17]. Bien que les séquences de l'intergène ribosomique 16S-23S (ITS) soient très diverses en taille et en séquence chez *Bradyrhizobium*, le groupage selon l'ITS s'est avéré cohérent avec celui trouvé par AFLP (*Amplified Fragment Length Polymorphism*) et avec les valeurs d'hybridations ADN/ADN et retenu comme moyen simple de distinguer les espèces génomiques dans ce groupe, dont 8 nouvelles représentées par les souches du Sénégal [17]. Parmi celles-ci, les souches isolées de *Aeschynomene* spp. ont suscité un intérêt particulier ces dernières années du fait de leur valeur de modèles. Certaines d'entre elles sont photosynthétiques et produisent un pigment semblable à la bacteriochlorophylle a et des caroténoïdes et certaines sont des endophytes naturels de la tige de riz sauvage au Sénégal (pour une revue voir [3]). Toutes les souches photosynthétiques forment un groupe monophylétique sur la branche *Bradyrhizobium*, séparé de *B. japonicum* et de *B. elkanii* formant pour partie les deux espèces génomiques VI et VIII, et pour partie des branches à part quelle que soit la technique, AFLP, ITS, hybridations ADN/ADN [15], [16], [17]. En outre, les données de génomique ont montré que les *Bradyrhizobium* photosynthétiques ne possèdent pas les gènes de nodulation *nod* jusque là décrits comme universels chez tous les rhizobia [4].

Dans cette étude, nous avons étudié des représentants de chacune des espèces génomiques de *Bradyrhizobium*, nommées ou non [15], [16], [17], pour les soumettre à une analyse par MLSA de 8 gènes de ménage. Parallèlement nous avons étudié (en ajoutant les gènes *glnII* et *nifH*) 38 souches représentant la diversité de *Bradyrhizobium* spp. (*Aeschynomene*) du Sénégal. Nous avons enfin étudié deux collections de nouveaux isolats. L'une d'elles est constituée de 131 souches de bradyrhizobia isolées en Inde de nodosités de soja (*Glycine max*) et de trois espèces de *Vigna* traditionnellement cultivées dans ce pays, considéré comme un centre secondaire de domestication de ces espèces. La deuxième collection comprend 67 souches de bradyrhizobia

isolées en Algérie (région de Béjaïa) de nodosités de *Retama retam* et *Retama sphaerocarpa*, deux légumineuses arbustives endémiques du bassin méditerranéen adaptées aux conditions environnementales des zones arides et semi-arides. Dans un premier temps, ces collections ont été entièrement criblées par PCR-RFLP de l'ITS 16S-23S de l'ADNr, de l'ADNr 16S et de gènes symbiotiques, le gène de nodulation *nodC* et le gène de structure de la nitrogénase *nifH* afin d'évaluer leur diversité génétique. Des représentants des principaux génotypes ont ensuite été analysés par séquençage de l'ITS 16S-23S de l'ADNr et des gènes *nifH* et *nodC*, puis de 3 gènes de ménage sélectionnés sur la base des études précédentes.

2. MATÉRIEL ET MÉTHODES

2.1. Souches et milieux de culture

Toutes les souches ont été maintenues et cultivées sur milieu Yeast Mannitol à 37 °C, et conservées à -80 °C dans du milieu YM ajusté à 20 % glycérol.

2.2. Amplification

Neuf gènes de ménage ont été considérés : le gène *recA* (recombinase A), le gène *dnaK* (Hsp70 molecular chaperone, heat shock protein *DnaK*), le gène *rpoB* (codant pour la sous-unité β de l'ARN polymérase, DNA-dépendante), *gyrB* (codant pour la sous-unité β de l'ADN gyrase), *atpD* (codant pour la sous-unité β de l'ATP synthétase), *glnII* (glutamine synthetase type II), *glnA* (glutamine synthetase type I), *gltA* (citrate synthetase I), *thrC* (threonine synthetase). Les amorces et les conditions de l'amplification de ces gènes, de même que pour l'ITS 16S-23S de l'ADNr et les gènes *nifH* et *nodC*, ont été reprises de la littérature ou d'études précédentes [7], [12], [13], [14], [17].

2.3. RFLP

Les nouveaux isolats ont été d'abord caractérisés par PCR-RFLP de l'ITS de l'ADNr16S-23S [5] de l'ADNr16S [6] et des gènes symbiotiques *nifH* et *nodC* [7].

2.4. Séquençage

Les produits de PCR ont été soit purifiés par le QIAquick PCR purification kit soit d'abord séparés par électrophorèse sur gel d'agarose 1 % en tampon TAE (40 mM Tris-Acetate, 1 mM EDTA, pH 8,3). L'ADN a été

ensuite purifié à partir de la bande de gel extraite à l'aide du QIAquick gel extraction kit (Qiagen, Courtaboeuf, France). Le séquençage a été sous-traité auprès de la société MacroGen (Corée du Sud), qui utilise les protocoles et la chimie de séquençage d'Applied Biosystems, sur séquenceurs ABI3730.

2.5. Analyse des séquences

Les séquences ont été assemblées à l'aide du programme ClustalX et corrigées manuellement sous Genedoc. Elles seront soumises à la banque EMBL. La longueur des alignements générés est la suivante (pb) : *recA*, 523 ; *atpD*, 542 ; *glnA*, 383 ; *glnII*, 600 ; *gyrB*, 630 ; *rpoB*, 403 ; *gltA*, 497 ; *thrC*, 345 ; *dnaK*, 659 ; ITS 1058 ; *nifH* 283 ; *nodC* 487.

Les séquences des mêmes gènes ont été déduites des génomes entiers de *Rhodospseudomonas palustris* CGA009 et de *Sinorhizobium meliloti* 1021 pour servir d'« outgroup » pour enraciner les arbres. Les alignements ont été analysés par différentes méthodes de reconstruction phylogénétiques : distances Neighbour joining, maximum de vraisemblance (ML) et maximum de parsimonie (MP) sous le logiciel PAUP4, avec estimation des modèles de vraisemblance sous Modeltest. Une approche Bayésienne (sous MrBayes) a été également menée sur chaque marqueur MLSA ainsi que sur les données concaténées entre elles, en utilisant les paramètres ML obtenus par modeltest pour chaque marqueur. Des tests de congruence Shimodaira-Hasegawa (SH tests, sous PAUP4) ont été utilisés pour comparer les topologies des arbres phylogénétiques obtenus par les différentes méthodes.

3. RÉSULTATS

3.1. MLSA sur les espèces de référence de *Bradyrhizobium*

Une trentaine de souches de référence de *Bradyrhizobium* ont été choisies comme représentantes des espèces nommées (*B. japonicum* I et Ia, *B. elkanii*, *B. liaoningense*, *B. yuanmingense*, *B. betae*, *B. canariense*) et des différentes espèces génomiques précédemment mises en évidence parmi des souches du Sénégal [15], [16], [17]. Deux à trois souches de chaque espèce ont été considérées.

Nous avons choisi 8 gènes de ménage en référence aux précédentes études MLSA concernant les rhizobia. Certains gènes comme *gltA* et *thrC*, n'ayant pu être amplifiés chez toutes les souches malgré des essais récurrents utilisant différents jeux d'amorces, nous n'avons retenu pour l'analyse que les marqueurs *gyrB*, *atpD*, *rpoB*, *dnaK*, *glnA* et *recA*. L'arbre phylogénétique préliminaire de l'ensemble des séquences concaténées (résultats non présentés ici) montre une image cohérente avec la taxonomie en cours. L'analyse phylogénétique utilisant différents algorithmes est en cours. Elle

permettra d'évaluer les niveaux d'homologie inter- et intra-espèces à prendre en compte pour délimiter les espèces.

3.2. Collection de *Bradyrhizobium* spp. (*Aeschynomene*)

Neuf gènes de ménage ont été séquencés chez 38 souches de *Bradyrhizobium* spp. (*Aeschynomene*) du Sénégal, représentatives à la fois des trois groupes d'inoculations croisées et des différents groupes génomiques décrits précédemment [15], [16], [17]. Les séquences ont été utilisées pour effectuer des reconstructions phylogénétiques par ML, MP et modèle Bayésien soit individuellement soit après concaténation de l'ensemble des marqueurs.

Concernant les données concaténées, les arbres Bayésien et ML montrent des groupements sensiblement identiques aux différents niveaux de branches ; seul l'arbre Bayésien est présenté figure 1. Les souches se répartissent globalement en deux grandes branches. La première, A, hétérogène, rassemble toutes les souches photosynthétiques plus, de façon surprenante, la souche non photosynthétique ORS377. Ces souches sont le plus souvent rattachées à l'espèce génomique VI, mais d'autres espèces sont représentées comme l'espèce II (ORS287), l'espèce VIII (BTAi1), et d'autres espèces indéterminées. La deuxième branche, B, plus homogène, contient l'ensemble des autres souches non photosynthétiques et la souche de référence *B. japonicum* Ia USDA 110. Cette branche renferme des souches des espèces VII et Ia, plusieurs souches isolées et, curieusement, ORS336 (espèce VI). Les groupements observés ici à différents niveaux sont globalement cohérents avec ceux obtenus précédemment [15], [16], [17]. Cependant, il existe des exceptions, où des souches groupées par AFLP et ITS sont séparées en MLSA (ex. ORS331 et ORS336) ou groupées par AFLP mais séparées en ITS et en MLSA (ex. ORS327 et ORS375).

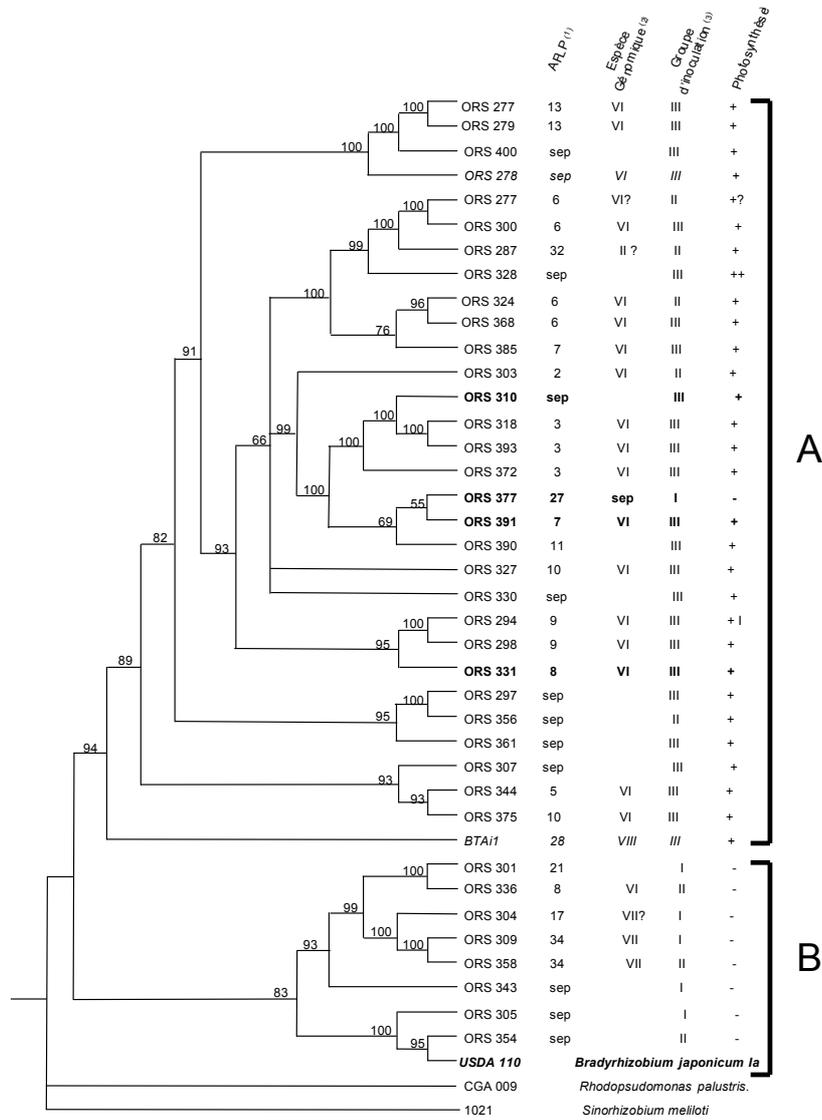


Figure 1 : Arbre Bayésien des données concaténées des séquences partielles de 9 gènes de ménage (*recA*, *atpD*, *glnA*, *glnII*, *gyrB*, *rpoB*, *gltA*, *thrC*, *dnaK*) des souches de *Bradyrhizobium* sp. (*Aeschynomene*). Burning of 10 000 first trees, 2.106 millions generations, 50 % Majority-rule consensus of 9950 trees. Les souches remarquables discutées dans le texte sont indiquées en caractères gras.

⁽¹⁾ Selon Willems *et al.* [16]

⁽²⁾ Selon Willems *et al.*, [15], [17].

⁽³⁾ Selon Alazard [1]

Mise à part ORS 377, toutes les souches du groupe d'inoculations croisées I de Alazard [1] appartiennent à la branche B. Une souche du groupe d'inoculation II appartient aussi à cette branche B ; les autres souches de ce groupe II se trouvent sur la branche A, ainsi que toutes les souches du groupe d'inoculation III.

Des tests de congruence Shimodaira-Hasegawa (SH tests, sous PAUP4) ont été effectués pour comparer les topologies des arbres phylogénétiques individuels pour chaque gène et après concaténation des 9 gènes. Avec trois des gènes étudiés, *glnII*, *dnaK* et *recA*, les arbres sont congruents avec celui obtenu après concaténation, le plus représentatif de l'ensemble étant *glnII*.

3.3. Étude phylogénétique des gènes *nifH* des souches d'*Aeschynomene* spp.

Les *Bradyrhizobium* photosynthétiques ne possédant pas les gènes *nod* classiques des rhizobia [4], nous avons limité notre étude des gènes symbiotiques à celle du gène *nifH* (fig. 2). Les souches se répartissent en quatre branches principales dont trois ont des séquences très voisines. La plupart des souches photosynthétiques appartiennent aux branches 1, 2 et 3. À l'inverse, la plupart des souches non photosynthétiques se groupent sur la branche 4, hétérogène avec 5 sous-branches. Les souches formant les branches « *nif* » 1, 2 et 3 correspondent globalement à la branche MLSA A et la branche « *nif* » 4 à la branche MLSA B. Cependant certaines souches font exception, comme les souches non photosynthétiques ORS301 et ORS343 et ORS377 dont les gènes *nifH* se groupent avec ceux des souches photosynthétiques (branches « *nif* » 1 et 2). À l'inverse, la souche photosynthétique présente une séquence *nifH* identique à la souche non photosynthétique ORS354 (branche 4).

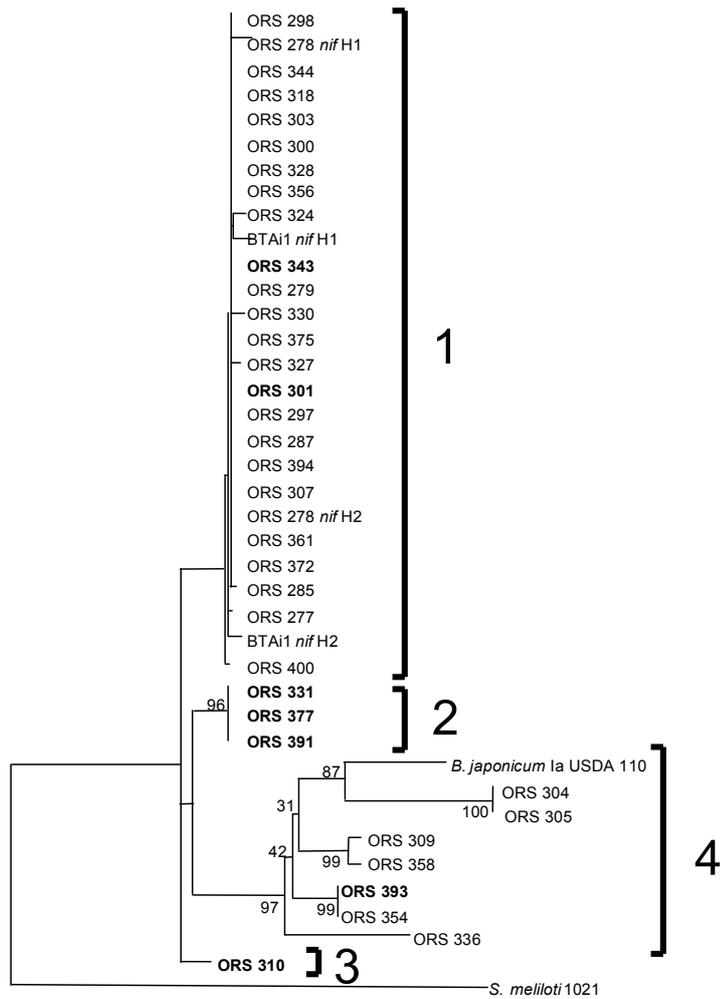


Figure 2 : Arbre phylogénétique construit par la méthode de neighbor-joining à partir des séquences du gène *nifH* des souches de *Bradyrhizobium* sp. (*Aeschynomene*). Les valeurs de bootstrap (1000 répliques) $\geq 80\%$ sont indiquées. Les souches remarquables discutées dans le texte sont indiquées en caractères gras.

3.4. Analyse de nouvelles collections

3.4.1. Collection de *bradyrhizobia* de *etama* spp. d'Algérie

3.4.1.1. Analyse de l'opéron ribosomique

Les 67 nouveaux isolats de nodules racinaires de *Retama* spp. forment un seul haplotype par PCR-RFLP de l'ADNr 16S (*Cfo*I et *Msp*I), distinct de ceux des espèces de référence de *Bradyrhizobium*, qui se divise en 9 haplotypes par PCR-RFLP de l'ITS (*Cfo*I et *Hae*III). L'arbre phylogénétique (NJ) construit à partir des séquences de l'ITS de 12 souches représentant chaque haplotype et des espèces de référence confirme la singularité des souches de *Retama* spp., qui forment un groupe monophylétique bien distinct subdivisé en 3 sous-groupes. Ces résultats suggèrent fortement que les bradyrhizobia de *Retama* spp. forment au moins une, voire trois, nouvelles espèces de *Bradyrhizobium*.

3.4.1.2. Analyse MLSA

Le séquençage des gènes *dnaK*, *glnII*, *recA* a été réalisé pour un représentant de chacun des trois sous-groupes délimités par l'analyse des séquences de l'ITS.

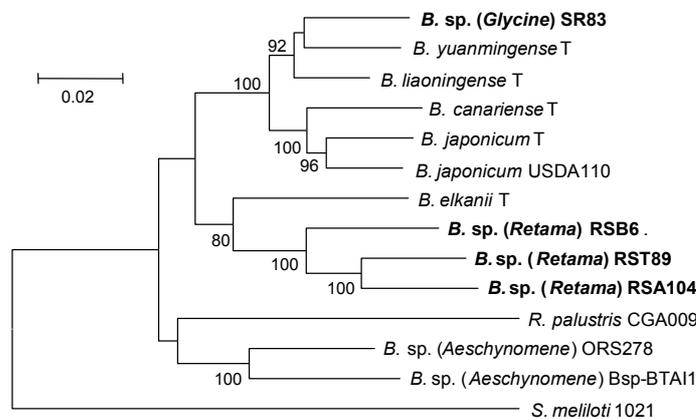


Figure 3 : Arbre phylogénétique construit par la méthode de neighbor-joining à partir des séquences partielles concaténées des gènes *dnaK*, *glnII* et *recA* des souches de *Bradyrhizobium* sp. de soja et de *Vigna* d'Inde et de *Retama* d'Algérie (un total de 1491 sites alignés). Les valeurs de bootstrap (1000 répliques) ≥ 80 % sont indiquées ; B., *Bradyrhizobium* ; R., *Rhodopseudomonas* ; S., *Sinorhizobium* ; T, souche type.

L'analyse phylogénétique de chacun des gènes, puis des trois gènes concaténés (fig. 3) confirme que les souches isolées de *Retama* forment un groupe distinct de ceux formés par les espèces décrites, qui se subdivise en

trois branches qui seraient phylogénétiquement plus proches de *B. elkanii* que des autres espèces (distance moyenne sur l'ensemble des séquences 0,119 ; distance avec *B. elkanii* 0,089 – 0,101 ; distance entre les séquences de *Retama* 0,050 – 0,072).

3.4.1.3. Analyse des gènes symbiotiques

Par PCR-RFLP (*Hae*III) les 67 souches forment un seul haplotype *nifH* et trois haplotypes *nodC* (*Hae*III et *Nde*II) dont un majoritaire (72 % des souches). La combinaison des haplotypes ITS, *nifH* et *nodC* a permis de classer les 67 souches dans 11 génotypes, dont un seul inclut 58 % des isolats.

Les 12 souches choisies pour le séquençage de l'ITS l'ont été aussi pour le séquençage des gènes *nifH* et *nodC*. L'analyse phylogénétique de ces deux gènes symbiotiques aboutit à la même conclusion que celle de la MLSA des gènes de ménage. Les isolats de *Retama* spp. forment à nouveau un groupe monophylétique bien distinct des séquences de référence de *Bradyrhizobium*.

3.4.2. Collections de souches de soja et *Vigna* spp. d'Inde

3.4.2.1. Analyse de l'opéron ribosomique

Nous avons enfin étudié 129 souches isolées de nodosités de légumineuses cultivées dans 5 agro-écosystèmes contrastés d'Inde, le soja (50 isolats), *Vigna unguiculata* (39 isolats), *V. radiata* (25 isolats) et *V. mungo* (15 isolats). Elles ont été classées dans deux haplotypes RFLP de l'ADNr 16S (7 enzymes de restriction), l'un identique à celui de *B. japonicum* et *B. liaoningense* et l'autre à *B. yuanmingense*, divisés respectivement en 3 et 5 haplotypes ITS (*Alu*I, *Cfo*I, et *Hae*III). L'un de ces derniers (ITS 1), associé au type ribosomique 16S *B. yuanmingense*, est majoritaire (65 % du total, 30 % des isolats de soja, 76 % des isolats de *V. radiata*, 97 % des isolats de *V. unguiculata*, 100 % des isolats de *V. mungo*). Les trois types ITS associés au type ribosomique 16S *B. japonicum*-*B. liaoningense* sont tous isolés du soja (64 %).

Un représentant de chaque haplotype ITS a été séquencé. L'arbre phylogénétique (NJ) montre que la plupart de ceux-ci ont des séquences identiques ou très proches de celles des espèces de référence. L'haplotype le plus fréquent (ITS 1) partage la séquence de *B. yuanmingense*. Tous les autres types ITS associés au génotype ribosomique 16S *B. yuanmingense* sont groupés avec *B. yuanmingense*. Deux des trois haplotypes ITS associés au type ribosomique 16S *B. japonicum*-*B. liaoningense* ont des séquences ITS identiques ou très similaires à la souche type de *B. liaoningense*. Le troisième haplotype ITS (ITS IV, souche SR83), groupant 38 % des isolats du soja, présente une

séquence distincte de celles des souches types et des représentants des espèces génomiques définies par Willems *et al.* [15], [17].

L'analyse phylogénétique des séquences ITS suggère donc que tous les isolats de *Vigna* spp. sont phylogénétiquement proches de *B. yuanmingense*, de même qu'environ 1/3 des isolats de soja. 26 % des isolats de soja sont rattachés à l'espèce *B. liaoningense*, mais 38 % n'ont pu être classés.

3.4.2.2. Analyse MLSA

Afin de confirmer la position phylogénétique du groupe ITS IV nous avons séquencé les gènes *dnaK*, *glnII*, et *recA* de la souche SR83. L'analyse individuelle de chaque marqueur donne des résultats congruents et aussi avec l'ensemble des trois marqueurs concaténés (fig. 3). SR83 forme un sous-groupe avec *B. yuanmingense* et *B. liaoningense*. Les distances génétiques observées suggèrent que ce groupe pourrait constituer une nouvelle espèce dans le genre *Bradyrhizobium*, mais davantage de données sont nécessaires pour estimer la distance moyenne intraspécifique.

3.4.2.3. Analyse des gènes symbiotiques

L'analyse par PCR-RFLP des deux gènes symbiotiques *nifH* et *nodC* a permis de définir 12 haplotypes *nifH*, 12 haplotypes *nodC*, constituant un total de 26 combinaisons d'haplotypes *nif-nod*, dont 25 sont associés aux isolats affiliés à *B. yuanmingense* sur la base des séquences ITS. Un génotype symbiotique unique, représentant 64 % des isolats, tous de soja, a été trouvé pour les isolats apparentés à *B. liaoningense* et au groupe ITS IV.

Le séquençage de plusieurs représentants de chaque haplotype *nifH* a été réalisé. L'analyse phylogénétique permet de définir deux groupes de séquences, celles apparentées à *B. yuanmingense* et celles apparentées à *B. japonicum* bv. *glycinearum* et *B. liaoningense*. Ces deux groupes correspondent bien aux deux haplotypes ribosomiques 16S et au regroupement obtenu par l'analyse phylogénétique de l'ITS, à l'exception de l'ITS IV. Pour ce dernier, la phylogénie des gènes symbiotiques le positionne avec *B. japonicum* bv. *glycinearum* et *B. liaoningense*.

4. CONCLUSION

Dans le genre *Bradyrhizobium*, les études précédentes avaient montré que le gène ADNr 16S est trop conservé et l'intergène ITS trop variable pour une phylogénie robuste. De plus il avait été montré que plusieurs souches (photosynthétiques) d'*Aeschynomene* possèdent plusieurs copies différentes de l'opéron ribosomique [17], gênant ainsi l'interprétation des analyses

phylogénétiques traditionnelles. Par exemple, différentes copies de l'ITS des souches ORS372 (AFLP 3) et ORS298 (AFLP 9) émergent sur plusieurs branches. Dans notre étude nous avons appliqué une approche MLSA en considérant 9 gènes de ménage aux espèces de référence et à différentes collections de ce genre, issues de différentes plantes et de différentes régions du monde : *Aeschynomene* spp. d'Afrique de l'Ouest, *Retama* spp. d'Algérie, *Vigna* spp. et soja d'Inde. Afin d'intégrer une dimension symbiotique dans la caractérisation, nous avons inclus l'analyse des gènes *nifH* et *nodC*.

Les gènes de ménage choisis se sont bien avérés présents en copie unique, ce qui facilite l'analyse. Pour les différentes collections étudiées, y compris les espèces de référence, la phylogénie obtenue avec les différents marqueurs, quel que soit l'algorithme de reconstruction, est globalement congruente avec celle issue de l'ensemble des marqueurs concaténés, avec certains présentant un niveau de congruence supérieur, comme *glnII*, *dnaK* et *recA*. L'image des relations phylogénétiques entre souches et espèces est globalement cohérente avec celle issue des techniques de taxonomie classique, en particulier l'analyse de l'ITS.

L'analyse des deux nouvelles collections de bradyrhizobia isolés en Algérie de *Retama* spp. et en Inde de soja et *Vigna* spp. par la même approche MLSA sur ces trois marqueurs a permis de vérifier les relations phylogénétiques observées par PCR-RFLP et séquençage de l'opéron ribosomique, et de confirmer certains groupes comme nouveaux.

En parallèle l'analyse des séquences des gènes symbiotiques *nifH* et *nodC* suggère leur transmission selon différents modes, soit verticale (de cellules mère à fille) soit horizontale (acquisition par transfert inter-espèces indépendamment du chromosome). Par exemple la diversité des gènes symbiotiques chez les isolats du soja d'Inde montre des gènes apparentés au niveau évolutif à *B. yuanmingense* (transmission verticale) et des gènes identiques à *B. japonicum* bv. *glycinearum* et *B. liaoningense* y compris pour une nouvelle espèce potentielle, suggérant donc une acquisition par transfert horizontal. Par contre, le nouveau groupe génomique de *Bradyrhizobium* mis en évidence parmi les souches de *Retama* spp. d'Algérie présente des gènes symbiotiques spécifiques (pas de transfert horizontal). Chez les souches de *Bradyrhizobium* spp. (*Aeschynomene*), la phylogénie du gène *nifH* suit globalement celle des gènes de ménage (transmission verticale), certains cas suggèrent des transmissions horizontales.

REMERCIEMENTS

Ce travail a été financé par le Bureau des Ressources Génétiques (projet 2005-45). Angèle N'Zoué a reçu une bourse de thèse du Département

Soutien-Formation de l'Institut de Recherche pour le Développement (IRD). Nous remercions Christine Le Roux et Odile Domergue pour leur aide et leurs conseils techniques, Adako Moudiongui pour les discussions constructives.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Alazard D., 1985. - Stem and root nodulation in *Aeschynomene* spp. Appl. Environ. Microbiol., 50: 732-734.
- [2] Eardly B.D., Nour S.M., van Berkum P., Selander R.K., 2005. - Rhizobial 16SrRNA and *dnaK* genes: mosaic and the uncertain phylogenetic placement of *Rhizobium galegae*. Appl. Environ. Microbiol., 27:1328-1335.
- [3] Giraud E., Fleischman D., 2004. - Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes. Photosynth. Res., 82: 115-130.
- [4] Giraud E., Moulin L., Vallenet D., et 31 autres, 2007. - Legume symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. Science, 316:1296-1297.
- [5] Laguerre G., Mavingui P., Allard M.R., Charnay M.P., Louvrier P., Mazurier S.I., Rigottier-Gois L., Amarger N., 1996. - Typing of rhizobia by PCR DNA fingerprinting and PCR-RFLP analysis of chromosomal and symbiotic gene regions: application to *Rhizobium leguminosarum* and its different biovars. Appl. Environ. Microbiol., 62: 2029-2036.
- [6] Laguerre G., Van Berkum P., Amarger N., Prévost D., 1997. - Genetic diversity of rhizobial symbionts isolated from legume species within the genera *Astragalus*, *Oxytropis* and *Onobrychis*. Appl. Environ. Microbiol., 63: 4748-4758.
- [7] Laguerre G., Nour S. M., Macheret V., Sanjuan J., Drouin P., Amarger N., 2001. - Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. Microbiology, 147: 981-993.
- [8] Stackebrandt E., Frederiksen W., Garrity G.M., Grimont P.A., Kampfer P., Maiden M.C., Nesme X., Rossello-Mora R., Swings J., Trupper H.G., Vauterin L., Ward A.C., Whitman W.B., 2002. - Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol., 52: 1043-1047.
- [9] Stepkowski T., Hughes C.E., Law I.J., Markiewicz L., Gurda D., Chlebicka A., Moulin L., 2007. - Diversification of lupine *Bradyrhizobium* strains : evidence from nodulation gene trees. Appl. Environ. Microbiol., 73:3254-3264.
- [10] Stepkowski T., Moulin L., Krzyzanska A., McInnes A., Law I.J., Howieson J., 2005. - European origin of *Bradyrhizobium* populations infecting lupins and *Serradella* in soils of Western Australia and South Africa. Appl. Environ. Microbiol., 71:7041-7052.
- [11] van Berkum P., Terefework Z., Paulin L., Suomalainen S., Lindström K., Eardly B.D., 2003. - Discordant phylogenies within the *rrn* loci of rhizobia. J. Bacteriol., 185: 2988-2998.
- [12] Vinuesa P., Leon-Barrios M., Silva C., Willems A., Jarabo-Lorenzo A., Pérez-Galdona R., Martínez-Romero E., 2005a. - *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (Papil-

- ionoideae : Genisteae) growing in the canary Islands, along with *B. japonicum* bv. *Genistearum*, *Bradyrhizobium* genospecies α and *Bradyrhizobium* genospecies β . Int. J. Syst. Evol. Microbiol., 55: 569-575.
- [13] Vinuesa P., Silva C., Lorite M.J., Izaguirre-Mayoral M.L., Bedmar J., Martinez-Romero, E., 2005b. - Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from *rrs*, *atpD*, *recA* and *nifH* sequences, and their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands. Syst. Appl. Microbiol., 28: 702-716.
- [14] Vinuesa P., Silva C., Werner D., Martinez-Romer E., 2005c. - Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. Mol. Phylogenet. Evol., 34: 29-54.
- [15] Willems A., Doignon-Bourcier F., Goris J., Coopman R., de Lajudie P., Gillis M., 2001. - DNA-DNA hybridization study of *Bradyrhizobium* strains. Int. J. Syst. Evol. Microbiol., 51: 1315-1322.
- [16] Willems A., Doignon-Bourcier F., Gillis M., de Lajudie P., 2001. - Evaluation of AFLP for the grouping of *Bradyrhizobium* strains. Genet. Sel. Evol., 33 (suppl.1): 365-377.
- [17] Willems A., Munive A., De Lajudie P., Gillis M., 2003. - In most *Bradyrhizobium* groups sequence comparison of 16S-23S rDNA Internal Transcribed Spacer regions corroborates DNA-DNA hybridizations. System. Appl. Microbiol., 26: 203-210.

Caractérisation de populations à partir de l'analyse de régions chromosomiques potentiellement sélectionnées en comparaison à la diversité neutre : étude conjointe de lignées expérimentales de poules et des quatre sous- populations ovines de race Lacaune

Isabelle PALHIÈRE^{(1)*}, Valérie LOYWYCK⁽²⁾, Magali SANCRISTOBAL⁽³⁾,
Denis LALOÉ⁽⁴⁾, Yves AMIGUES⁽⁵⁾, Bertrand BED'HOM⁽²⁾,
Bernard BIBÉ⁽¹⁾, Piter BIJMA⁽⁶⁾, Loys BODIN⁽¹⁾, Daïa BOUDARENE⁽³⁾,
Claude CHEVALET⁽³⁾, Sophie LEROUX⁽³⁾, Katayoun MOAZAMI-
GOUDARZI⁽⁷⁾, Philippe MULSANT⁽³⁾, Frédérique PITEL⁽³⁾,
Bertrand SERVIN⁽³⁾, Etienne VERRIER⁽²⁾, Florence VIGNOLES⁽³⁾

⁽¹⁾INRA, Station d'Amélioration Génétique des Animaux,
31326 Castanet-Tolosan cedex, France

⁽²⁾UMR INRA-AgroParisTech Génétique et Diversité Animale,
78352 Jouy-en-Josas cedex, France

⁽³⁾INRA, UMR 444 Laboratoire de Génétique Cellulaire,
31326 Castanet-Tolosan cedex, France

⁽⁴⁾INRA Station de Génétique Quantitative Appliquée,
78352 Jouy-en-Josas cedex, France

⁽⁵⁾LABOGENA, 78352 Jouy-en-Josas cedex, France

⁽⁶⁾Animal Breeding and Genomics Centre, Wageningen University,
6700 AH Wageningen, Pays-Bas

⁽⁷⁾INRA Laboratoire Génétique Biochimique et Cytogénétique,
78352 Jouy-en-Josas cedex, France

Abstract: Characterization of populations by detecting genomic regions supposed to be under selection, compared to the neutral diversity: example of chicken experimental lines and the four breeding populations in Lacaune sheep breed. The management of animal populations under selection can be improved by the molecular characterization of their genetic diversity, in particular to identify regions affected by selection. The purpose of this study was to analyse the genetic diversity of populations initially sharing a common genetic background but selected both independently and intensively for more than 10 generations. Two selection designs were considered: (i) three experimental lines of chicken selected on different immune traits, and a control line, deriving from a single founder population;

* Correspondance et tirés à part : isabelle.palhiere@toulouse.inra.fr

(ii) four independently selected lines of a same breeding population, the Lacaune sheep breed: two of the lines were selected on dairy traits and the two others on prolificacy and muscular development. The adaptative genetic diversity was characterized by analysing and comparing the polymorphism evolution of 50 supposedly selected and 20 supposedly neutral microsatellite markers. The supposedly selected markers were located in 9 regions of the genome that were supposed to be under selection (8 QTL regions and 1 gene). For both designs, about 50 animals from the last generation were sampled and genotyped within each line. In the chicken design, founders were genotyped as well. Within each species, the population structure is well characterized by genotypes at the neutral markers. The comparison of allele frequencies at neutral markers and at markers located in QTL regions does not seem to indicate any significant differences between these regions with respect to the population structure. Using simulations of population evolving under drift alone, we tested for each marker the departure from this evolutionary scenario. Surprisingly, only two markers in chicken, and one marker in the dairy QTL region in sheep showed polymorphism evolution that could not only be explained by drift and let us assume evidences of selection signatures. From these results, we suggest possible improvements for the design of studies aimed at detecting signatures of selection: (i) to have a large density of markers in the supposedly selected regions; (ii) to study samples that have been separated since a limited number of generations; (iii) to check the validity of the supposedly selected regions. From our results, it seems that the adaptative genetic diversity is more difficult to characterise than the neutral genetic diversity. To improve this characterization, the development of new methods for the molecular data analysis is likely necessary.

diversity/ signature of selection/ microsatellite markers/ sheep/ chicken

Résumé – Cette étude vise à analyser la diversité génétique de populations ayant un important fonds génétique commun et sélectionnées indépendamment depuis plus de 10 générations. Elle s'appuie sur deux dispositifs : (i) trois lignées expérimentales de poules issues des mêmes fondateurs, sélectionnées sur des caractères différents de réponse immunitaire ; (ii) quatre populations ovines de race Lacaune, sélectionnées sur des caractères laitiers pour les deux unités de sélection laitière et sur des caractères de prolificité ou de développement musculaire pour les populations allaitantes. La diversité génétique fonctionnelle a été caractérisée par l'analyse du polymorphisme à 50 marqueurs microsatellites situés dans 9 régions chromosomiques, en comparaison à la diversité neutre (20 microsatellites). 50 animaux de la dernière génération ont été échantillonnés pour chaque population ainsi que les fondateurs des lignées de poules. Des analyses descriptives et les distances de Reynolds ont permis de montrer que les populations sont clairement différenciées sur la base des marqueurs neutres. L'analyse de la diversité génétique fonctionnelle, fondée sur des analyses multidimensionnelles ou sur des tests d'écart à la dérive génétique, a permis de mettre en évidence de faibles différences de comportement entre marqueurs. Seuls trois marqueurs ont une variabilité incompatible avec une simple dérive génétique.

diversité/ signature de sélection/ microsatellites/ ovin/ poule

1. INTRODUCTION

L'objectif de l'étude est d'analyser la diversité génétique et de détecter des signatures de sélection au sein de populations différentes mais ayant un important fonds génétique commun. Les investigations ont porté sur deux cas concrets dans deux espèces différentes, la poule et le mouton.

L'essor de la génétique moléculaire a permis une meilleure connaissance du génome des animaux, notamment par le développement de marqueurs génétiques. Ces marqueurs génétiques, répartis sur tout le génome, permettent de caractériser la diversité génétique neutre des populations animales. Des premières études de la diversité de populations de poules [20] et d'ovins [3], [7] ont pu ainsi être menées à partir de microsatellites. Par ailleurs, ces mêmes marqueurs ont également contribué à repérer des régions chromosomiques influençant fortement les performances des animaux. Ainsi, des gènes ou des QTL ont été identifiés récemment, en poules comme en ovins, concernant des caractères de production ou fonctionnels [1], [4]. A côté de la diversité génétique neutre, il est donc possible de caractériser la diversité fonctionnelle des populations en étudiant leur variabilité au niveau de ces zones du génome soumises à sélection. Celle-ci est d'un intérêt majeur pour différencier les populations ayant un important fonds génétique commun, qui sont très semblables sur une grande partie du génome (apprécié par la diversité neutre) mais peuvent être différentes sur une petite région chromosomique [10], [16].

La présente étude vise à caractériser la diversité de populations génétiquement proches à partir de l'analyse de signatures de sélection sur quelques régions chromosomiques supposées soumises à sélection, en comparaison de la diversité issue de marqueurs microsatellites neutres.

2. MATÉRIEL ET MÉTHODES

2.1. Animaux

Les dispositifs d'étude sont similaires dans les deux espèces. En premier lieu, il s'agit de populations ou lignées sélectionnées de façon indépendante depuis une dizaine de générations, à partir d'une population ancestrale. Chez la poule, trois lignées ont été créées en 1994 à partir du croisement de deux lignées commerciales « White Leghorn ». 42 fondateurs (G-2) ont été croisés au hasard pour constituer une population F2, représentant la population de base avant sélection (G0). Chaque lignée correspond à un protocole de sélection individuelle sur un caractère de réponse immunitaire différent, appliqué depuis 11 générations (une génération par an). Le progrès

génétique réalisé n'est pas le même selon le caractère considéré [13] : il est fort dans la lignée 1 (0,48 écart-type génétique/génération) et plus modéré dans les deux autres lignées (0,07 et 0,02 écart-type génétique/génération, respectivement, dans les lignées 2 et 3). Chaque année, 15 mâles et 30 femelles, choisis parmi les 200 animaux de chaque lignée, sont accouplés au hasard pour créer la génération suivante. Toutes les généalogies sont disponibles depuis la population de base jusqu'à la dernière génération étudiée.

En ovins, quatre populations d'une même race, la race Lacaune, sont sélectionnées en ferme depuis une quarantaine d'années (intervalle de génération compris entre 4 et 5 ans). La race est divisée en deux rameaux, un laitier et un allaitant, chaque rameau faisant l'objet de deux programmes de sélection reposant sur des éleveurs et appliqués à des ensembles d'animaux différents. Les deux programmes laitiers, conduits par la Confédération Générale de Roquefort (conf) et OVITEST (ovitL) respectivement, sont issus d'une même population d'origine qui s'est divisée depuis 1972. Ils s'appuient sur des bases de sélection de taille identique (environ 85000 femelles), portent sur les mêmes caractères et réalisent des progrès génétiques importants et similaires : environ 0,2 écart-type génétique/an [5]. Les deux programmes allaitants sont fondés sur des bases de sélection plus réduites (de l'ordre de 15000 femelles) et ont des origines moins clairement établies. Officiellement, ces programmes ont été mis en place au début des années 80, et appliquent une sélection efficace sur des caractères prioritaires différents, à savoir la prolificité et les aptitudes bouchères, pour les programmes conduits par OVITEST (ovitV) et par GEBRO respectivement.

D'autre part, il existe pour chaque espèce une population témoin pas ou peu sélectionnée sur les caractères étudiés. En poules, il s'agit d'une lignée témoin au sens strict, constituée à partir des mêmes fondateurs que les lignées sélectionnées et conduite dans les mêmes conditions. En ovins, on considère que la race Blanc du Massif Central (BMC) fait office de population témoin car, d'une part, elle est très proche sur les plans géographique, phénotypique et génétique de la race Lacaune et, d'autre part, elle est non sélectionnée sur les caractères laitiers et fait l'objet d'une pression de sélection modérée sur les caractères allaitants.

En ce qui concerne les animaux échantillonnés afin d'être génotypés, l'objectif était de disposer de 50 animaux par population, choisis parmi les individus nés lors de la dernière génération. L'échantillonnage a été réalisé au hasard parmi les 200 animaux de la onzième génération pour les poules, et en ovins, parmi les jeunes mâles élites candidats à la sélection de façon à être représentatif des origines de la population active. De plus, pour les

poules, tous les fondateurs (G-2) et 55 individus de la population F1 (G-1) ont également été retenus.

2.2. Marqueurs moléculaires

Les marqueurs microsatellites ont été génotypés à LABOGENA et sur la plate-forme génomique de la génopole "Toulouse Midi-Pyrénées" à l'aide d'un séquenceur d'ADN 3100 ABI PRISM® (Applied Biosystems, Foster City, CA, USA). Les génotypes ont été déterminés en utilisant les logiciels GeneScan Analysis 3.7 et Genotyper Analysis 3.7 (Applied Biosystems, Foster City, CA, USA). La base GEMMA a été utilisée pour gérer les typages ovins et les tests d'informativité chez la poule [8].

Une vingtaine de marqueurs supposés neutres, répartis sur le génome, ont été choisis dans le panel AVIANDIV pour les poules (21 microsatellites) et le panel FAO-ISAG pour les ovins (22 microsatellites). Par ailleurs, 50 marqueurs localisés dans 9 régions supposées soumises à sélection ont également été génotypés au total pour les deux espèces. 8 régions correspondent à des QTL issus de protocoles de primo-détection et présentent une densité de marqueurs assez faible avec, en moyenne, un marqueur tous les 7 cM :

- 7 QTL concernent des caractères de réponse immunitaire chez les poules, détectés dans des lignées expérimentales néerlandaises [22], [23], [24]. Leur taille varie de 17 à 43 cM sauf pour le QTL situé sur le chromosome 16 (microchromosome). En effet, la position génétique sur la carte consensus (2006) est estimée à 0 cM pour les 3 marqueurs situés dans cette zone QTL. Il s'agit de la région du CMH (Complexe Majeur d'Histocompatibilité) qui présente un niveau de polymorphisme extrêmement élevé ;
- 1 QTL, localisé grâce à un dispositif de familles en ferme de race Lacaune, concerne la quantité de lait qui est un caractère sélectionné depuis les années 50 dans cette race [4]. Les marqueurs sont répartis sur une zone de 65 cM ce qui couvre une région plus large que le QTL détecté (45 cM).

Une autre région, la zone FeCL, est également étudiée. Elle contient un gène majeur d'hyper-ovulation détecté dans l'unité de sélection OVITEST viande [6], [12]. A l'inverse des autres régions retenues, elle se caractérise par une densité en marqueurs importante : 6 marqueurs espacés de 0,5 cM et 1 marqueur à 14 cM ; et un effet fort : +1 ovule pour les femelles porteuses de la mutation à l'état hétérozygote par rapport aux femelles homozygotes non porteuses.

2.3. Analyse de la diversité neutre intra et inter population

Afin d'avoir un premier aperçu de la diversité génétique neutre intra et inter populations, pour une espèce donnée, des analyses descriptives ont été effectuées. Une Analyse Factorielle des Correspondances (AFC) a été appliquée sur les données de poules et un diagramme « Heatmap » a été obtenu à partir de la matrice des indices de similarités de Nei [17] : $\sum f_{Xi} f_{Yi}$, f_{Xi} étant la fréquence de l'allèle i pour la population X et f_{Yi} étant la fréquence de l'allèle i pour la population Y . De plus, pour caractériser la différenciation génétique entre les populations, les distances de Reynolds ont été estimées pour chaque locus, comme décrit par Laval *et al.*, [9]. Un arbre a été construit à partir de ces distances par la méthode du « Neighbor-Joining ». Un ré-échantillonnage de type bootstrap a été effectué afin d'évaluer la robustesse de la topologie de l'arbre (1000 répétitions).

2.4. Analyse multidimensionnelle à partir des marqueurs neutres et des marqueurs sur les zones sélectionnées

Quand les analyses portent sur des populations proches, les marqueurs inclus dans l'analyse peuvent ne pas être congruents : les distances génétiques entre populations calculées marqueur par marqueur peuvent être négatives et donc produire des structures contradictoires entre populations. Dans ce cas, une structure consensuelle n'a pas de sens. Nous avons donc testé la congruence entre marqueurs selon la procédure développée par Moazami-Goudarzi et Laloé [15]. Pour chaque locus, la matrice de distances entre populations est calculée; les corrélations entre les matrices de distances calculées pour chaque locus sont obtenues selon la procédure de Mantel [14].

Une ACP sur la matrice des corrélations entre matrice de distances est réalisée. Le cercle des corrélations qui en résulte et qui représente les relations entre marqueurs permet de juger visuellement de la congruence entre marqueurs. Plus les marqueurs sont proches, plus les structures correspondantes sont proches. Ainsi, si les corrélations entre distances sont positives, les marqueurs seront tous situés sur un demi-cercle.

2.5. Test d'écart à la dérive

Tester l'écart à la dérive génétique, posée comme hypothèse nulle (H_0), requiert la connaissance de la distribution sous H_0 d'une statistique de test choisie au préalable. Cette statistique sera typiquement une distance génétique entre populations ou une variation de fréquences alléliques entre générations. Il n'existe pas de résultat théorique donnant la distribution sous dérive de telles statistiques. Une distribution empirique a donc été obtenue à l'aide de simulations.

Dans le cas de populations ayant divergé indépendamment, à effectifs constants, depuis une population fondatrice entièrement génotypée (cas des poules), la statistique de test utilisée correspond à l'estimateur f_c proposé par Nei et Tajima [18], c'est-à-dire la variation temporelle des fréquences alléliques. Pour chaque marqueur, 2000 itérations ont été faites pour simuler l'évolution sous dérive pure.

Pour des populations échantillonnées une seule fois dans le temps (population fondatrice inconnue) et ayant une histoire évolutive ramifiée (cas des ovins), nous avons procédé en deux temps. Tout d'abord, la topologie de cette histoire évolutive a été inférée à partir de l'arbre « Neighbor-Joining » fondé sur des marqueurs supposés neutres, ainsi que sur une connaissance historique de la structure des populations (connaissance imparfaite, mais existante). Dans un second temps, conditionnellement à l'histoire évolutive inférée des populations, des simulations ont été menées permettant l'obtention de distributions empiriques de fréquences alléliques, donc de distances génétiques de Reynolds entre paires de populations notamment.

Dans les 2 types de schémas évolutifs, des « P-values » empiriques ont été calculées permettant de tester l'écart à l'hypothèse nulle de dérive génétique, pour chacun des marqueurs.

3. RÉSULTATS

3.1. Différenciation génétique des populations étudiées, à partir des marqueurs neutres

Les figures 1 et 2 illustrent la différenciation des populations étudiées, chez la poule et les ovins respectivement, sur la base des génotypes aux marqueurs neutres. Par l'AFC (fig. 1), il apparaît que les fondateurs (G-2), les animaux des populations F1 (G-1), issues du croisement des fondateurs, ainsi que les animaux de la lignée témoin sont superposés au centre de la figure. Du fait du croisement, la population G-1 est plus homogène que la population des fondateurs (G-2). Les individus des trois lignées sélectionnées sont, quant à eux, assez distincts et répartis selon les deux axes, ce qui reflète la divergence après 11 générations. Du fait de la sélection, les populations de G11 sont plus homogènes que celles de G-1 et G-2. La matrice des similarités (fig. 2) indique que les cinq populations ovines se distinguent clairement. Il existe une forte ressemblance des animaux intra mais aussi entre rameaux laitiers, imputable à leur origine commune et leur séparation récente. Par ailleurs, la population GEBRO est assez éloignée des autres populations de race Lacaune comme de la population témoin. Celle-ci semble disposer, d'une part, d'une plus grande

variabilité intra-population et, d'autre part, d'une ressemblance assez forte avec les populations de race Lacaune.

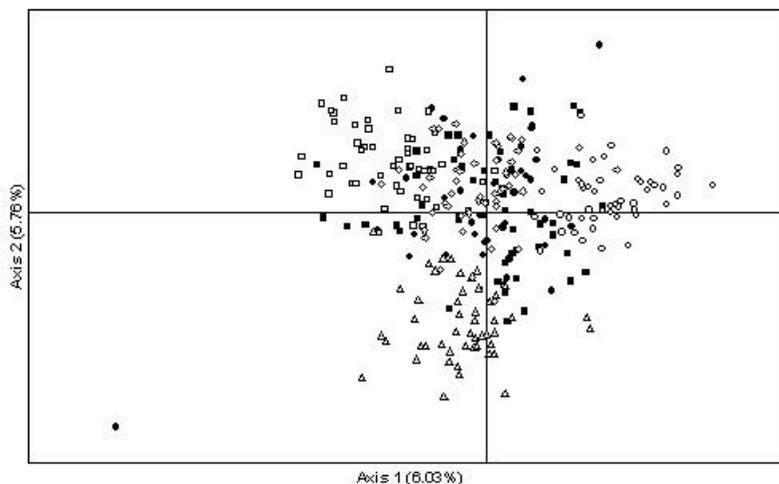


Figure 1 : AFC des individus de G-2, G-1 et G11 à partir des génotypes aux marqueurs supposés neutres (poules). Les points noirs correspondent aux animaux de G-2, les carrés noirs de G-1 et les figures blanches de G11 : les cercles correspondent à la lignée 1, les carrés à la lignée 2, les triangles à la lignée 3 et les losanges à la lignée témoin.

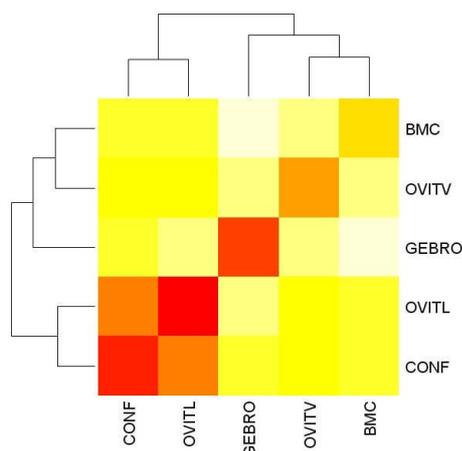


Figure 2 : Matrice des similarités entre les populations ovines sur les 22 locus supposés neutres. Sur la diagonale, la valeur correspond à l'homozygotie attendue de la population. Une couleur foncée indique une forte similarité entre les populations. Les noms des populations sont abrégés de la façon suivante : CONF : Confédération Générale de Roquefort ; OVITL : OVITEST programme laitier ; GEBRO : GEBRO programme allaitant ; OVITV : OVITEST programme allaitant BMC : Race Blanc du Massif central.

3.2. Distances de Reynolds sur les marqueurs neutres

Les valeurs des distances moyennes estimées avec les marqueurs neutres (21 chez la poule et 22 en ovins) sont présentées dans le tableau I. Comme attendu pour des populations génétiquement proches, les distances obtenues sont faibles. Elles se situent dans l'intervalle (0,02-0,06) entre populations ovines, c'est-à-dire 5 à 10 fois inférieures à celles obtenues dans une étude de comparaison entre races ovines espagnoles [2]. Les distances entre populations de poules, comprises entre 0,07 et 0,15, sont similaires à celles obtenues par Laval *et al.*, [10] entre deux générations de lignées divergentes de poules séparées par 12 années de sélection.

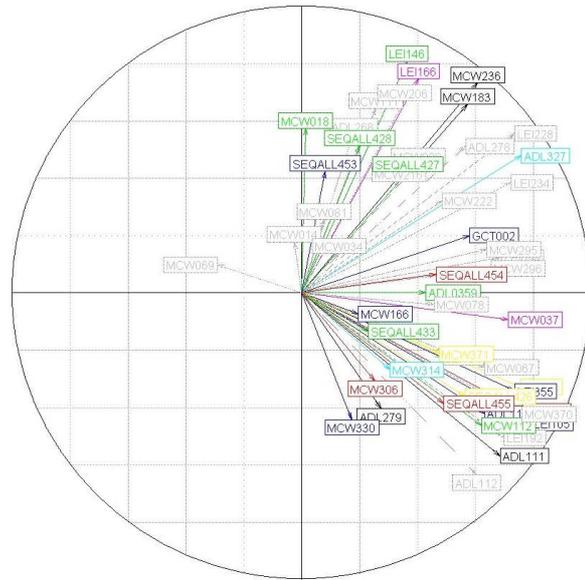
Tableau I : Matrice de distances génétiques de Reynolds à partir des fréquences alléliques des marqueurs supposés neutres, pour les poules (au-dessus de la diagonale) et pour les ovins (en-dessous de la diagonale).

	(1)	(2)	(3)	(4)	(5)	(6)
Fondateurs (1)		0,07	0,10	0,07	0,07	
Témoin / BMC (2)			0,11	0,12	0,07	
Lignée 1 / Confédération (3)		0,03		0,14	0,15	
Lignée 2 / GEBRO (4)		0,04	0,05		0,12	
Lignée 3 / OVITEST lait (5)		0,03	0,02	0,06		
OVITEST viande (6)		0,02	0,03	0,04	0,03	

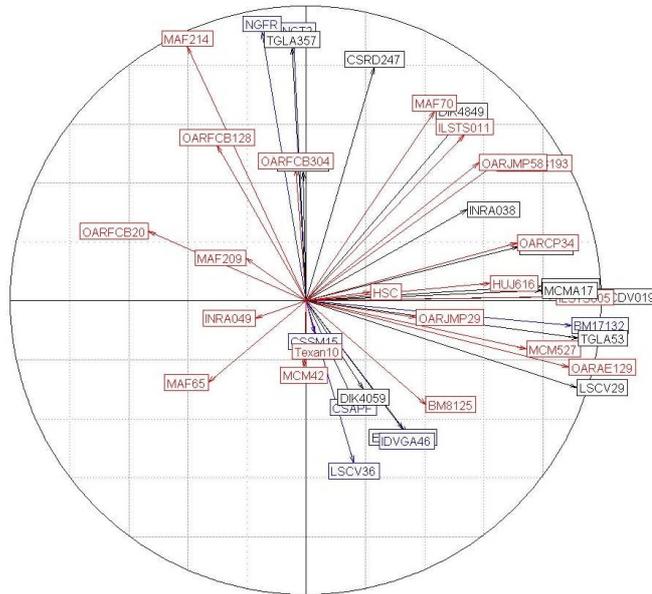
Des arbres ont été construits dans chaque espèce à partir de ces distances mais ne sont pas présentés ici, aucun nœud n'étant significatif pour les populations de poules (schéma en étoile) et seul le nœud de séparation des populations laitières et allaitantes étant significatif en ovins (bootstrap 99%). Chez les ovins, l'observation précédente, couplée à nos connaissances préalables de l'histoire des populations, a conduit à l'inférence de la topologie évolutive suivante : séparation entre races BMC et Lacaune, puis séparation des rameaux lait et viande, enfin séparation entre OVITEST viande et GEBRO d'une part, et OVITEST lait et Confédération d'autre part.

3.3. Comparaison de la structuration des populations selon l'information des marqueurs neutres et des marqueurs sur les zones sélectionnées

La structure des lignées de poules apparaît assez consensuelle, les marqueurs étant dans leur très grande majorité regroupés dans une moitié du cercle (fig. 3 (a)). Par contre, on n'observe pas de regroupement selon la nature des marqueurs (neutres ou situés sur des zones sélectionnées).



a)



b)

Figure 3 : Cercle des corrélations entre les marqueurs à partir des distances génétiques de Reynolds, pour les différents types de marqueurs, (a) pour les lignées de poules, (b) pour les populations ovines. Les marqueurs neutres et les marqueurs de chaque région chromosomique étudiée sont représentés par des couleurs différentes : les marqueurs neutres sont en rouge pour les ovins et en gris pour les poules.

Concernant les ovins, la congruence entre marqueurs, telle qu'elle ressort du cercle des corrélations, est assez lâche (figure 3 (b)). On observe une dispersion importante des marqueurs dans le cercle, même si un test sur la somme des coefficients de corrélation montre que celle-ci est significativement positive, et indique donc l'existence d'un consensus significatif. L'examen des structures de chaque population indique que les populations Confédération et OVITEST Lait sont très proches l'une de l'autre, et que la population GEBRO est isolée (résultat non présenté). Enfin, comme pour le cas des poules, on n'observe pas de regroupement selon la nature des marqueurs.

3.4. Identification de régions chromosomiques ayant répondu à la sélection

Les P-values des tests d'écart à la dérive, basés sur les distances génétiques entre populations ovines, sont présentées dans la figure 4, pour chacun des marqueurs. Après correction pour la multiplicité des tests (Bonferroni), un seul marqueur est significatif : INRA038 se trouvant dans la zone du QTL laitier. On peut cependant noter une accumulation de valeurs relativement élevées de signification ($-\log_{10}(\text{P-values})$ sur la figure 4) dans la zone d'hyperovulation.

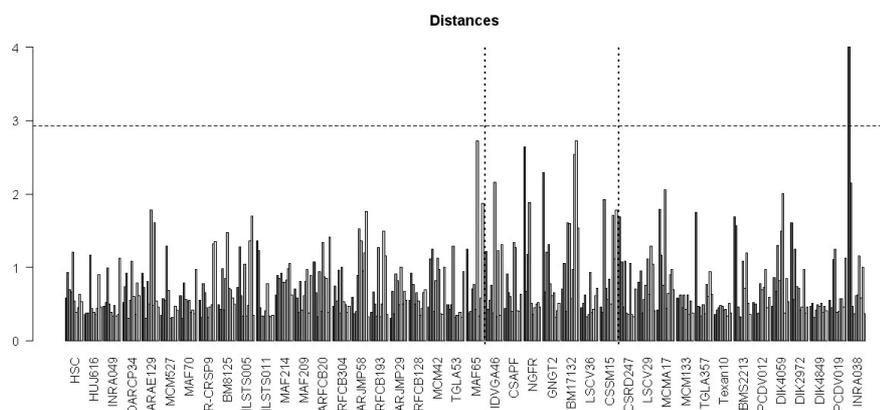
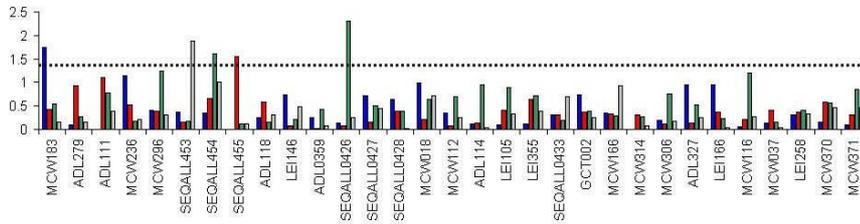


Figure 4 : Tests d'écarts à la dérive génétique dans les populations ovines : P-values après transformation $-\log_{10}$. Pour chaque marqueur, 10 barres sont représentées, chacune correspondant à une distance génétique entre 2 populations. La ligne pointillée horizontale désigne la valeur seuil au niveau 5% après correction selon Bonferroni. Les 2 lignes verticales délimitent les 3 ensembles de marqueurs : les 22 supposés neutres, les 7 de la zone Fecl, les 13 de la zone QTL laitier.

Les P-values des tests d'écart à la dérive, basés sur les variations temporelles des fréquences alléliques dans les lignées de poules sont présentées dans la figure 5, pour chacun des marqueurs situés dans les zones

QTL (a) et chacun des marqueurs neutres (b). On peut noter une suite de valeurs relativement élevées de signification dans la zone QTL située sur le chromosome 14. Après correction pour la multiplicité des tests (taux de faux positifs égal à 5% pour l'ensemble des tests dans une lignée), seuls deux marqueurs présentent des variations de fréquences alléliques qui ne peuvent pas être expliquées par la seule action de la dérive : ADL278, marqueur supposé neutre et SEQALL454, marqueur situé dans la zone 2, sur le chromosome 14.

a) marqueurs des zones QTL ;



b) marqueurs supposés neutres

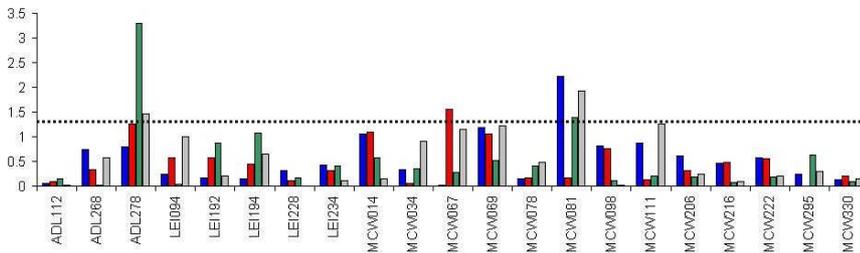


Figure 5. Tests d'écart à la dérive génétique dans les populations de poule : P-values après transformation $-\log_{10}$. Pour chaque marqueur, 4 barres sont représentées, chacune correspondant à la variation temporelle des fréquences alléliques entre la génération G-2 (fondateurs) et la génération G11 (lignée 1 en bleu, lignée 2 en rouge, lignée 3 en vert et lignée témoin en gris). La ligne pointillée horizontale désigne la valeur seuil au niveau 5% avant correction pour la multiplicité des tests.

4. DISCUSSION

Notre étude montre la difficulté de comparer l'évolution des marqueurs neutres et des marqueurs soumis à sélection dans le cas de populations à important fonds génétique commun. En s'appuyant sur les deux dispositifs d'étude, différents éléments peuvent être discutés pour comprendre le fait que des signatures de sélection n'ont pas été clairement mises en évidence.

En dehors de l'espèce considérée, la différence entre les 2 dispositifs de notre étude repose principalement dans le génotypage ou non de la population fondatrice. Laval *et al.*, [11] ont montré sur la base de simulations l'intérêt majeur du génotypage exhaustif des fondateurs.

La densité des marqueurs n'était pas uniforme : plus importante dans la zone FecL d'hyperovulation, par rapport aux autres zones. Notre approche étant basée sur chaque marqueur pris indépendamment, c'est l'accumulation des valeurs significatives dans une zone qui conforte notre conviction que cette zone aurait pu être soumise à la sélection [10], [16]. C'est ce qui semble ressortir de la zone d'hyperovulation, bien qu'aucun des marqueurs ne dépasse le seuil de significativité.

Les marqueurs de type SNP étant moins polymorphes que les microsatellites, leur utilisation avec l'approche choisie dans ce projet, chaque marqueur pris indépendamment, conduirait à des puissances de détection encore plus faibles. Cependant, on peut envisager dans le futur des génotypages beaucoup plus denses et des méthodes de détection utilisant le déséquilibre de liaison [25].

Il est important de noter la difficulté de trouver des zones chromosomiques soumises à sélection et fiables. Au début de notre projet, nous avions une confiance relativement bonne dans les zones proposées ; pourtant, après analyses supplémentaires, il semble que, par exemple, les QTL de poule soient en fait pas ou peu sélectionnés. Toutefois, dans le cas d'une sélection assistée par gène, avérée et forte, comme pour le gène PrP, Palhière *et al.*, [19] ont montré que, en l'espace d'une génération de sélection, une signature de sélection est observée sur les marqueurs proches du gène.

Le nombre de générations entre deux échantillonnages d'une même population, ou depuis la scission d'une population fondatrice en 2 populations filles, était-il optimal dans notre étude ? On peut penser qu'étudier les signatures sur un pas de temps plus court pourrait limiter le bruit de fond lié à la dérive (variance moins grande), comme cela avait été fait [16] chez le porc. Dans ces travaux cependant, les marqueurs avaient pu être sélectionnés au préalable sur la base d'allèles spécifiques des populations parentales de la lignée synthétique étudiée (sino-européenne). Des études de puissance par simulation seraient utiles pour déterminer le nombre optimal de générations pour détecter de la sélection, en fonction de l'intensité de la sélection, de la taille efficace de la population et de la densité du marquage autour des zones sélectionnées. Ces études seront une aide au choix de la statistique la plus appropriée : on a choisi ici les distances génétiques en ovins et la variation temporelle des fréquences alléliques en poules, mais d'autres statistiques sont possibles (F_{st} , longueurs de branches, hétérozygoties, ...).

De nombreuses méthodes statistiques ont été proposées pour la détection de zones chromosomiques ayant été soumises à sélection (voir la revue de [21]). Cependant, les études menées sur des populations ayant divergé depuis très peu de générations sont rares. Dans ce cadre, la supériorité d'une approche par rapport à une autre n'est pas encore démontrée et devra être étudiée. En particulier, les approches utilisant le déséquilibre de liaison devront être évaluées, en fonction de la densité de marqueurs disponibles dans la région d'intérêt.

Une explication supplémentaire des résultats négatifs de notre étude est la suivante. Il est possible que certains allèles aux marqueurs soient partagés entre haplotypes porteurs et non porteurs de l'allèle favorable au QTL. La fréquence de ces allèles a pu paraître stable du fait de la sélection avec l'allèle favorable et de la contre-sélection avec l'allèle défavorable simultanément.

De plus, on peut penser que la sélection massale appliquée sur ces populations a sans doute porté sur un nombre non négligeable de locus, réduisant ainsi la pression de sélection sur chacun d'eux.

Notre étude confirme l'intérêt majeur de caractériser les ressources non seulement au travers de la variabilité neutre mais également au travers de la variabilité fonctionnelle. Il semble donc important de gérer à l'avenir les ressources en tenant compte de cette variabilité fonctionnelle, d'autant plus que les populations sont soumises à une forte pression de sélection. Par ailleurs, l'étude montre que la caractérisation "fonctionnelle" est plus délicate que la caractérisation à partir du polymorphisme neutre : cela renforce l'utilité de consacrer des moyens à cette caractérisation sur du polymorphisme fonctionnel.

REMERCIEMENTS

Les auteurs remercient l'Organisme de Sélection « UPRa Lacaune » et l'Organisme de Sélection « Races Ovines des Massifs – Sélection », pour la fourniture d'ADN des populations ovines, ainsi que l'Unité Expérimentale de Génétique Factorielle Avicole de l'INRA, à Nouzilly, pour l'entretien des lignées sélectionnées de poules. Les auteurs expriment leur reconnaissance au BRG pour le soutien financier apporté à la réalisation de ce projet et au Département de Génétique Animale de l'INRA pour le co-financement des génotypages.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Abasht B., Dekkers J.C., Lamont S.J., 2006. - Review of quantitative trait loci identified in the chicken, *Poult. Sci.* 85: 2079-2096.
- [2] Alvarez I., Gutiérrez J.P., Royo L.J., Fernandez I., Gomez E., Arranz J.J., Goyache F., 2005. - Testing the usefulness of the molecular coancestry information to assess genetic relationships in livestock using a set of Spanish sheep breeds, *J. Anim. Sci.* 83(4): 737-744.
- [3] Arranz J.J., Bayon Y., San Primitivo F., 1998. - Genetic relationships among Spanish sheep using microsatellites. *Anim. Genet.* 29: 434-440.
- [4] Barillet F., Arranz J.J., Carta A., Jacquiet P., Stear M., Bishop S., 2006. - Final consolidated report of the European Union contract of acronym "genesheep-safety". (QLK5-CT-2000-00656), 145 pp.
- [5] Barillet F., 2007. - Genetic improvement for dairy production in sheep and goats. *Small Rumin. Res.* 70: 60-75.
- [6] Bodin L., SanCristobal M., Lecerf F., Mulsant P., Bibé B., Lajous D., Belloc J.P., Eychenne F., Amigues Y., Elsen J.M., 2002. - Segregation of a major gene influencing ovulation in progeny of Lacaune meat sheep. *Genet. Sel. Evol.* 34: 447-464.
- [7] Diez-Tascon C., Litteljohn R.P., Almeida P.A.R., Crawford A.M., 2000. - Genetic variation of closely related populations using microsatellites. *Anim. Genet.* 31(4): 243-251.
- [8] Iannuccelli E., Woloszyn N., Arhainx J., Gellin J. et Milan D., 1996. - GEMMA: A database to automate microsatellite genotyping., *Proceedings of the 25th International Conference on Animal Genetics, Tours-France, Animal Genetics 27, suppl.:* 55.
- [9] Laval G., SanCristobal M., Chevalet C., 2002. - Measuring genetic distances between breeds: use of some distances in various short term evolution models, *Genet. Sel. Evol.* 34: 481-507.
- [10] Laval G., Chevalet C., Pitel F., Amigues Y., Bordas A., Coville J.L., Tixier-Boichard M., 2003. - Modélisation et évaluation expérimentale de la diversité génétique neutre ou adaptative dans le cadre de lignées sélectionnées chez la poule, *Les Actes du BRG*, 4: 35-47.
- [11] Laval G., SanCristobal M., Chevalet C., 2003. - Maximum-likelihood and Markov Chain Monte Carlo approaches to estimate inbreeding and effective size from allele frequency changes. *Genetics* 164 : 1189-1204.
- [12] Lecerf F., Mulsant P., Elsen J.M., Bodin L., 2002. - Localization and mapping of a major gene controlling ovulation rate in Lacaune sheep. In: *Proceedings of the 7th WGCALP, Montpellier, France*, 30: 753-756.
- [13] Loywyck V., 2005. - A comparison of two methods for prediction of response and rates of inbreeding in selected populations with the results obtained in two selection experiments. *Genet. Sel. Evol.* 37: 273-289.
- [14] Mantel, N., 1967. - The detection of disease clustering and generalized regression approach. *Cancer Res*, 27: 209-220.
- [15] Moazami-Goudarzi K., Laloë D., 2002. - Is a multivariate consensus representation of genetic relationships among populations always meaningful. *Genetics* 162: 473-484.

- [16] Milan D., Demeure O., Laval G., Iannucelli N., Genet C., Bonnet M., Burgaud G., Riquet J., Gasnier C., Bidanel J.P., 2003. - Identification de régions du génome répondant à la sélection dans une lignée porcine sino-européenne : la Tai zumu. In: Journées de la recherche porcine en France, 35 : 309-316.
- [17] Nei M., 1973. - Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA, 70: 3321-3323.
- [18] Nei M., Tajima F., 1981. - Genetic drift and estimation of effective population size. Genetics 98: 625-640.
- [19] Palhière I., Brochard M., Verrier E., Moazami-Goudarzi K., Barillet F., Bed'Hom B., Bibé B., Bouix J., Francois D., Leymarie C., Pantano T., 2006. - Did the selection for scrapie resistance impact the genetic variability? Preliminary results on four French sheep breeds, in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, CD-Rom n°30-10.
- [20] Rosenberg N. A., Burke T., Elo K., Feldman M. W., Freidlin P. J., Groenen M. A. M., Hillel J., Mäki-Tanila A., Tixier-Boichard M., Vignal A., Wimmers K., Weigend S., 2001. - Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159: 699-713.
- [21] Sabeti P.C., Schaffner S.F., Fry B., Lohmueller J., Varilly P., Shamovsky O., Palma A., Mikkelen T.S., Altschuler D., Lander E.S., 2006. - Positive natural selection in the human lineage. Science 312 (5780): 1614-1620.
- [22] Siwek M., Cornelissen S.J.B., Nieuwland M.G.B., Buitenhuis A.J., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K., van der Poel J.J., 2003a. - Detection of QTL for immune response to Sheep Red Blood Cells in laying hens, Animal Genetics 34: 422-428.
- [23] Siwek M., Buitenhuis A.J., Cornelissen S.J.B., Nieuwland M.G.B., Bovenhuis H., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K., van der Poel J.J., 2003b. - Detection of different QTL for antibody response to keyhole limpet hemocyanin and *Mycobacterium butyricum* in two unrelated populations of laying hens, Poultry Science 82: 1845-1852. Erratum in: Poultry Science 85 (2006) 1118-1119.
- [24] Siwek M., Buitenhuis A.J., Cornelissen S.J.B., Nieuwland M.G.B., Knol E.F., Bovenhuis H., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K., van der Poel J.J., 2006. - Detection of QTL for innate: non specific antibody levels binding LPS and LTA in two independent populations of laying hens, Developmental and Comparative Immunology 30: 659-666.
- [25] Voight B.F., Kudravalli S., Wen X., Protchard J.K., 2006. - A map of recent positive selection in the human genome. PLoS Biol. Mar. 4(3): e72.

Démasquage des gènes spécifiques d'une espèce génomique du complexe *Agrobacterium tumefaciens* par AFLP et multicapteur à ADN

Perrine PORTIER⁽¹⁾, Denis COSTECHAREYRE⁽¹⁾, Daniel MULLER⁽¹⁾,
David CHAPULLIOT⁽¹⁾, Christine OGER⁽³⁾, Guy PERRIÈRE⁽²⁾,
Laurent GUÉGUEN⁽²⁾, Vincent DAUBIN⁽²⁾, Xavier NESME^{(1)*}.

⁽¹⁾Université de Lyon ; Université Lyon 1 ; CNRS, INRA, Ecologie Microbienne, UMR 5557, USC 1193, 16 rue Raphaël Dubois, Domaine Scientifique de La Doua, 69622 Villeurbanne cedex, France

⁽²⁾ Université de Lyon ; Université Lyon 1 ; CNRS, Biométrie et Biologie Evolutive, UMR 5558, 16 rue Raphaël Dubois, Domaine Scientifique de La Doua, 69622 Villeurbanne cedex, France

⁽³⁾ Université de Lyon ; Université Lyon 1, IFR 41 « Biosciences, Santé et Environnement », PRABI, 16 rue Raphaël Dubois, Domaine Scientifique de La Doua, 69622 Villeurbanne cedex, France

Abstract: Unmasking species specific genes in *Agrobacterium tumefaciens* species G8 by AFLP and microarray. In the frame of studies undertaken to associate a biological concept to the present genomic definition of the bacterial species, a study was done to find species specific genes common to all members of a model species, the genomic species G8 of the species complex *Agrobacterium tumefaciens*. The study was firstly based upon the identification of the genomic origin of AFLP markers of the species by comparing experimental results to *in silico* simulations of AFLP of the complete genome sequence of the G8 reference strain C58. Due to size determination uncertainties and possible close co-migration of predicted fragments, only a little number of amplified AFLP markers could be localized without ambiguities in the genome of C58. Nevertheless, the study delivered a catalog of the genes common to all G8 members. Along to standard house-keeping genes, 25% of the G8 core genome also consisted of hypothetical non-conserved or "ORFan" genes. ORFans, which encode unknown functions, are good candidate to encode species specific functions and thus to allow for further studies about the ecological specialization of the species. The study also revealed that AFLP experimental results were influenced by epigenetic determinants that could have artefactually reduced the number of species specific genes delivered in the study, but not the confidence about those readily found common. The still unknown epigenetic phenomenon is thought to prevent the digestion of certain parts of the genomic DNA by endonucleases required for AFLP. It is likely the cause of the significantly higher number of genes experimentally found common to all species members around the centre of

* Correspondance et tirés à part : nesme@univ-lyon1.fr

the linear chromosome, while the epigenetic phenomenon would have hampered digestions and then delivered artefactual not-common genes in the chromosome lateral branches. The second approach was based on a microarray constructed against the whole genome of C58. A comparative genome hybridization study conducted with members of all the *A. tumefaciens* complex allowed us to determine genes specific of the genomic species G8. Most of these genes are ORFans and knowing their functions requires additional studies. Nevertheless, the study showed that the species specific genes as well as other accessory genes were much more abundant on the linear than on the circular chromosome. The linear chromosome of *A. tumefaciens* therefore looks to be the preferred genome place for genetic innovations.

***Agrobacterium tumefaciens*/ bacterial species/ comparative genomics/ core genome/ AFLP**

Résumé : Dans le contexte de la recherche générale sur le déterminisme génétique des espèces bactériennes, des gènes spécifiques communs à tous les membres d'une espèce modèle ont été recherchés chez l'espèce génomique G8 du complexe *Agrobacterium tumefaciens*. L'étude était basée sur l'identification de l'origine génomique des fragments AFLP marqueurs de l'espèce réalisée par simulation bioinformatique de l'AFLP à partir de la séquence du génome de la souche de référence C58. L'analyse a permis de dresser un catalogue de gènes communs à tous les membres de l'espèce. L'étude montre cependant que les résultats de l'AFLP expérimentale dépendent aussi de caractères épigénétiques qui en réduisent l'efficacité. Ce phénomène, dont le déterminisme n'est pas connu, ne gêne cependant pas l'identification des gènes communs de l'espèce mais a pu en réduire considérablement leur nombre. L'utilisation d'un multicapteur à ADN (ou biopuce) basé sur le génome complet de C58 avec des représentants de toutes les espèces du taxon a permis d'établir le catalogue de tous les gènes vraiment spécifiques à l'espèce G8 d'*A. tumefaciens*. Parmi ceux-ci, il y avait beaucoup de gènes non-conservés (i.e. sans orthologues connus dans les bases de données). Ces gènes orphelins qui codent a priori des fonctions inconnues et originales sont de bons candidats pour déterminer la niche écologique spécifique de l'espèce. L'étude montre en outre que le chromosome linéaire est le lieu privilégié des innovations génétiques dans ce taxon.

***Agrobacterium tumefaciens*/ espèce bactérienne/ génomique comparative/ génome-noyau/ AFLP**

1. INTRODUCTION

Même s'il fait toujours l'objet de discussion, un concept basé sur l'isolement sexuel - le concept d'espèce biologique ou BSC [14] - est associé à la notion d'espèce chez les eucaryotes. Il n'en est pas de même pour les procaryotes où l'espèce est pour l'instant définie uniquement de manière opérationnelle, sans

qu'un concept biologique y soit associé. En effet, bien que sa définition ait grandement évolué en fonction des développements méthodologiques permettant d'accéder à des caractères morpho-biochimiques puis moléculaires, l'espèce bactérienne est actuellement déterminée expérimentalement. Les espèces « génomiques » correspondent à des groupes de souches présentant une forte similitude génomique déterminée classiquement par des mesures d'hybridations de génomes entiers (hybridation ADN-ADN; [22]) ou des méthodes alternatives comme l'amplification sélective de fragments de restriction (AFLP; [15] ou, de plus en plus souvent maintenant, par le séquençage de plusieurs gènes de ménage (MLSA; [7]). Le grand intérêt et l'avantage de cette définition est que, chez la plupart des taxons, les espèces génomiques sont expérimentalement distinguables des autres du fait de l'existence d'une nette discontinuité dans la distribution des fréquences de similitudes génomiques entre espèces [9]. C'est sur cette base empirique (mais non arbitraire) qu'est encore définie l'espèce bactérienne utilisée actuellement en taxonomie bactérienne [18]. Néanmoins, le comité à l'origine de cette définition encourage vivement les études permettant d'associer un concept biologique à cette définition empirique.

L'existence d'espèces génomiques individualisées implique l'existence de forces causant ou maintenant leurs cohésions. Ces forces pourraient résulter soit, comme chez les eucaryotes, d'un phénomène d'isolement sexuel opérant au niveau du taux de recombinaison homologue, soit de la sélection en écotypes comme dans le concept d'espèce écologique de Van Valen [20]. Nous avons étudié la première alternative [6], ce qui nous a permis de montrer que la chute du taux de recombinaison homologue - *i.e.* l'isolement sexuel - entre espèces apparentées était peu importante. En conséquence, au contraire des eucaryotes, l'isolement sexuel chez les bactéries est plus vraisemblablement la conséquence que la cause de la différenciation des espèces. Il n'y a donc pas de réconciliation entre concepts d'espèces eucaryote et procaryote. La force de cohésion alternative, celle résultant d'une relation entre espèce génomique et écotype doit maintenant être étudiée. Pour Cohan [5], l'existence de discontinuités dans la distribution de la diversité génomique est le résultat de sélections adaptatives. Dans les populations asexuées ou faiblement sexuées, une mutation adaptative arrivant dans une cellule individuelle est seulement transmise à sa descendance clonale qui va alors être plus compétitive et va remplacer les autres clones. En conséquence, la population perd sa diversité génétique à tous les locus simultanément. Ce processus est connu sous le nom de « coup de balai » sélectif [10] ou, s'il est répété, de sélection périodique [3]. A notre avis, c'est ce mécanisme qui est également impliqué dans la spéciation bactérienne, c'est-à-dire dans l'apparition d'espèces génomiques différenciées. Dans cette hypothèse, la spéciation bactérienne résulterait de l'acquisition et de la sélection de gènes « spécifiques d'espèce » déterminant

l'adaptation à la niche écologique spécifique de l'espèce. Comme les gènes spécifiques d'espèce sont *a priori* communs à tous les membres de l'espèce, une approche de génomique comparative doit permettre de les découvrir, de les localiser dans le génome et d'identifier les fonctions qu'ils déterminent. *In fine*, la connaissance des gènes spécifiques d'espèce doit permettre d'accéder au type de niche écologique spécifique de l'espèce étudiée.

Pour découvrir des gènes spécifiques d'espèce et établir une correspondance entre espèce génomique et écotype, nous avons étudié un taxon modèle qui possède une grande diversité intra- et inter-spécifique et dont l'un des membres a été complètement séquencé. Ce choix permet de déterminer et d'identifier de façon optimale les gènes réellement communs à tous les membres de l'espèce étudiée, puis, parmi les gènes communs ceux qui sont absents chez les espèces apparentées et donc réellement spécifiques de l'espèce. Le modèle choisi est l'espèce génomique G8 appartenant au complexe d'espèces *Agrobacterium tumefaciens*. En effet, en référence à la définition actuelle, *Agrobacterium tumefaciens* n'est pas une espèce valide mais un taxon complexe qui englobe une dizaine d'espèces génomiques étroitement apparentées appartenant au biovar 1 du genre *Agrobacterium* décrit par Keane *et al.*, [13]. Dans ce complexe, l'espèce génomique G8 présente une importante diversité mise en évidence par hybridation ADN/ADN [16], par AFLP [17] et récemment par MLSA [6]. C58 est la souche de référence de G8. Son génome comprend quatre réplicons : un chromosome circulaire (CcC58, 2 841 420 bp), un chromosome linéaire (LcC58, 2 075 560 bp) et deux plasmides, un plasmide de fonction inconnue dit plasmide « cryptique » (pAtC58, 542 779 bp) et le plasmide Ti (pTiC58, 214 233 bp) déterminant du phénotype tumoral induit par les agrobactéries phytopathogènes [1], [8], [23]. Cependant, comme les plasmides, notamment le plasmide Ti [19], sont des éléments accessoires et mobiles du génome, la recherche de gènes spécifiques d'espèce doit être limitée aux seuls chromosomes.

L'identification de gènes communs aux membres de l'espèce G8 peut être obtenue en couplant AFLP expérimentale et prédictive. L'AFLP permet de produire un ensemble de fragments de restriction de tailles variables provenant aléatoirement de toutes les régions du génome étudié, dont la séparation par électrophorèse fournit un profil caractéristique à chaque souche. L'analyse bioinformatique des profils AFLP montre que cette méthode permet de classer sans ambiguïté les agrobactéries dans des espèces génomiques préalablement définies par hybridation ADN/ADN [15]. Cette capacité discriminante résulte de la présence de fragments AFLP communs à tous les membres d'une espèce, ce qui nous a permis d'établir un catalogue de fragments AFLP marqueurs de l'espèce G8 [17]. En outre, comme l'AFLP est basée sur la reconnaissance de séquences nucléotidiques bien

identifiées (site de restriction et PCR discriminante), il est possible de prédire le résultat d'une AFLP à partir d'une séquence de génome [2].

Dans la présente étude, nous avons couplé les résultats d'AFLP expérimentale et prédictive pour mettre en évidence, identifier et localiser les gènes communs aux agrobactéries de l'espèce génomique G8 d'*Agrobacterium tumefaciens*. Pour cela, un programme permettant de simuler des AFLP à partir de génomes bactériens a été développé et utilisé pour prédire, localiser et identifier les gènes correspondants aux fragments AFLP caractéristiques de la souche C58. Les résultats obtenus ont ensuite été comparés à ceux obtenus à l'aide d'un multicateur à ADN ciblant tous les gènes de la souche C58.

2. MATÉRIEL ET MÉTHODES

2.1. Souches bactériennes

Les souches analysées par AFLP dans cette étude, C58, T37, TT9, LMG 75, LMG 46, 6-Mushin, J-07 et AW137 représentent tous les membres connus à ce jour de l'espèce génomique G8 d'*Agrobacterium tumefaciens*. Elles ont été analysées par AFLP expérimentale par Portier *et al.* [17] en utilisant les conditions *EcoRI*+CA / *MseI*+0, *EcoRI* +CC / *MseI*+0, *EcoRI*+CT / *MseI*+0 et *EcoRI*+CG / *MseI*+0 essentiellement comme décrit par De Vos *et al.* [21] pour établir le catalogue des fragments AFLP communs aux membres de G8.

Les essais d'hybridations génomiques comparatives (CGH) à l'aide d'un multicateur à ADN ont été réalisés, outre les souches G8, avec des représentants de toutes les autres espèces génomiques d'*A. tumefaciens*, TT111 et CFBP 5771 (G1), CIP 497-74 (G2), CIP 107333 (G3), B6 et Kerr 14 (G4), NCPPB 925 (G6), CIP 107445 (G5), Zutra 3-1 (G7), 0363 (G9) et PGF01-22 (G13).

2.2. AFLP prédictive

Le programme Digest a été développé pour prédire à partir de génomes bactériens les fragments AFLP attendus en recherchant les chaînes de caractères correspondants aux sites de restriction et aux nucléotides discriminants. Le programme fonctionne réplicon par réplicon. Après introduction des séquences des enzymes de restriction et des nucléotides discriminants, le programme fournit la position en paires de bases (pb) depuis l'origine du réplicon du début et de la fin du fragment prédit, ainsi que sa taille. Cette taille, augmentée de 27 pb pour tenir compte des adaptateurs nécessaires à l'AFLP expérimentale, donne la taille des

fragments AFLP attendus. Le programme permet de caractériser les plus petits fragments AFLP qui puissent être produits (± 34 pb) sachant que la limite de résolution des électrophorèses, ± 30 pb, permet leur détection. Le programme est disponible sur demande.

Digest a été utilisé pour prédire les résultats des AFLP aux conditions décrites plus haut sur les quatre réplicons de C58 en utilisant les numéros d'accès de GenBank : chromosome circulaire (CcC58), NC_003304 ; chromosome linéaire (LcC58), NC_003305 ; plasmide At (pAtC58), NC_003306 ; et plasmide Ti (pTiC58), NC_003308. L'affectation des fragments prédits à un gène de C58 a ensuite été réalisée en comparant les sorties de Digest à l'annotation du génome de C58 (<http://depts.washington.edu/agro/>).

2.3. Statistiques

L'effet de la proximité de l'origine de réplication du chromosome circulaire CcC58 sur la fréquence des gènes communs a été testée en comparant le nombre d'occurrences des gènes dans les régions autour de l'origine (entre 75 and 25 centisomes) et dans la région du terminus (entre 25 et 75 centisomes). Sur le chromosome linéaire, l'effet a été testé en comparant le nombre d'occurrences dans la région centrale (entre 33,3 et 66,6 centisomes) autour de l'origine de réplication (située à 52 centisomes chez LcC58, cf. <http://depts.washington.edu/agro/>) à la somme des nombres d'occurrences dans les deux extrémités du chromosome (de 0 à 33,3 centisomes et de 66,6 à 100 centisomes). Le test du Chi2 a été réalisé avec le programme Stat View 4.02 software (Abacus Concepts).

2.4. Séquençage de gènes du chromosome linéaire

Les régions génomiques correspondants aux fragments prédits par l'AFLP prédictive ont été amplifiées par PCR en utilisant des amorces définies à partir du génome de C58 en amont et en aval des sites de restrictions *EcoRI* et *MseI* putatifs. Les amorces amont et aval de chaque région ciblée présentaient respectivement une séquence T7 ou T3. Chaque région a ensuite été séquencée sur chaque brin en utilisant les oligonucléotides T3 ou T7. Les séquences réalisées par la société GeneScreen (Lille, France) ont été comparées à la séquence de référence de C58 afin de vérifier la présence des sites de restriction et la taille attendue des fragments afin de prédire la possibilité d'obtenir le fragment pour chaque région et chaque souche.

2.5. Multicapteur à ADN

Les essais d'hybridation comparatives de génome ont été réalisés à l'aide d'un multicapteur à ADN (ou « biopuce ») original constitué de près de 390 000 sondes nucléotidiques couvrant l'intégralité des quatre réplicons de C58 à raison d'une sonde de 50 nucléotides tous les 50 pb, sur les deux brins avec un décalage de 25 nucléotides entre les brins. Afin de détecter d'éventuels transferts et afin de disposer de gènes de contrôles absents des souches testées, le multicapteur présentait également des sondes dessinées de la même façon de plasmides de diverses Rhizobiacées correspondant aux accessions : NC_002377 (pTiA6), DQ058764 (pTiB0542; par courtoisie de P. Oger, accès en ligne non autorisé), NC_002575 (pRi1724), NC_006277 (pAgK84), AJ271050 (pRi2659), AF065242 (pTiChry5), et les plasmides de *Rhizobium loti*: NC_007762, NC_007763, NC_007764, NC_004041, NC_007765, NC_007766. Afin de normaliser les données, le multicapteur comporte également des sondes dessinées à partir de tous les allèles connus à ce jour chez *A. tumefaciens* des gènes *recA*, *mutS* et *gyrB*. Le multicapteur a été produit par NimbleGen Systems Inc. (Madison, USA) qui a procédé également aux étapes de marquage et d'hybridation des ADN génomiques. La pureté spécifique de chaque ADN hybridé a été préalablement contrôlée par séquençage du gène *recA*.

2.6. Normalisation des données de génomique comparative par multicapteur à ADN

Les données brutes d'intensité des sondes (I_i) comprises entre 300 et 65 000 unités environ ont d'abord été transformées en logarithmes base 2, donnant alors une gamme de 3 à 16 environ pour les valeurs de $\log_2(I_i)$, avant extraction du bruit de fond, normalisation et calcul des intensités relatives d'hybridation par rapport aux hybridations homologues.

2.6.1. Extraction du bruit de fond

Pour chaque multicapteur, m , hybridé avec un ADN, a , l'intensité du bruit de fond résultant des hybridations aspécifiques a été déterminée empiriquement à partir de l'histogramme de distribution de $\log_2(I_{im})$. En considérant d'une part que la distribution du bruit de fond est symétrique de part et d'autre du premier mode qui correspond de fait aux hybridations aspécifiques (A_m) et que d'autre part la partie gauche de ce pic correspond uniquement à des hybridations aspécifiques (alors que la partie droite du pic correspond à un mélange d'hybridations aspécifiques et d'hybridations spécifiques de faibles intensités, on fait donc l'hypothèse que les hybridations aspécifiques se distribuent symétriquement à droite du mode. En pratique, on a reporté à droite de A_m l'étendue, Δ_m , de l'étalement de la

courbe observé à gauche de A_m . Ceci permet de calculer l'intensité effective d'hybridation de chaque sonde, $E_{im} = \log_2(I_{im}) - (A_m + \Delta_m)$, en bornant à 0 les valeurs inférieures.

2.6.2. Normalisation des données

Les données de chaque multicateur ont été ensuite normalisées à partir de la moyenne des valeurs de E_{im} obtenues en condition d'identité totale entre sondes et ADN (*i.e.* « *perfect match* »), M_m , obtenues à partir des données d'hybridations des allèles *recA*, *mutS* et *gyrB* de la souche testée pour donner les valeurs normalisées: $N_i = E_{im} / M_m$.

2.6.3. Hybridation relative par gène

Pour chaque souche *a* et chaque gène *g*, l'intensité relative d'hybridation par rapport à celle obtenue en condition d'identité totale entre ADN hybridé et sondes (*e.g.* comparativement à C58 pour les réplicons de C58) a été calculée par : $R_{g,a} = \text{moyenne (des } N_{i,a} \text{ du gène } g) / \text{moyenne (des } N_{i,C58} \text{ du gène } g)$.

La présence d'un gène est considérée comme certaine quand son hybridation représente au moins 5 % de celle obtenue avec C58. Un gène est considéré comme absent ou indétectable quand il hybride à moins de 2 % de la valeur obtenue avec C58. Entre 2 et 5 %, la présence du gène est possible mais il est vraisemblablement ou très divergent ou partiellement supprimé.

3. RÉSULTATS

3.1. Comparaison des AFLP expérimentale et prédictive chez C58

La longueur des fragments obtenus par AFLP expérimentale est sujette à des variations dues aux conditions expérimentales et aux séquences nucléotidiques elles-mêmes. De ce fait, la taille des fragments détectés expérimentalement ne correspond pas exactement aux longueurs prédites par Digest. En comparant visuellement les données expérimentales et prédictives chez C58, nous avons vu que la plupart des fragments détectés pouvaient être alloués à une taille de fragment prédit en considérant un intervalle de confiance correspondant à la taille du fragment ± 1 pb ou ± 1 % pour les fragments de plus de 100 pb. En outre, comme l'incertitude sur la taille augmente fortement avec la taille, nous avons limité l'étude aux fragments de moins de 600 pb.

Pour les quatre conditions *EcoRI*+CA / *MseI*+0, *EcoRI*+CC / *MseI*+0, *EcoRI*+CT / *MseI*+0 et *EcoRI*+CG / *MseI*+0, le nombre de fragments

obtenus par AFLP expérimentales chez C58 était de 44, 34, 64 et 57, respectivement (données non montrées). En considérant l'intervalle de confiance décrit précédemment, le nombre de fragments obtenus expérimentalement présentant une taille compatible avec une taille prédite était respectivement 37, 30, 59 et 48 (données non montrées). Ceci indique que plus de 86 % des fragments expérimentaux peuvent être alloués à au moins un fragment prédit.

L'AFLP simulée montre que pour les conditions décrites précédemment, des fragments originaires de régions génomiques différentes peuvent avoir la même taille ou des tailles trop proches pour déterminer sans équivoque la région d'origine du fragment prédit. Ainsi, le nombre de fragments prédits chez C58 était respectivement 84, 44, 106 et 91, correspondant respectivement à 76, 39, 90 et 87 fragments de tailles différentes, et à 43, 35, 47 et 50 fragments prédits dont la position génomique pouvait être déterminée sans ambiguïté (données non montrées), soit environ 54 % des fragments prédits.

Enfin, le nombre de fragments obtenus expérimentalement chez C58 dont la région génomique d'origine a pu être déterminée sans ambiguïté était respectivement de 20, 26, 34 et 31 pour les conditions AFLP testées (données non montrées), montrant que seulement 55 % des fragments obtenus par AFLP expérimentale ont pu être positionnés sans ambiguïté dans le génome.

Au total, les 111 fragments obtenus par AFLP expérimentale chez C58 ont pu être localisés sur CcC58, LcC58, pAtC58 et pTiC58 (respectivement 58, 33, 9 et 11 fragments). Leur répartition par réplicons (respectivement 50, 30, 9 et 11 %), reflète assez bien la taille relative des réplicons : 50, 37, 10 et 4 % pour 2 841 490, 2 075 960, 542 779 et 214 233 pb, respectivement. La répartition équitable et régulière de l'ensemble des fragments sur les chromosomes linéaire et circulaire (fig. 1) montre que l'AFLP effectue un échantillonnage quasi aléatoire de toutes les régions du génome.

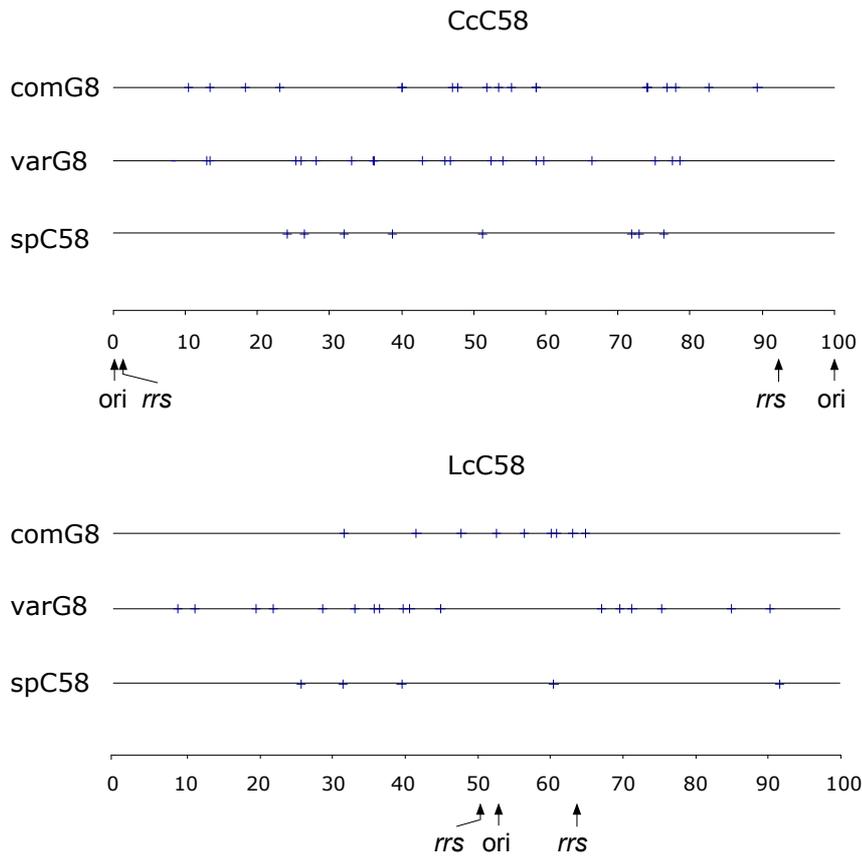


Figure 1 : Localisation génomique des fragments AFLP expérimentalement trouvés communs (comG8) ou variables (varG8) chez les souches de l'espèce génomique G8 d'*Agrobacterium tumefaciens*, ou spécifique à la souche C58 (spC58). La position est donnée en centisome par rapport à l'origine de réplication pour le chromosome circulaire (CcC58) et à partir de l'extrémité du bras gauche pour le chromosome linéaire (LcC58). ori, origine de réplication des chromosomes; rrs, position des opérons ribosomiques.

3.2. Localisation génomique des fragments AFLP communs aux souches de G8.

Parmi les 57 marqueurs AFLP de l'espèce G8 déterminés par Portier *et al.* [17], 35 correspondent à un des 111 fragments positionnés sans ambiguïté dans le génome de C58. Parmi les 22 marqueurs restants, deux proviennent des opérons ribosomiques dont les origines génomiques peuvent également être données sans ambiguïté. Au total 22 (+ 2 pour les opérons ribosomiques 1 et 2), 9 (+ 2 pour les opérons ribosomiques 3 et 4),

2 et 2 fragments communs aux membres de G8 ont été positionnés respectivement sur CcC58, LcC58, pAtC58 et pTiC58. En conséquence, en se basant sur les 111 fragments détectés et positionnés chez C58, il y a donc respectivement 36, 24, 7 et 9 fragments AFLP variables qui n'ont pas été obtenus avec tous les membres de G8.

La répartition relative des gènes d'origine des fragments communs et variables n'est pas significativement différente en fonction des réplicons (Chi 2,816; 2 ddl; $p=0,24$; en regroupant les données des deux plasmides). Par contre, l'examen de leur répartition le long des chromosomes circulaire et linéaire suggère une abondance accrue des gènes correspondants aux fragments communs au centre du chromosome linéaire, c'est-à-dire proche de l'origine de réplication (fig. 1). Les gènes ayant délivré des fragments avec toutes les souches de G8 sont significativement plus abondants autour de l'origine de réplication de LcC58 (Chi2 corrigé = 9,172; 1 ddl; $p = 0,0025$), mais pas autour de l'origine de CcC58 (Chi2 = 0,102; 1 ddl; $p=0,75$).

3.3. Séquençage des régions d'origine des fragments communs et variables de LcC58.

Les zones d'origine de fragments communs et variables du chromosome linéaire ont été amplifiées par PCR et séquencées chez les 8 souches de G8. La majorité des régions ciblées, y compris des régions pour lesquelles des fragments AFLP n'avaient pas été trouvés chez toutes les souches, ont été amplifiées par PCR. Ceci indique que l'absence d'un fragment en AFLP expérimentale n'est généralement pas imputable à une délétion mais suggère plutôt l'intervention de mutations dans les sites de restriction ou les nucléotides discriminants. L'analyse des séquences des produits des PCR a permis de prédire la production ou non de fragment pour chaque gène et chaque souche analysée. Les concordances entre prédictions et résultats expérimentaux (tabl. I) montrent que, lorsque les critères d'allocation entre fragments prédits et fragments amplifiés sont strictement respectés, généralement les fragments non prédits n'ont pas été détectés chez une souche donnée. Par contre, à peine 72 % des fragments prédits ont été obtenus expérimentalement. En conséquence, malgré la surestimation du nombre de fragments variables, l'AFLP permet de localiser correctement les gènes communs aux membres de G8.

Tableau I : Concordance entre AFLP prédictive et expérimentale pour 21 gènes du chromosome linéaire séquencés chez les souches de l'espèce génomique G8 d'*A. tumefaciens*.

Fragment AFLP	Position du fragment	Souches de G8					
		C58	T37	J07	Mushin6	LMG75	LMG46
CT148	5	P-D	P-ND	P-ND	P-ND	P-D	P-ND
CT420	9	P-D	P-D	P-D	P-D	P-D	P-ND
CC585	11	P-D	P-D	P-ND	P-D	P-D	P-D
CC106	20	P-D	P-D	P-ND	P-D	P-D	P-D
CA531	22	P-D	P-D	P-D	P-ND	P-D	P-D
CG147	32	P-D	NP-ND	NP-ND	P-ND	NP-ND	NP-ND
CT89	32	P-D	P-D	P-D	P-D	P-D	P-D
CC168	36	P-D	NP-ND	NP-ND	P-D	P-D	P-ND
CG167	40	P-D	NP-ND	NP-ND	P-ND	P-ND	P-ND
CT593	53	P-D	P-D	P-D	P-D	P-D	P-D
CG405	57	P-D	P-D	P-D	P-D	P-D	P-D
CT303	60	P-D	P-D	P-D	P-D	P-D	P-D
CC261	61	P-D	P-ND	P-ND	P-ND	P-ND	P-ND
CG511	61	P-D	P-D	P-D	P-D	P-D	P-D
CA208	63	P-D	P-D	P-D	P-D	P-D	P-D
CT265	65	P-D	P-D	P-D	P-D	P-D	P-D
CA104	67	P-D	NP-ND	P-D	P-D	NP-ND	NP-ND
CT305	70	P-D	P-D	P-ND	P-D	P-D	P-ND
CC42	85	P-D	P-D	P-D	NP-ND	P-ND	P-ND
CT57	91	P-D	NP-ND	P-D	P-D	P-D	P-D
CC179	99	P-D	P-ND	P-ND	P-ND	P-ND	P-D

La position des fragments est donnée en centisome par rapport à partir de l'extrémité du bras gauche du chromosome linéaire. P, fragment prédit par l'analyse de la séquence; NP, fragment non prédit; D, fragment détecté par AFLP expérimentale; ND, fragment non détecté. Les souches T19 et T37 d'une part, et AW137 et C58 d'autre part présentent des séquences identiques pour tous les gènes testés.

3.4. Annotation des gènes communs de G8

L'annotation des gènes communs aux membres de G8 a été prise à partir du site <http://depts.washington.edu/agro/>. A côté de quelques gènes connus, près de la moitié codent des protéines hypothétiques (tabl. II). Dix de celles-ci correspondent à des protéines conservées chez des genres apparentés: *Sinorhizobium meliloti*, *Mesorhizobium loti*, *Bradyrhizobium japonicum* ou *Rhodospseudomonas palustris* (www.kazusa.or.jp/rhizobase). Dix autres correspondent à des gènes « orphelins » (*ORFan*), sans similarité significative dans les bases de données. Relativement à l'analyse faite via le site <http://cbcsrv.watson.ibm.com/hgt/>, il apparaît que peu de gènes associés à des fragments communs pourraient résulter de transferts horizontaux.

Tableau II : Annotation des gènes trouvés communs aux souches de G8 par AFLP.

CDS	annotation	produit du gène
CcC58		
Atu0307	<i>ugpE</i>	ABC transporter membrane spanning protein [sn-Glycerol-3-phosphate]
Atu0389	tRNA-Arg	tRNA-Arg
Atu0528	<i>mscL</i>	large conductance mechanosensitive channel proteins
Atu0661		glycosyltransferase
Atu1148		hypothetical protein
Atu1148		hypothetical protein
Atu1348		hypothetical protein
Atu1365		hypothetical protein
Atu1480		hypothetical protein
Atu1526	<i>gnd</i>	6-phosphogluconate dehydrogenase
Atu1587		hypothetical protein
Atu1683		hypothetical protein
Atu1685		hypothetical protein
Atu2137		hypothetical protein
Atu2142	<i>glnA</i>	glutamine synthetase
Atu2214		hypothetical protein
Atu2243		hypothetical protein
Atu2375	<i>exoT</i>	UDP-hexose transferase
Atu2571		acyl carrier protein
Atu2734		dehydrogenase
Atu2832	<i>tbdF</i>	tRNA modification GTPase
Atu2834		hypothetical protein
LcC58		
Atu3605		hypothetical protein
Atu3783	<i>gltD</i>	glutamate synthase small subunit
Atu3899	<i>cysG</i>	siroheme synthase
Atu3980		hydrolase
Atu4058	<i>exoW</i>	succinoglycan biosynthesis protein
Atu4133		hypothetical protein
Atu4149	<i>ugdH</i>	UDP-glucose 6-dehydrogenase
Atu4192		ABC transporter substrate binding protein [oligopeptide]
Atu4224	<i>soxB</i>	sarcosine oxidase beta subunit

3.5. Génomique comparative complète par multicapteur à ADN

Un multicapteur à ADN a été utilisé pour détecter par hybridation tous les gènes similaires à ceux de C58. La recherche a été effectuée en hybridant des ADN génomiques extraits des souches de G8 ainsi que de représentants de la plupart des autres espèces génomiques d'*A. tumefaciens*. La prise en compte des autres espèces du complexe permet de discriminer à l'intérieur des gènes communs aux souches de G8, d'une part, les gènes strictement spécifiques à G8 (spG8) et donc absents chez les autres espèces, et, d'autre

part, les gènes qui sont communs (comAt) ou variables (varAt) dans le complexe d'espèce *A. tumefaciens* (fig. 2). La répartition des gènes entre les classes de souches dépend, de façon limitée, de l'intensité d'hybridation minimale choisie pour déterminer la présence d'un gène dans une souche donnée. Au seuil 5 % de l'intensité d'hybridation, la très grande majorité des gènes de C58, 60,6 %, ont été détectés dans toutes les souches de toutes les espèces d'*A. tumefaciens* (tabl. III). Ces gènes constituent le génome-noyau (*core genome*) de ce taxon. Les gènes répondant strictement aux critères fixés pour définir des gènes spécifiques à G8 (*species specific genes*) sont rares: 35 et 53 respectivement sur les chromosomes circulaire et linéaire, soit au total seulement 1,9 % des gènes chromosomiques. Enfin, parmi le génome accessoire (*accessory genome*) quelques gènes n'ont été détectés que chez la souche C58 : 33 et 63 respectivement sur les chromosomes circulaire et linéaire, soit seulement 1,9 % des gènes chromosomiques. Ces derniers gènes abondent dans la branche droite du chromosome linéaire.

Près de 47 % des gènes spécifiques aux membres de G8 codent des protéines hypothétiques. Ils sont le plus souvent groupés par îlots de deux ou trois gènes. Les deux plus gros îlots qui contiennent respectivement 13 (Atu1410-Atu1422) et 12 (Atu4295-Atu4307) gènes ont été trouvés respectivement sur les chromosomes circulaire et linéaire. Le premier contient des gènes ABC transporteurs et le second des éléments de système à deux composants, sans qu'il soit possible de mieux les décrire pour l'instant.

Tableau III : Proportion dans les chromosomes de C58 des gènes trouvés communs, variables ou spécifiques à G8 et aux autres espèces du complexe *A. tumefaciens*.

	Chromosome		Total
	CcC58	LcC58	
spC58	1,2 %	3,3 %	2,0 %
varG8	2,5 %	9,7 %	5,4 %
spG8	1,2 %	2,8 %	1,9 %
varAt	21,5 %	43,0 %	30,1 %
comAt	73,6 %	41,2 %	60,6 %

spC58, gènes spécifiques à la souche C58; varG8, gènes variables dans l'espèce génomique G8; spG8, gènes spécifiques à l'espèce génomique G8 ; varAt, gènes variables chez les autres espèces du complexe *A. tumefaciens*; comAt, gènes communs à tous les membres du complexe d'espèce *A. tumefaciens*. Le critère d'attribution d'un gène à une souche donnée a été fait au seuil 5% de la valeur d'hybridation obtenue avec C58.

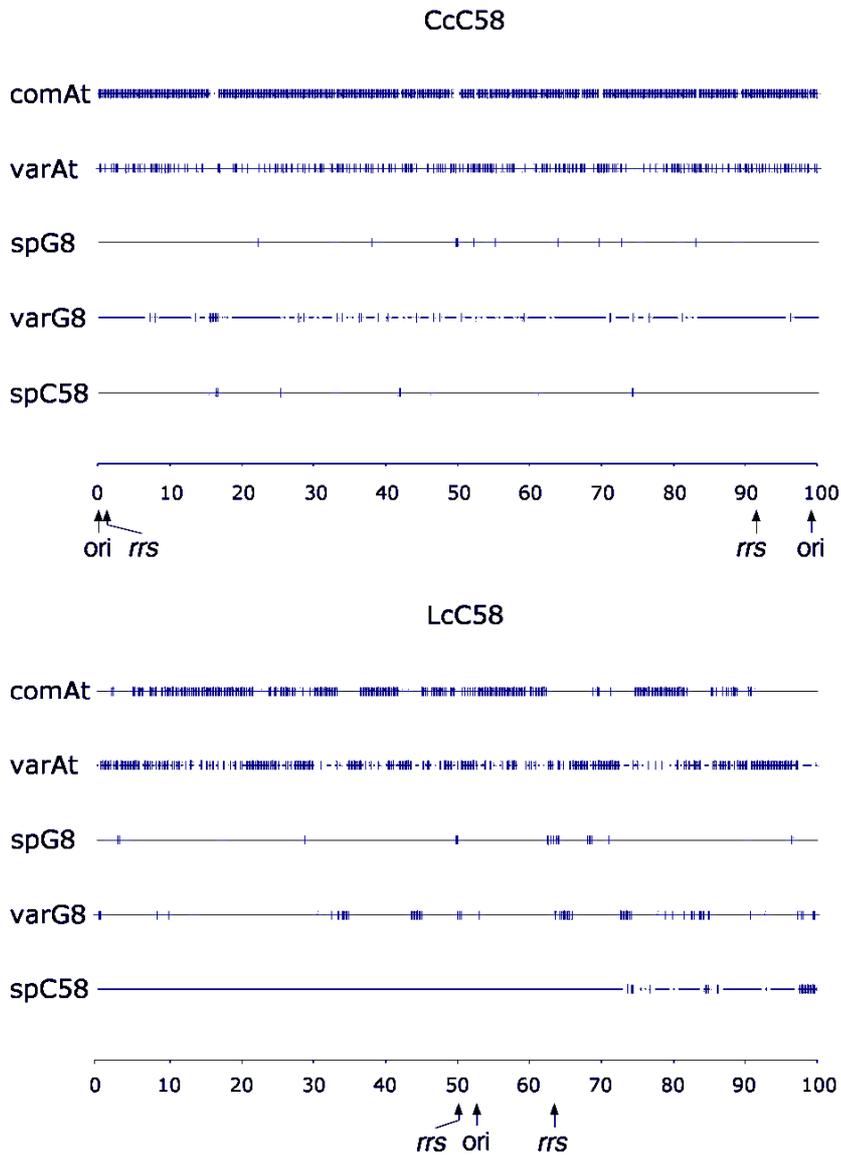


Figure 2 : Localisation des gènes communs (comAt) ou variables (varAt) dans les différentes espèces génomiques d'*Agrobacterium tumefaciens*, ou spécifiques (spG8) ou variable (varG8) dans l'espèce G8, ou spécifique à la souche C58 (spC58).

La position est donnée en centisome par rapport à l'origine de répliation pour le chromosome circulaire (CcC58) et à partir de l'extrémité du bras gauche pour le chromosome linéaire (LcC58). ori, origine de répliation des chromosomes; rrs, position des opérons ribosomiques. de G8. Le critère d'attribution d'un gène à une souche donnée a été fait au seuil 2% de la valeur d'hybridation obtenue avec C58.

4. DISCUSSION

Dans le présent travail, la recherche des parties du génome qui caractérise une espèce bactérienne modèle a été entreprise par deux approches de génomique comparative de simplicité et de coût, mais aussi d'efficacité et de précision croissants. Dans un premier temps, Mougel *et al.* [15] avaient montré que l'AFLP permet de définir clairement et sans ambiguïté les espèces génomiques du complexe *Agrobacterium tumefaciens*. Cette propriété résulte de la capacité de l'AFLP à délivrer un ensemble de fragments communs à toutes les souches d'une espèce donnée [17]. A la suite d'Arnold *et al.* [2], qui avaient montré la possibilité de simuler le résultat de l'AFLP à partir de la séquence nucléotidique d'un génome complet, nous avons tenté de caractériser les gènes d'origine des fragments AFLP marqueurs d'une espèce dont nous connaissons la séquence génomique d'un représentant. Les résultats obtenus lors de la présente étude montrent qu'il est effectivement possible de déterminer une partie des gènes d'origine des marqueurs AFLP d'une espèce. Cependant, l'AFLP présente des contraintes qui sont susceptibles de modifier l'interprétation et la portée des résultats obtenus.

Un résultat constant et surprenant de la présente étude est que la simulation de l'AFLP à partir de séquences nucléotidiques prédit toujours plus de fragments ($\pm 20\%$) qu'il en est observé expérimentalement. Cette observation n'a été reportée, à notre connaissance, ni par Arnold *et al.* [2] travaillant avec *Escherichia coli*, ni par d'autres auteurs. Elle pourrait résulter du modèle bactérien utilisé, et des conditions de réalisation des AFLP expérimentales. En effet, pour obtenir un fragment AFLP, la double digestion de l'ADN génomique est absolument requise. Si, pour une raison ou une autre, un des deux sites de restriction n'est pas accessible à l'endonucléase, il n'y a pas libération d'un fragment de taille attendue. Le phénomène en cause dans la présente étude n'est pas connu. On sait cependant que certaines endonucléases sont sensibles à la méthylation de l'ADN et qu'elles ne digèrent pas alors les sites méthylés. Or, chez *Agrobacterium tumefaciens*, et plus précisément chez C58, la méthylation de l'ADN génomique a été récemment décrite comme un nouveau mécanisme de régulation de l'expression des gènes [12]. L'abondance significative des fragments AFLP communs dans la zone centrale du chromosome linéaire, concomitamment à une abondance des fragments variables dans les branches du chromosome (fig. 1) est probablement également en relation avec ce phénomène. En effet, le séquençage prédit également souvent la présence de fragments communs dans les branches latérales. La discordance entre prédiction et expérience (tabl. I) pourrait résulter d'une inhibition de l'action des endonucléases dans les branches du chromosome linéaire. Si on assume un rôle particulier de la méthylation, on notera que Kahng *et al.*, [12]

ont montré que la méthylation du génome de C58 varie au cours du cycle cellulaire inhibant l'expression des gènes qui ne sont pas indispensables à ce moment. Comme les ADN génomiques utilisés en AFLP ont été extraits au début de la phase plateau, il est probable qu'effectivement une grande partie des bras du chromosome linéaire ait été plus ou moins méthylée à cette étape du cycle. On suppose donc que la variabilité de la détection des fragments originaires des bras du chromosome linéaire résulte essentiellement d'une hétérogénéité des stades de cycle cellulaire entre les souches au moment du prélèvement des ADN. Faute de preuve expérimentale, on ne peut pousser plus loin la spéculation. Cependant, les discordances entre AFLP prédictive et expérimentale suggèrent fortement que cette dernière est déterminée à la fois par des caractères génétiques et des caractères épigénétiques. Nous pensons donc qu'en termes de génomique comparative, les résultats les plus pertinents obtenus à la suite de notre étude concernent uniquement ceux portant sur les fragments AFLP communs à toutes les souches de G8, car la variabilité des autres fragments pourrait avoir un déterminisme aussi bien génétique qu'épigénétique.

La détermination par AFLP des gènes correspondant aux fragments expérimentalement trouvés communs à G8 est en revanche correcte. Un nombre important de gènes codant des protéines hypothétiques sans homologues connus ont été trouvés. Ces derniers gènes sont des candidats de choix pour trouver des fonctions spécifiques de l'espèce. Cependant, et c'est une autre de ses limites, l'AFLP ne permet pas de distinguer parmi les gènes communs ceux qui sont spécifiques à l'espèce G8. En fait, suite à l'analyse ultérieure par multicapteur, nous n'avons trouvé aucun gène réellement spécifique à G8 parmi les marqueurs AFLP de l'espèce G8 déterminés par Portier *et al.* [17].

L'analyse par multicapteur à ADN a été performante car elle nous a permis non seulement d'analyser l'ensemble des gènes identifiés chez la souche de référence mais elle a également permis de rechercher ces gènes dans les autres espèces du complexe *A. tumefaciens*. Un premier type de multicapteur basé sur des sondes de 29 nucléotides avait d'abord été testé. Les résultats obtenus avec les souches de G8 étaient excellents, mais les signaux d'hybridation étaient trop peu distincts du bruit de fond avec les souches des autres espèces. Le signal dépend en effet directement du nombre de mésappariements entre la sonde et l'ADN cible, et la chute du signal ne permettait pas d'analyser des ADN trop divergents. Le choix de sondes de 50 nucléotides s'est révélé correct, car des sondes plus longues donnent un signal d'hybridation exploitable pour une gamme plus importante de taux de mésappariement. Ce nouveau multicapteur a pu être utilisé pour analyser les espèces très apparentées à G8, mais les espèces plus divergentes telles que *A. larrymoorei* ou *A. vitis* et les genres apparentés

Rhizobium et *Sinorhizobium* étaient trop divergents pour être convenablement analysés.

Un peu moins d'une centaine de gènes se sont révélés spécifiques de G8. Leur étude devrait nous permettre d'approcher la connaissance de la spécialisation écologique de l'espèce G8. Néanmoins, près de la moitié de ces gènes sont orphelins, et requièrent donc au préalable des analyses génétiques poussées pour connaître leurs fonctions. Il est donc difficile d'aller plus loin à ce stade. Par contre, l'organisation génomique de ces gènes est révélatrice. D'une part, leur organisation en îlots suggère fortement qu'ils fonctionnent en ensembles discrets de gènes qui effectuent ou régulent la réalisation de fonctions complexes bien définies. Il y aurait donc bien des fonctions/propriétés spécifiques à l'espèce qui méritent d'être étudiées en profondeur. Ceci est à mettre en relation avec l'opinion de Young *et al.* [24], pour qui le génome bactérien peut-être considéré comme ayant deux composantes: un génome-noyau largement chromosomique, partagé avec différents organismes et un génome accessoire dont la distribution est sporadique et localisée dans les plasmides ou dans des îlots génomiques sur les chromosomes.

Enfin, il est tout à fait remarquable que le chromosome linéaire - qui compte environ 40 % du total des gènes chromosomiques - compte plus des deux tiers des gènes spécifiques ou variables (tabl. III). Notre étude montre donc que le chromosome linéaire est le lieu privilégié de la diversité et de l'innovation génétique chez les agrobactéries. Pour Wood *et al.*, [23], il s'agit bien d'un chromosome mais avec des caractéristiques de plasmide. Ce qui conforte l'opinion de Young *et al.* [24]. Enfin, ce sont les bras du chromosome linéaire qui sont particulièrement riches en gènes accessoires. Si les choses sont similaires à ce qui se passe chez *Streptomyces coelicolor* [14], on peut penser que chez *Agrobacterium* également les gènes situés dans les extrémités des branches pourraient déterminer d'importantes adaptations écologiques.

5. CONCLUSIONS - PERSPECTIVES

La présente étude a permis de comparer les génomes d'un ensemble représentatif de la diversité d'une espèce génomique. Une première méthodologie basée sur l'AFLP nous a permis de positionner dans le génome des fragments marqueurs de l'espèce. L'étude a révélé que l'AFLP est sensible à des facteurs épigénétiques qui peuvent biaiser le caractère quasi aléatoire de la distribution des sites de restriction dans le génome. En outre, la comparaison des fragments expérimentaux et prédits est délicate, et un nombre très limité de marqueurs peuvent être positionnés sans

ambiguïté. Cette méthode apparaît donc délicate à mettre en oeuvre et de trop faible puissance pour déterminer des gènes spécifiques d'espèce. Toutefois, si on ne cherche pas à déterminer l'origine génomique des marqueurs, l'AFLP reste une méthode de choix pour l'identification rapide et efficace de nombreux isolats. Pour améliorer la qualité de l'AFLP, une amplification préalable de tout le génome par des procédés enzymatiques permettrait de produire un ADN génomique sans méthylation, palliant ainsi les problèmes de masquage des sites de restriction.

La deuxième méthodologie était basée sur l'utilisation d'un multicapteur à ADN saturant toutes les zones du génome d'une souche de référence de l'espèce étudiée. Cette technique a permis de déterminer les gènes spécifiques d'une espèce, et l'étude approfondie de ces gènes va faire l'objet de la suite de nos travaux. L'approche multicapteur à ADN a cependant deux importantes limites. La première est qu'elle ne permet que la comparaison de génomes très apparentés. En pratique, seules peuvent être analysées des souches d'un même complexe d'espèces, mais pas d'espèces plus éloignées et surtout pas de genres différents. La deuxième limite est que la construction du multicapteur dépend de la disponibilité de la séquence d'une souche de référence par espèce étudiée. Ici, nous n'avons pu déterminer les gènes spécifiques que de l'espèce G8 d'*A. tumefaciens* alors que l'étude de l'ensemble du taxon serait utile. L'offre croissante, et à coût de plus en plus abordable, de séquençages à haut débit devrait cependant permettre d'obtenir rapidement les séquences de la plupart des souches de références des espèces. Il est alors vraisemblable que l'analyse des séquences de génomes complets devienne la méthode de choix pour la détermination des gènes spécifiques de chaque taxon. Les multicapteurs à ADN garderont tout leur intérêt pour étudier l'expression différentielle des gènes, en particulier pour étudier les conditions de milieu susceptibles de constituer la niche écologique spécifique des espèces étudiées.

REMERCIEMENTS

Les auteurs tiennent à remercier le BRG, l'IFR 41 de l'Université Lyon 1 ainsi que le département SPE de l'INRA pour leurs soutiens financiers. Perrine Portier et Denis Costechareyre ont chacun reçu une allocation de recherche du ministère de l'éducation nationale et de la recherche et Daniel Muller un contrat postdoctoral de l'INRA.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Allardet-Servent A., Michaux-Charachon S., Jumas-Bilak E., Karayan L., Ramuz M., 1993. - Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. *J. Bacteriol.* 175:7869-7874.
- [2] Arnold C., Metherell I., Willshaw G., Maggs A. Stanley J., 1999. - Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high-resolution typing method with phylogenetic significance. *J. Clin. Microbiol.* 37: 1274-1279.
- [3] Atwood K. C., Schneider L. K., Ryan F. J., 1951. - Periodic selection in *Escherichia coli*. *PNAS*, 37:146–155
- [4] Choulet F., Aigle B., Gallois A., Mangenot S., Gerbaud C., Truong C., Francou F.X., Fourrier C., Guerinéau M., Decaris B., Barbe V., Pernodet J.L., Leblond P., 2006. - Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol. Biol. Evol.* 23:2361-2369
- [5] Cohan F.M., 2001. - Bacterial species and speciation. *Syst. Biol.* 50: 513-524.
- [6] Costechareyre D., Bertolla F., Nesme X., 2008. Homologous recombination in *Agrobacterium*: potential implication for the genomic species concept in bacteria. *Mol. Biol. Evol.* *Accepté pour publication.*
- [7] Gevers D., Cohan F.M., Lawrence J.G., Spratt B.G., Coenye T., Feil E.J., Stackebrandt E., Van de Peer Y., Vandamme P., Thompson F.L., Swings J., 2005. - Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3:733-739.
- [8] Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Quorollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F., Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B., Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G., Cielo C. Slater S., 2001. - Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science.* 294: 2323-2328.
- [9] Grimont P.A., 1988. - Use of DNA reassociation in bacterial classification. *Can J Microbiol.* 34:541-546.
- [10] Guttman D. S., Dykhuizen D. E., 1994. - Detecting selective sweeps in naturally occurring *Escherichia coli*. *Genetics*, 138:993–1003.
- [11] Jumas-Bilak E., Michaux-Charachon S., Bourg G., Ramuz M. Allardet-Servent A., 1998. - Unconventional genomic organization in the alpha subgroup of the proteobacteria. *J. Bacteriol.* 180: 2749-2755.
- [12] Kahng L.S., Shapiro L., 2001. - The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. *J. Bacteriol.*, 183:3065-3075.
- [13] Keane, P.J., Kerr A., New P.B., 1970. - Crown gall on stone fruits: identification and nomenclature of *Agrobacterium* isolates. *Aust. J. Biol. Sci.* 23:585–595.
- [14] Mayr E., 1942. - Systematics and the origin of species from the viewpoint of a zoologist, Columbia University Press, New York. 344p.
- [15] Mougél C., Thioulouse J., Perriere G. Nesme X., 2002. - A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* 52: 573-586.

- [16] Popoff M.Y., Kersters K., Kiredjian M., Miras I. Coynault C., 1984. - Taxonomic position of *Agrobacterium* strains of hospital origin. *Ann. Microbiol.* 135a:427-442.
- [17] Portier, P., Fischer-Le Saux M., Mougél C., Lerondelle C., Chapulliot D., Thioulouse J., Nesme X., 2006. - Identification of genomic species in *Agrobacterium* biovar 1 by AFLP genomic markers. *Appl. Environ. Microbiol.* 72:7123-7131.
- [18] Stackebrandt E., Frederiksen W., Garrity G.M., Grimont P.A., Kämpfer P., Maiden M.C., Nesme X., Rossello-Mora R., Swings J., Truper H.G., Vauterin L., Ward A.C. Whitman W.B., 2002. - Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52: 1043-1047.
- [19] Teyssier-Cuvellé S., Oger P., Mougél C., Groud K., Farrand S.K., Nesme X., 2004. - A highly selectable and highly transferable Ti plasmid to study conjugal host range and Ti plasmid dissemination in complex ecosystems. *Microbiol. Ecol.* 48:10-18.
- [20] van Valen L., 1976. - Ecological species, multispecies, and oaks. *Taxon* 25:233-239.
- [21] de Vos P., Hogers R., Bleeker M., Reijans M., Van De Lee T., Hornes M., Fritjers A., Pot J., Peleman J., Kuiper M., 1995. - AFLP: a new technique for DNA fingerprinting. *Nucleic acids res.* 23: 4407-4414.
- [22] Wayne L.G., Brenner D.J., Colwell R.R., Grimont P.A.D., Kandler O., Krichevsky M.I., Moore L.H., Moore W.E.C., Murray R.G.E., Stackebrandt E., Starr M.P. Truper H.G., 1987. - Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *IJSB.* 37: 463-464.
- [23] Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida Jr N.F., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee Sr D., Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C., Kutuyavin T., Levy R., Li M.J., McClelland E., Palmieri A., Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.Y., Dolan M., Chumley F., Tingey S.V., Tomb J.F. Milton P.G., Maynard V.O. And Nester W.E., 2001. - The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science*, 294: 2317-2323.
- [24] Young J.P., Crossman L.C., Johnston A.W., Thomson N.R., Ghazoui Z.F., Hull K.H., Wexler M., Curson A.R., Todd J.D., Poole P.S., Mauchline T.H., East A.K., Quail M.A., Churcher C., Arrowsmith C., Cherevach I., Chillingworth T., Clarke K., Cronin A., Davis P., Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Mungall K., Norbertczak H., Rabinowitsch E., Sanders M., Simmonds M., Whitehead S., Parkhill J., 2006. - The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* 7:R34.

Identification des gènes candidats de la tolérance au zinc chez la pseudométallophyte *Arabidopsis halleri* par l'intégration des données de génomique comparative et de transcriptomique disponibles chez *Arabidopsis*

Nancy ROOSENS⁽¹⁾, Zaigham SHAHZAD⁽²⁾, Eric LACOMBE⁽²⁾,
Hélène FRÉROT⁽¹⁾, Anne CREACH⁽¹⁾, Pierre BERTHOMIEU⁽²⁾,
Pierre SAUMITOU-LAPRADE^{(1)*}

⁽¹⁾Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016, Université des Sciences et Technologies de Lille - Lille1, 59655 Villeneuve d'Ascq Cedex, France

⁽²⁾Biochimie et Physiologie Moléculaire des Plantes – UMR CNRS-INRA-Montpellier SupAgro-Université Montpellier 2, Campus SupAgro-INRA, Place Viala, 34060 Montpellier Cedex 1

Abstract: Zn hyperaccumulators have attracted considerable interest and represent an important genetic resource, mainly because Zn hyperaccumulation can potentially be used in phytoremediation and biofortification. Moreover this specialized flora displays naturally selected tolerance to extremely high metal concentrations—a trait firmly established in the literature as one of the best examples of adaptation. In the present study we constructed a genetic linkage with 81 markers anchored on *Arabidopsis thaliana*, including 25 genes known to be involved in metal homeostasis. Using these markers, we compared the genetic linkage map of *A. halleri* × *A. l. petraea* to the annotated *A. thaliana* genome sequence to generate a tool for *A. halleri* genomic approaches. First this provided an extensive overview of the chromosomal rearrangements that have occurred since the divergence between *A. thaliana* and its closest relative *A. halleri*. Second, on the basis of the syntenic relationships assessed experimentally through this work, we transferred the quantitative trait regions (QTR) confidence intervals for Zn tolerance to the *A. thaliana* physical map, allowing access to all the genes localized in the corresponding regions. Third, we validated from the 25 genes involved in metal homeostasis the 3 ones localized in the QTR (*HMA4*, *MTP1-a* and *MTP1-b*) that can be considered the best candidates for conferring Zn tolerance. Using these genes as probes, we selected from a *A. halleri* BAC library the clones covering, at least partially, the pics of the QTL of zinc tolerance in order to detect possible candidate genes present in *A. halleri* specifically. The sequencing of the BAC containing *HMA4* did not detect additional genes with a known function in metal ho-

* Correspondance et tirés à part : Pierre.saumitou@univ-lille1.fr

meostasis, confirming *HMA4* as the best candidate on which additional investigations have to be performed.

zinc tolerance/ comparative genomics/ *Arabidopsis halleri*/ QTL mapping/ synteny

Résumé : Les espèces tolérantes et hyperaccumulatives de métaux représentent une ressource génétique importante dans le développement des stratégies de phytoremédiation. En particulier elles représentent un outil de choix pour l'identification et la caractérisation des gènes de l'homéostasie des métaux. Dans cette étude nous avons complété par 20 nouveaux marqueurs la carte génétique d'*Arabidopsis halleri* composée aujourd'hui de 81 marqueurs ancrés chez *A. thaliana*. Grâce à ces marqueurs, la colinéarité entre génomes a été vérifiée, les intervalles de confiance des 3 régions associées au caractère quantitatif (QTR) de tolérance au zinc ont été recalculés et transférés sur la carte physique d'*A. thaliana* en tenant compte des zones de syntenie. Ce transfert a permis de vérifier la correspondance de position de 25 gènes connus pour leur rôle dans l'homéostasie des métaux et de détecter parmi eux trois gènes localisés dans les QTR : *HMA4*, *MTP1-a* et *MTP1-b*. Ces gènes peuvent être considérés comme candidats de la tolérance au zinc. Les clones BAC construits à partir du génome d'*A. halleri* portant trois de ces gènes candidats ont été identifiés pour réaliser la séquence. Le séquençage du BAC contenant *HMA4* confirme l'absence d'autre gène de l'homéostasie des métaux à proximité.

tolérance au zinc/ génomique comparative/ *Arabidopsis halleri*/ cartographie QTL / syntenie

1. INTRODUCTION

1.1. La phytoremédiation : une méthode prometteuse dont le bon développement dépend d'une meilleure connaissance des mécanismes fondamentaux de l'homéostasie des métaux.

Il existe en France de nombreux sites et sols pollués par les métaux. A ce jour, les méthodes de traitements physiques ou chimiques et l'incinération de ces matériaux classés comme déchets ne sont pas réalistes en termes de coût financier et environnemental. Les méthodes douces de remédiation utilisant la capacité de certaines plantes à tolérer et à accumuler les métaux et groupées sous le terme générique de phytoremédiation, peuvent fournir une solution alternative à ce problème. La mise en œuvre de la phytoremédiation intègre les critères du développement durable. C'est une méthodologie *écologique* qui permet dans le temps de restaurer les performances des sites en matière de biodiversité de la faune et de la flore. Elle est *économique* car elle présente un coût d'investissement restreint. La phytoextraction apparaît comme une méthode prometteuse pour le traitement des sols moyennement pollués. Son

inconvenient majeur réside dans la durée du processus (qui peut aller au delà de la dizaine d'années) nécessaire pour ramener les taux de contamination en dessous des seuils admis. Cependant cet inconvenient peut être réduit voir annulé si la phytoextraction met en œuvre des espèces permettant à la fois l'extraction des métaux du sol et la production d'énergies renouvelables comme par exemple le bois de chauffage ou des agro-carburants. Aujourd'hui la phytoextraction peine à se développer en particulier sur les sites pollués par le zinc et le cadmium parce qu'il n'a pas été possible d'identifier de plante à forte biomasse qui accumulerait les métaux en fortes concentrations. Les espèces hyperaccumulatrices identifiées dans la nature sont petites et n'ont jamais fait l'objet d'amélioration de leur potentiel agronomique et les collections d'espèces d'intérêt agronomique n'ont en général jamais été sélectionnées pour leur potentiel d'accumulation (ou d'exclusion) des métaux du sol. Un des principaux freins pour disposer de la plante « idéale » est le manque de connaissances fondamentales sur les mécanismes responsables des processus d'homéostasie des métaux et en particulier ceux conduisant à la tolérance et à l'hyperaccumulation des métaux.

D'un point de vue fondamental, l'évolution de la tolérance aux métaux lourds constitue, depuis les travaux de Bradshaw [5], un modèle de choix pour l'étude de l'effet de la sélection sur l'établissement et le maintien de caractères adaptatifs dans les populations naturelles. Il s'agit d'une sélection directionnelle de très forte intensité, agissant à une échelle micro-évolutive en raison de la colonisation de sites pollués d'origine anthropique, et dont le phénotype peut être mesuré expérimentalement, par exemple par un test de tolérance en condition hydroponique fondé sur la mesure de la dose minimale de zinc dans la solution de culture conduisant à l'arrêt de la croissance racinaire [20]. Cependant, les travaux réalisés jusqu'à présent ont concerné soit le polymorphisme du trait phénotypique « tolérance aux métaux lourds », soit l'influence de l'évolution de populations tolérantes sur l'organisation du polymorphisme au niveau de marqueurs neutres. La mise en évidence et la caractérisation de gènes directement impliqués dans la tolérance aux métaux permettrait d'étudier l'effet de la sélection sur le polymorphisme des gènes cibles.

1.2. *Arabidopsis halleri*: une ressource génétique pour l'identification des mécanismes fondamentaux de l'homéostasie des métaux et de l'adaptation aux milieux pollués par les métaux

Au cours des dix dernières années, l'intérêt pour les plantes hyperaccumulatrices a fortement augmenté en raison de leur utilisation potentielle en phytoremédiation [1]. Ces plantes ont acquis au cours de leur évolution la capacité d'accumuler dans leurs parties aériennes des

concentrations exceptionnellement élevées de métaux [17]. Les mécanismes qui permettent aux plantes accumulatrices de tolérer de fortes concentrations de métaux dans leurs tissus ne sont pas encore élucidés bien que l'on puisse penser qu'ils ont évolué dans le réseau des gènes de l'homéostasie des métaux [7]. Pour étudier les bases physiologiques et moléculaires de la tolérance aux métaux chez les plantes hyperaccumulatrices plusieurs modèles biologiques ont été décrits en particulier dans la famille des Brassicacées [1], [16], [3], [24]. Parmi les espèces tolérantes hyperaccumulatrices de zinc et de cadmium, *Arabidopsis halleri* (Brassicaceae) est une des plus prometteuses. En effet, il s'agit de la pseudo-métallophyte (c'est-à-dire présente sur sites non pollués et pollués par les métaux) la plus proche phylogénétiquement du modèle biologique *A. thaliana* dont elle n'aurait divergé que depuis 5 millions d'années. Même si les deux espèces diffèrent pour le nombre de chromosomes ($n=5$ pour *A. thaliana* et $n=8$ pour *A. halleri*), elles présentent 94 % d'identité nucléotidique dans les régions codantes de leurs gènes [3]. Cette identité a permis d'utiliser les puces à ADN d'*A. thaliana* pour comparer les niveaux d'expression de leurs gènes respectifs en réponse aux concentrations de métal dans le milieu de culture et d'identifier des gènes présentant une réponse différentielle au métal chez la pseudo-métallophyte [3], [9], [25].

1.3. L'analyse génétique par croisements inter-spécifiques a permis de préciser l'architecture génétique de la tolérance au zinc et de localiser les régions du génome responsables de la variation du trait.

Différents travaux fondés sur l'étude de descendances produites par croisements inter-spécifiques d'*A. halleri* (*Ab*) avec sa plus proche parente *A. lyrata petraea* (*Alp*), espèce non tolérante et non accumulatrice ($n=8$ chromosomes), ont permis de contourner la contrainte liée à la nature constitutive de la tolérance au zinc [15] et de mettre ce trait en ségrégation [14]. Très récemment une étude fondée sur l'analyse d'une première génération de backcross (BC1) à partir d'un croisement *Ab* x *Alp* a permis d'établir l'architecture génétique de la tolérance au zinc chez *A. halleri* [26]. Trois régions associées à la variation quantitative de la tolérance au zinc (Quantitative Trait Regions ou QTR) ont été identifiées et leurs contributions respectives (17,7, 16,2 et 8,1 % de la variance génétique) à l'expression du trait ont été établies. L'étude a également montré que, de par leur taille importante (24, 4, et 13 cM respectivement), chacune des régions serait susceptible de contenir des centaines voire des milliers de gènes.

1.4. Objectifs de l'étude

L'objectif de cette étude est d'accroître notre connaissance des QTR de la tolérance au zinc grâce à l'analyse comparée des génomes d'*A. halleri* et d'*A. thaliana*, l'intégration des données de transcriptomique acquises chez *A. thaliana* et l'analyse des séquences d'*A. halleri* dans les QTR. Nous avons développé une cartographie comparée d'*A. halleri* avec *A. thaliana* dont la complète annotation du génome d'*A. thaliana* (Arabidopsis Genome Initiative 2000) fournit une base très solide pour les études de cartographie comparée au sein de la famille des Brassicacées [11]. La carte génétique d'*A. halleri* [26] composée pour partie de 61 marqueurs ancrés chez *A. thaliana* a été complétée par 20 nouveaux marqueurs [18] dont la position est connue chez *A. thaliana*. La colinéarité entre génomes a été vérifiée et les intervalles de confiance des 3 QTR recalculés et transférés chez *A. thaliana* en tenant compte des zones de syntenie, ce qui a permis de réduire la taille des QTR de la tolérance au zinc. La correspondance de position de 25 gènes impliqués dans l'homéostasie des métaux a été effectivement vérifiée entre les deux espèces et parmi ces gènes trois ont pu être localisés dans les QTR. Cependant, parce qu'il est possible qu'une partie des gènes des QTR d'*A. halleri* ne soient pas présents chez *A. thaliana* dans les régions transposées, les clones BAC d'*A. halleri* portant trois de ces gènes candidats ont été identifiés et pour partie séquencés.

2. MATÉRIEL ET MÉTHODES

2.1. Matériel végétal et extraction d'ADN

L'étude a été réalisée à partir d'un croisement inter-spécifique de type back-cross (BC1) entre un père *A. halleri* (*P1*) du site pollué d'Auby (France) et une mère *A. lyrata petraea* (*P2*), dont un descendant F1 (*P3*) a été utilisé comme père dans un rétrocroisement sur une mère de l'espèce *A. l. petraea* (*P4*). Les génotypes *P2* et *P4* ont été choisis différents pour éviter tout problème de dépression de consanguinité chez une espèce allogame. L'ADN génomique des quatre génotypes parentaux et de leurs 336 descendants en BC1 caractérisés pour leur tolérance au zinc [26] a été extrait pour la recherche de marqueurs moléculaires selon la méthode de Dellaporta modifiée [19].

2.2. Augmentation de la densité de marqueurs sur la carte de liaison et analyse QTL

Afin de densifier la carte génétique *Ab* x *Alp*, nous avons défini des couples d'amorce d'amplification homologues de la séquence génomique d'*A.*

thaliana à partir des bases de données publiques (Base de données TAIR à <http://www.arabidopsis.org>). Les nouvelles amorces ciblent des régions exoniques (de tailles comprises entre 650 et 800 paires de bases) dans des gènes choisis en fonction de leur rôle dans l'homéostasie des métaux (tabl. I) et/ou de leur position chez *A. thaliana*. Les amorces Tc-up3, Tc-ip2, Ah-IRT2 ont été définies chez la pseudo-métallophyte *Thlaspi caerulescens* (Brassicacée) [2]. La production et la détection des fragments d'amplification et/ou de restriction ont été réalisées selon un protocole décrit précédemment [18].

Tableau I : Liste des marqueurs ancrés chez *A. thaliana* utilisés pour la construction de la carte de liaison (*Ab x Alp*) et pour la comparaison avec la carte physique d'*A. thaliana*.

Localisation chez <i>A. thaliana</i> (en kb)	Locus	Identification chez <i>A. thaliana</i>	Fonction putative de la protéine chez <i>A. thaliana</i>
Groupe de liaison LG1			
Chr1-00240	AthACS ^a	At1g01660	
Chr1-01019	Ah-PCS2 ^b	At1g03980	Phytochelatine synthase 2
Chr1-01498	AXR1 ^a	At1g05180	
Chr1-02985	Ah-NAS1 ^b	At1g09240	Nicotiamine synthase related protein
Chr1-03095	PhyA ^a	At1g09570	
Chr1-03665	Ah-ZNT1 ^b	At1g10970	Transporteur de métal (ZIP4)
Chr1-04266	ICE10 ^a	F12F1	
Chr1-04516	ICE13 ^a	At1g13220	
Chr1-05639	Tc-up3 ^b	At1g16500	
Chr1-08062	GI ^a	At1g22770	
Chr1-10858	ATTS0392 ^a	At1g30630	
Chr1-16487	VIP1 ^a	At1g43700	
Chr1-17268	At1g46768 ^a	At1g46768	
Groupe de liaison LG2			
Chr1-22940	F19K23 ^a	At1g62050	
Chr1-21141	Ah-NAS2 ^b	At1g56430	Nicotiamine synthase related protein
Chr1-24870	SLL2 ^a	At1g66680	
Chr1-25543	Lyr132 ^a	T23K23	
Chr1-26613	Nga111 ^a	F28P22	
Chr1-28476	At1g75830 ^a	At1g75830	
Chr1-28980	ADH1 ^a	At1g77120	
Chr1-30378	Ah-NRAMP1 ^b	At1g80830	Transporteur de métal (NRAMP1)
Groupe de liaison LG3			
Chr3-00303	Tc-up2 ^b	At3g01860	
Chr3-01936	Ah-FRD3 ^b	At3g08040	MATE efflux family protein; similar to FRD3 (Ferric Reductase defective 3)
Chr3-02889	Ah-MT2a ^b	At3g09390	Métallothionéine (MT2a)
Chr3-02624	Ah-ZIP1 ^b	At3g08650	Transporteur de métal (ZIP1)
Chr3-04577	MDC16 ^a	MDC16	
Chr3-08098	DMC1 ^a	At3g22880	
Chr3-08806	At3g24300 ^b	At3g24300	
Chr2-06207	At2TCA ^a 1	T9F8	
Chr2-04317	At2g10940	At2g10940	
Chr2-06252	At2g14620	At2g14620	
Chr2-07228	Ah-CCH	At2g18196	Copper Chaperone related protein
Chr2-08286	Ah-HMA4 ^a	At2g19110	ATPase (HMA4) transporteur de Cd / Zn
Chr2-10216	ICE14 ^a	F11A3	
Groupe de liaison LG4			
Chr2-09133	Con ^a	At2g21320	

Génétique de la tolérance au zinc chez *Arabidopsis thaliana*

Chr2-09533	At2g22430 ^a	At2g22430	
Chr2-09834	CK2alpha2 ^a	At2g23080	
Chr2-09863	Ah-NRAMP3 ^a	At2g23150	Transporteur de métal (NRAMP3)
Chr2-11066	ELF3 ^a	At2g25930	
Chr2-12324	At2g28700 ^a	At2g28700	
Chr2-12846	Ah-ZIP6 ^b	At2g30080	Transporteur de métal (ZIP6)
Chr2-14018	At2g33010 ^a	At2g33010	
Chr2-15786	ACT3 ^a	At2g37620	
Chr2-15997	Ah-CAX1 ^b	At2g38170	Echangeur de Ca (CAX1)
Chr2-16299	ICE12 ^a	At2g39010	
Chr2-16779	At2g40140 ^a	At2g40140	
Chr2-17894	At2g43010 ^a	At2g43010	
Chr2-19272	ICE11 ^a	F11C10	
Chr2-19245	Ah-MTP1A ^a	At2g46800	Transporteur de Zn (ZAT1, MTP1)
Chr2-19598	Ah-MTPc3 ^b	At2g47830	Transporteur de métal chloroplastique ; protéine putative de transport de métal / tolérance au métal (MTPc1)
Chr2-19531	Ah-MHX1 ^b	At2g47600	Echangeur Magnesium/proton (MHX1)
Groupe de liaison LG5			
Chr2-04411	Nga1145 ^a	T16F16	
Chr3-10526	At3g28220 ^a	At3g28220	
Chr3-14096	At3g33530 ^a	At3g33530	
Chr3-16844	At3g45810 ^a	At3g45810	
Chr3-19036	F3H ^a	At3g51240	
Chr3-21509	Ah-MTPc3 ^a	At3g58060	Famille de protéine d'efflux de cation / protéine de tolérance au métal MTPc3)
Chr3-23179	Nga112 ^a	At3g62650	
Groupe de liaison LG6			
Chr4-04240	At4GA2 ^a	T18A10	
Chr4-01123	LD ^a	At4g02560	
Chr4-06346	AthDET1 ^a	At4g10180	
Chr5-08819	ICE2 ^a	K18P6	
Chr5-08352	AthCDPK9 ^a	MQM1	
Chr5-04824	Ah-HMADP2 ^b	At5g14910	Protéine contenant un domaine associé aux métaux lourds
Chr5-02626	At5g08160 ^a	At5g08160	
Ch 5-00980	ATHCTR1 ^b	CTR1.2	
Chr5-00506	Ah-MT2b ^a	At5g02380	Métallothionéine (MT2b)
Groupe de liaison LG7			
Chr4-17925	At4g38220 ^a	At4g38220	
Chr4-17540	Ah-HMA1 ^b	At4g37270	ATPase transporteur de Cd / Zn (HMA1)
Chr4-15994	At4g33160 ^a	At4g33160	
Chr4-17718/19456	Cha15 ^a	F11C18	
Chr4-13587	TSB2 ^a	At4g27070	
Chr4-16669	At4-TC1 ^a	T19F6	
Chr4-11384	SRK ^a	At4g21366	
Chr4-09449	HAT4	At4g16780	
Chr4-09207	FCA ^a	At4g16280	
Chr4-10703	Ah-IRT2 ^b	At4g19680	Transporteur de métal Fe / Zn (IRT2)
Chr5-14025	PhyC ^a	At5g35840	
Chr5-15022	ATTSO191 ^a	At5g37780	
Chr5-16149	ICE9 ^a	At5g40340	
Groupe de liaison LG8			
Chr5-17648	ATCLH2 ^a	At5g43860	
Chr5-20841	EMF2 ^a	At5g51230	
Chr5-24008	ZIP2 ^a	At5g59520	Transporteur de Zn (ZIP2)
Chr5-24702	MHJ24 ^a	MHJ24	
Chr5-26879	Ah-NRAMP4 ^a	At5g67330	Transporteur de métal (NRAMP4)

La position du marqueur sur la carte physique d'*A. thaliana*, l'identifiant du locus chez *thaliana*, la référence de l'article dans lequel sont décrites les conditions d'amplification PCR et le type de polymorphisme détecté sont présentés : ^a [25], ^b [18], ^c . Les gènes ayant une fonction avérée dans l'homéostasie des métaux sont indiqués dans les cases grisées.

Pour les marqueurs présentant une différence de taille du produit d'amplification chez les parents (*Ab-ZNT1*, *Tc-up3*, *Ab-NRAMP1*, *Tc-up2*, *Ab-ZIP6*, *Ab-HMADP2*, *AtHCTR1*, *Ab-IRT2*) le poids moléculaire a été estimé par le logiciel Saga-GT en référence au marqueur de taille Li-Cor 50–700 pb. Pour les marqueurs ne présentant pas de polymorphisme de taille chez les parents (*Ab-PCS2*, *Ab-NAS1*, *Ab-NAS2*, *At3g24300*, *Ab-CAX1*, *Ab-MTPc1*, *Ab-MHX1*, *HAT4*), une analyse de séquence des fragments parentaux a permis d'identifier des variations nucléotidiques ciblant des sites de restriction susceptibles d'être détectées chez les descendants par une analyse du polymorphisme de restriction (RFLP) sur gel d'agarose coloré au bromure d'éthidium. Enfin, pour les marqueurs *Ab-MT2a*, *Ab-ZIP1* et *Ab-ZIP2*, le polymorphisme a été détecté par la méthode de Southern selon un protocole décrit précédemment [26]. Les sondes d'hybridation ont été produites à partir clones d'*A. thaliana* correspondant aux gènes ciblés chez *A. balleri* et référencés dans GenBank (110G24T7, AF033535, et AF033536)

2.3. Analyse comparée de la carte de liaison *Ah* x *Alp* et de la carte physique d'*A. thaliana*

La carte de liaison *Ab* x *Alp* produite dans ce travail a été réalisée à partir des 61 marqueurs d'ancrage décrits précédemment [26] enrichie de 20 nouveaux marqueurs également ancrés. La carte de liaison a été construite à l'aide du logiciel JoinMap 3.0 [22]. En appliquant les paramètres d'analyses définis par Willems *et al.* [26], un intervalle de confiance de 2 Lod des régions QTR a pu être défini par le logiciel Mapchart [23] à partir des valeurs de Lodscore calculées par MapQTL [22]. La position en kilobase (kb) sur la carte physique d'*A. thaliana* a été définie pour chaque marqueur de la carte de liaison *Ab* x *Alp*.

2.4. Sélection des BACs portant les gènes candidats

Les clones BACs portant les gènes candidats ont été identifiés par hybridation moléculaire de type Southern sur des filtres construits à partir d'une banque BAC ordonnée représentant 4 équivalents génomes [13] et selon un protocole décrit dans cette publication. Les sondes ont été réalisées à partir de fragments amplifiés par PCR avec des amorces spécifiques à partir d'ADN génomique d'*A. balleri*. Les couples d'amorces spécifiques étaient: 5'-CAGCTTGCGTCGCCATTGTC-3' et 5'-CCAGCTAGTCTTTCACCTAC-3' pour *AbHMA4*, et 5'-CGAGTCTTCAATTTCTGCAACT-3' et 5'-AACTTTATTGATTTATTGTTAA-3' pour *AbMTP1*. L'ADN des BACs

présentant un signal d'hybridation a été extrait à l'aide du kit Nucléobond (Macherey Nagel), mais sans utiliser de colonne de purification. La présence dans le BAC d'une séquence révélée par hybridation a été confirmée par PCR et par séquençage (GATC Biotech). L'analyse des profils de restriction des clones BAC a été réalisée par digestion d'un à trois microgrammes d'ADN par différentes enzymes de restriction puis par électrophorèse sur gel d'agarose 0,8 % (p/v) en tampon TAE 0,5X (Tris 20 mM, acide acétique 10 mM, EDTA 0,5 mM). Pour l'hybridation moléculaire des profils de restriction, le gel a été traité pendant 15 minutes dans HCl 0,25N, puis rincé à l'eau. Les fragments d'ADN ont alors été transférés du gel vers des membranes chargées positivement (Hybond N+, Amersham) par capillarité en utilisant une solution de NaOH 0,4M. Les hybridations ont ensuite été réalisées selon le même protocole et avec les mêmes sondes qu'indiqué ci-dessus.

3. RÉSULTATS

3.1. Augmentation de la densité de la carte de liaison, réduction des intervalles des QTR et cartographie de gènes candidats potentiels

Vingt nouveaux marqueurs ont été ajoutés à la carte *Ab* x *Alp* publiée récemment [26]. La taille des groupes de liaison varie entre 53 et 80,2 cM pour une longueur de génome totale qui s'établit à 550 cM. Au total ce sont 83 marqueurs ancrés chez *A. thaliana* qui sont localisés sur la carte en moyenne tous les 6,6 cM (les espaces entre marqueurs allant de 0,4 à 23,5 cM).

Nous avons confirmé les trois QTL de tolérance au zinc définis précédemment [26] sur les groupes de liaison LG3, LG4 et LG6 mais la taille des intervalles de confiance a été considérablement réduite, passant de 24 cM, 4 cM et 13 cM à 5 cM, 3 cM et 12 cM pour Zntol-1, Zntol-2 and Zntol-3 respectivement. Sur les 23 marqueurs choisis pour leur rôle connu dans l'homéostasie des métaux et qui se positionnent à la position attendue sous l'hypothèse de sythénie entre génomes, trois se situent clairement dans les zones QTR définies et vingt se situent en dehors.

3.2. Analyse comparée de la carte de liaison (*A. halleri* x *A. l. petraea*) et du génome d'*A. thaliana*

L'analyse des positions des marqueurs ancrés sur la carte *Ab* x *Alp* confirme les cinq grands réarrangements génomiques qui ont accompagné le passage de 8 chromosomes observés chez la plupart des espèces du genre *Arabidopsis* à 5 chez *A. thaliana* [11], [18]. Trois fusions impliquant les

groupes de liaison LG1 et LG2, LG3 et LG4 et enfin LG7 et LG8 ont produit respectivement les chromosomes *At* Chr.I, *At* Chr.II et la partie basse du chromosome *At* Chr.V. Un événement de translocation se serait produit entre LG3 et LG5, produisant les bras supérieurs de *At* Chr.II et *At* Chr.III ; un second événement de duplication impliquant LG6 et LG7 aurait généré le chromosome *At* Chr.IV et le bras supérieur de *At* Chr.V. En plus de ces événements majeurs nous avons détecté trois inversions impliquant de petites régions (fig. 1) et deux événements de duplication détectés spécifiquement chez *A. halleri* qui concernent le marqueur 2-19245 qui correspond au gène *MTP1* décrit sous le nom de *ZAT* [21] et connu pour son rôle dans le transport du zinc à travers la membrane de la vacuole. Ce gène présent en copie unique chez *A. l. petraea* et *A. thaliana* présenterait chez *A. halleri*, au moins, trois copies —*MTP1-a*, *MTP1-b* and *MTP1-c*— localisées respectivement sur les groupes de liaison LG4, LG6 and LG1 [8], [26].

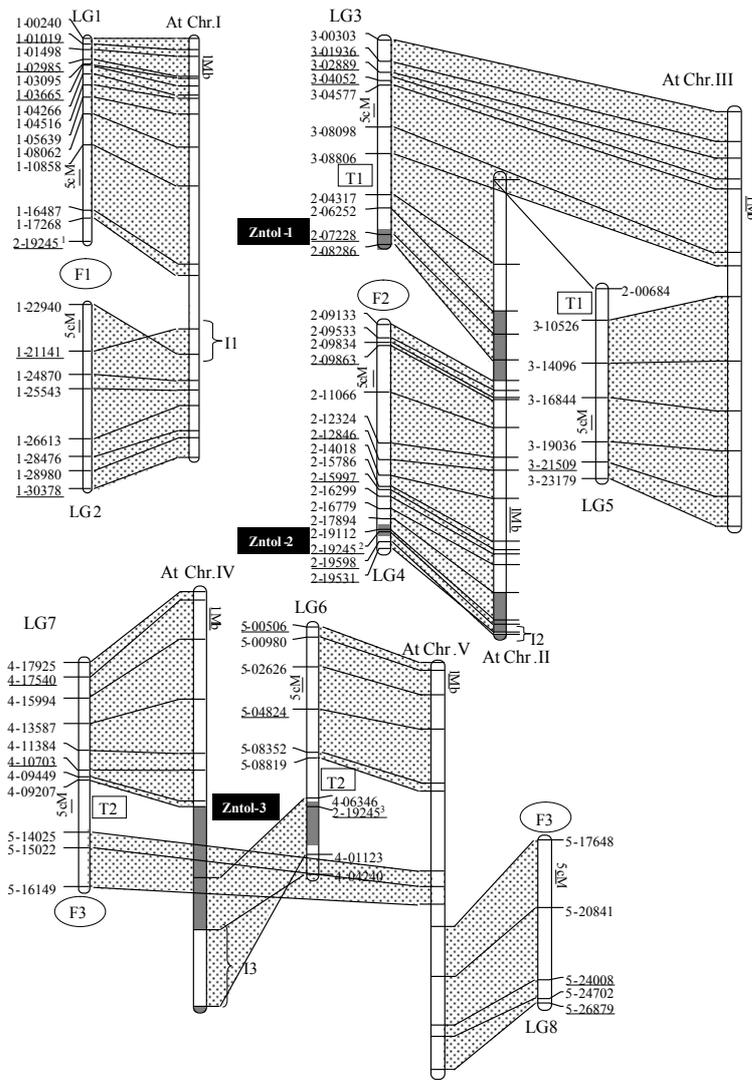


Figure 1 : Comparaison de la carte de liaison *Ab* × *Alp* (groupes de liaison LG1 à 8) et de la carte physique d'*A. thaliana* (chromosomes *At* Chr.I à V). T1 et T2 indiquent les points de translocations réciproques entre les groupes (LG3 LG5), et (LG6 LG7). F1, F2 et F3 indiquent les événements de fusion entre les groupes (LG1 LG2), (LG3 LG4), et (LG7 LG8). I1, I2 et I3 indiquent les inversions observées entre la carte de liaison *Ab* × *Alp* et la carte physique *A. thaliana*. Les intervalles de confiance des QTR de la tolérance au zinc (Zntol-1, Zntol-2 and Zntol-3) définis par l'analyse QTL à partir du croisement inter-spécifique entre *A. balleri* et *A. lyrata petraea* sont présentés ainsi que leur transposition sur la carte physique d'*A. thaliana*.

3.3. Transfert des intervalles QTR pour la tolérance au zinc de la carte (*A. halleri* x *A. l. petraea*) vers la carte d'*A. thaliana* et détection chez *A. thaliana* des gènes candidats de la tolérance au zinc d'*A. halleri*.

Grâce aux connaissances acquises au point précédent sur les zones de synthèse et de modification entre *A. halleri* et *A. thaliana*, nous avons pu transférer les bornes extérieures des intervalles de confiance des zones QTR de la tolérance au zinc définies sur la carte *Ab* x *Alp* sur la carte physique d'*A. thaliana* [18]. Le transfert sur le génome d'*A. thaliana* de la région contenant ZnTol1 nous permet, grâce aux données de génomique accumulée sur cette espèce modèle, d'identifier l'ensemble des gènes présents entre les bornes Chr 2-06252 (correspondant au gène référencé At2g14620) et Chr 2-09133 (correspondant au gène référencé At2g21320). Au total cette région contient 739 gènes parmi lesquels 11 ont présenté une expression différentielle entre *A. thaliana* et *A. halleri* dans les expériences de transcriptomique [25], [9]. Parmi ces 11 gènes, un seul exprime une protéine ayant un rôle avéré dans l'homéostasie des métaux (*HMA4* référencé At2g19110). Le gène *HMA4* peut être considéré comme candidat puisque chez *A. thaliana*, la protéine a été décrite comme appartenant à la famille des ATPases de type-P et serait impliquée dans la translocation du zinc de la racine vers les feuilles probablement *via* sa capacité à charger le xylème. De la même façon les gènes *MTP1-a* et *MTP1-b* exprimés spécifiquement dans la membrane de la vacuole [8] et décrits pour leur rôle potentiel dans la séquestration du zinc dans la vacuole apparaissent comme les candidats les plus pertinents dans les régions contenant ZnTol2 et ZnTol3.

3.4. Sélection de clones BAC portant les gènes candidats

Les clones BACs portant les gènes candidats *HMA4*, *MTP1-a* et *MTP1-b*, ont été recherchés dans une banque préalablement construite au laboratoire [13]. Les sondes utilisées pour les hybridations ont été réalisées à partir de fragments d'ADN amplifiés par PCR à partir d'ADN génomique d'*A. halleri* avec des amorces spécifiques. Une seule sonde a été utilisée pour identifier les BACs portant les gènes *MTP1-a* et *MTP1-b* car il n'existe qu'une seule séquence disponible du gène *MTP1*. Trois et huit BACs ont été identifiés puis validés par PCR avec les sondes *HMA4* et *MTP1*, respectivement (résultats non présentés). Une analyse des profils de restriction de ces BACs a permis de montrer que les trois BACs portant *HMA4* portaient des régions chevauchantes (fig. 2), suggérant qu'ils correspondent à un même locus. La situation est plus complexe pour *MTP1* puisqu'il existe trois gènes *MTP1*, localisés en des endroits différents sur la carte génétique [8], [26]. Au moins trois profils de restriction différents ont

été identifiés. Certains résultats d'hybridation suggèrent qu'il pourrait y avoir d'autres locus portant une copie du gène *MTP1* (l'étude est en cours).

Le BAC 14O3 portant le gène *AbHMA4* a été séquencé sur une longueur de 121 kb et sa séquence annotée ne révèle pas la présence d'autre protéine dont la fonction connue serait liée à l'homéostasie des métaux.

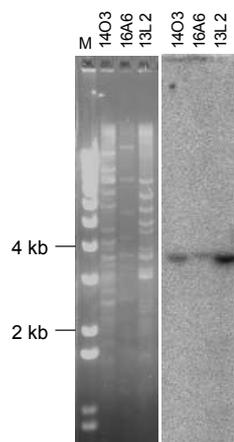


Figure 2 : Profil de restriction et hybridation moléculaire de type Southern des clones BAC portant le gène *AbHMA4*. L'enzyme de restriction utilisée est l'enzyme HindIII. La sonde a été réalisée à partir d'un fragment amplifié par PCR à partir d'ADN génomique d'*A. balleri* avec des amorces spécifiques d'*AbHMA4*. 14O3, 16A6 et 13L2 = noms des clones BAC. M= échelle de poids moléculaires.

4. DISCUSSION ET PERSPECTIVES

La comparaison de génomes d'espèces appartenant à un même genre voire à une même famille peut se révéler très intéressante non seulement pour aborder la question de l'évolution de la structure des génomes mais aussi pour détecter des gènes impliqués dans des mécanismes adaptatifs. L'identification de gènes candidats peut être conduite *via* la localisation de régions QTR responsables de caractéristiques écologiques pertinentes détectées chez des espèces non modèles, et leur transfert sur la carte physique d'une espèce modèle. Cependant avant qu'une telle approche ne puisse être appliquée de façon satisfaisante il faut pouvoir répondre à la question clé du niveau de macro- et micro-synténie entre espèces apparentées. Dans notre étude nous avons appliqué cette approche à la recherche des gènes candidats de la tolérance aux fortes concentrations de zinc dans les sols pollués à partir des

connaissances acquises sur le modèle *A. thaliana* et sur l'architecture génétique de la tolérance chez la pseudo-métallophyte *A. halleri*.

Nous avons tout d'abord démontré que des amorces d'amplification définies grâce à la séquence complète du génome d'*A. thaliana* peuvent être utilisées sur le génome de ses proches parentes *A. halleri* et *A. lyrata petraea*, en particulier pour densifier la carte génétique des espèces non modèles au niveau des régions d'intérêt ou pour le positionnement rapide de gènes dont la fonction connue pourrait avoir permis l'évolution du trait adaptatif.

Nous avons dans un second temps évalué la colinéarité de la carte de liaison $Ab \times Alp$ et de la carte physique d'*A. thaliana*. Tout en confirmant les cinq grands événements de réarrangement chromosomique ayant accompagné le passage de huit chromosomes à cinq, nous avons détecté trois remaniements mineurs (sous la forme de trois inversions). La première inversion (I1) déjà décrite dans la littérature différencie *A. thaliana* de toutes ses espèces sœurs et se serait produite chez *A. thaliana* depuis sa séparation des autres espèces du genre [4], [10]. La seconde inversion (I2) localisée sur LG4 n'a été décrite chez aucune autre espèce du genre *Arabidopsis*. Cependant la plus faible densité de marqueurs sur les cartes de liaison publiées d'*A. lyrata lyrata* et d'*A. lyrata petraea* pourrait expliquer la non détection de cette inversion et ne permet pas de conclure si cette inversion est spécifique à *A. halleri*. La troisième inversion (I3) localisée à l'extrémité inférieure du groupe LG6 de la carte $Ab \times Alp$ correspond à une double inversion déjà décrite dans la carte d'*A. lyrata petraea* [12]. On peut penser qu'il existe dans cette région des remaniements spécifiques à *A. halleri* puisque la zone Chr 4-05724 - Chr 4-06945 inclut la duplication à l'origine de *Ab-MTP1-b*.

Dans un troisième temps nous avons mis à profit les connaissances acquises entre génomes et nous avons démontré les possibilités offertes par le transfert des zones QTR déterminées chez une espèce possédant des caractéristiques écologiques intéressantes mais dont la structure du génome est peu connue vers une espèce modèle dont le génome est séquencé. Ce transfert a permis d'identifier un petit nombre de gènes candidats pouvant faire l'objet d'étude plus approfondies pour validation, en particulier *Ab-HMA4* qui co-localise avec *Zntol-1* et *Ab-MTP1* dont la copie orthologue (*Ab-MTP1-a*) co-localise avec *Zntol-2* alors qu'une copie paralogue co-localise avec *Zntol-3*. Cependant deux limites majeures de cette démarche doivent être mentionnées : la première concerne l'existence de réarrangements locaux qui peuvent entraîner un transfert des QTR sur des régions erronées et la seconde concerne le risque que des régions de génomes aient pu être perdues chez *A. thaliana* au moment des profonds réarrangements qui ont accompagné et suivi le passage de huit chromosomes à cinq ; la taille réduite du génome d'*A. thaliana* au regard de ses parentes sauvages [6] souligne ce risque de perte d'information chez le modèle *A. thaliana*. Dans le cas du BAC portant le gène *AbHMA4* nous avons constaté une très bonne

conservation de la structure physique du génome entre les deux espèces. En revanche le séquençage d'extrémités de BACs révèle souvent des similitudes faibles avec la séquence du génome d'*A. thaliana* (résultats non présentés) soulignant le grand intérêt que représente aujourd'hui le projet en cours de séquençage complet du génome d'*A. lyrata* (<http://www.jgi.doe.gov/sequencing/why/CSP2006/AlyrataCrubella.html>).

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Assuncao, A. G. L., H. Schat, M. G. M. Aarts. 2003. *Thlaspi caerulescens*, an attractive model species to study heavy metal hyperaccumulation in plants. *New Phytologist*, 159:351-360.
- [2] Basic, N., G. Besnard. 2006. Gene polymorphisms for elucidating the genetic structure of the heavy-metal hyperaccumulating trait in *Thlaspi caerulescens* and their cross-genera amplification in Brassicaceae. *Journal of Plant Research*, 119:479-487.
- [3] Becher, M., I. N. Talke, L. Krall, U. Kramer. 2004. Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *The Plant Journal*, 37:251-268.
- [4] Boivin, K., A. Acarkan, R. S. Mbulu, O. Clarenz, R. Schmidt. 2004. The *Arabidopsis* genome sequence as a tool for genome analysis in Brassicaceae. A comparison of the *Arabidopsis* and *Capsella rubella* genomes. *Plant Physiology*, 135:735-744.
- [5] Bradshaw, A. D. 1952. Populations of *Agrostis tenuis* resistant to lead and zinc poisoning. *Nature*, 169:1098.
- [6] Clauss, M. J., M. A. Koch. 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends in Plant Science*, 11:449-459.
- [7] Clemens, S., M. G. Palmgren, U. Krämer. 2002. A long way ahead: understanding and engineering plant metal accumulation. *Trends in Plant Science*, 7:309-315.
- [8] Dräger, D. B., A.-G. Desbrosses-Fonrouge, C. Krach, A. N. Chardonnens, R. C. Meyer, P. Saumitou-Laprade, U. Kramer. 2004. Two genes encoding *Arabidopsis halleri* MTP1 metal transport proteins co-segregate with zinc tolerance and account for high *MTP1* transcript levels. *The Plant Journal* 39:425-439.
- [9] Filatov, V., J. Dowdle, N. Smirnoff, B. Ford-Lloyd, H. J. Newbury, M. R. Macnair. 2007. A quantitative trait loci analysis of zinc hyperaccumulation in *Arabidopsis halleri*. *New Phytologist*, 174:580-590.
- [10] Hansson, B., A. Kawabe, S. Preuss, H. Kuittinen, D. Charlesworth. 2006. Comparative gene mapping in *Arabidopsis lyrata* chromosomes 1 and 2 and the corresponding *A. thaliana* chromosome 1: recombination rates, rearrangements and centromere location. *Genetical Research*, 87:75-85.
- [11] Koch, M., M. Kiefer. 2005. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species--

- Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. American Journal of Botany, 92:761-767.
- [12] Kuitinen, H., A. de Haan, C. Vogl, S. Oikarinen, J. Leppala, M. Koch, T. Mitchell-Olds, C. H. Langley, O. Savolainen. 2004. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics, 168:1575-1584.
- [13] Lacombe, E., M. Cossegal, M. Mirouze, T. Adam, F. Varoquaux, S. Loubet, P. Piffanelli, M. Lebrun, P. Berthomieu. 2008. Construction and characterization of a BAC library from *Arabidopsis halleri*: evaluation of physical mapping based on conserved synteny with *A. thaliana*. Plant Science, 174: 634-640.
- [14] Macnair, M. R., V. Bert, S. B. Huitson, P. Saumitou-Laprade, D. Petit. 1999. Zinc tolerance and hyperaccumulation are genetically independent characters. Proceeding of the Royal Society of London 30:1127-1136.
- [15] Pauwels, M., H. Frérot, I. Bonnin, P. Saumitou-Laprade. 2006. A broad-scale analysis of population differentiation for Zn-tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). Journal of Evolutionary Biology, 266:2175-2179.
- [16] Peer, W. A., M. Mamoudian, B. Lahner, R. D. Reeves, A. S. Murphy, D. E. Salt. 2003. Identifying model metal hyperaccumulating plants: germplasm analysis of 20 Brassicaceae accessions from a wide geographical area. New Phytologist, 159:421-430.
- [17] Reeves, R. D., C. Schwartz, J. Morel, J. Edmonson. 2001. Distribution and metal accumulating behavior of *Thlaspi caerulescens* and associated metallophytes in France. International Journal of Phytoremediation, 3:145-172.
- [18] Roosens, N., G. Willems, C. Godé, A. Courseaux, and P. Saumitou-Laprade. 2007. The use of comparative genome analysis and syntenic relationships allows extrapolating the position of Zn tolerance QTL regions from *Arabidopsis halleri* into *Arabidopsis thaliana*. Plant and Soil: in press.
- [19] Saumitou-Laprade, P., Y. Piquot, O. Raspé, J. Bernard, K. Vrieling. 1999. Plant DNA fingerprinting and profiling. in J. T. Epplen, and T. Lubjuhn, eds. DNA profiling and DNA fingerprinting. Birhäuser Verlag, Basel, p. 17-38.
- [20] Schat, H., and W. M. Ten Bookum, 1992. Genetic control of copper tolerance in *Silene vulgaris*. Heredity, 68, 219 - 229.
- [21] van der Zaal, B. J., L. W. Neuteboom, J. E. Pinas, A. N. Chardonens, H. Schat, J. A. Verkleij, P. J. Hooykaas. 1999. Overexpression of a novel Arabidopsis gene related to putative zinc-transporter genes from animals can lead to enhanced zinc resistance and accumulation. Plant Physiology, 119:1047-1055.
- [22] Van Ooijen, J. W., M. P. Boer, R. C. Jansen, C. Maliepaard. 2002. MapQTL 4.0, Software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, The Netherlands.
- [23] Voorrips, R. E. 2002. Software for the graphical presentation of linkage maps and QTLs. Heredity, 93:77-78.
- [24] Weber, M., E. Harada, C. Vess, E. Roepenack-Lahaye, S. Clemens. 2004. Comparative microarray analysis of *Arabidopsis thaliana* and *Arabidopsis halleri* roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. The Plant Journal, 37:269-281.
- [25] Weber, M., A. Trampczynska, S. Clemens. 2006. Comparative transcriptome analysis of toxic metal responses in *Arabidopsis thaliana* and the Cd²⁺-

- hypertolerant facultative metallophyte *Arabidopsis halleri*. *Plant, Cell and Environment*, 29:950-963.
- [26] Willems, G., D. B. Dräger, M. Courbot, C. Godé, N. Verbruggen, P. Saumitou-Laprade. 2007. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics*, 176:659-674.

La Cryobanque Nationale Lapin : un outil de gestion dynamique des ressources génétiques cunicoles

Pascal SALVETTI^{(1)*}, Daniel ALLAIN⁽²⁾, Gérard BOLET⁽²⁾,
Jacques HURTAUD⁽³⁾, Samuel BOUCHER⁽⁴⁾, Thierry JOLY⁽¹⁾

⁽¹⁾Université de Lyon, Unité CRYOBIO, UPSP ENVL/ISARA-Lyon, ISARA-Lyon,
1 avenue Bourgelat, 69280 Marcy L'Étoile, France

⁽²⁾INRA, Station d'Amélioration Génétique des Animaux, B.P. 52627,
31326 Castanet Tolosan Cedex, France

⁽³⁾Grimaud Frères Sélection SAS, La Corbière, 49450 Roussay, France

⁽⁴⁾Fédération Française de Cuniculture, Labovet conseil - BP 539,
85505 Les Herbiers, France

Abstract: The French Rabbit Cryobank: a tool for management of rabbit genetic resources. Cryobanking programs concerning rabbit species began in France during the nineties with first embryos freezing. Nowadays, about 18,000 frozen rabbit embryos were stored in liquid nitrogen in the French National Cryobank but these cryopreservation programs have not been used yet for a dynamic management of rabbit genetic resources. First, this project aimed to define and eliminate the restricting social factors related to cryobanking actions particularly for the private breeders who didn't want to provide their genetic resources. The situation has changed with the evolution of legal texts protecting their genetics and a sudden awareness of the advantages of cryopreservation methods. All parties involved in the French rabbit breeding are now ready to contribute actively to programs for the cryopreservation of rabbit genetic resources. Currently, the genetic resources of all protagonists are stored in the French National Cryobank and their motivations are very diverse: they use cryopreservation as an economic management tool for commercial strains, as a security in case of sanitary problems or physical destructions... In the future, it would be necessary to develop the communication with impassioned and amateur breeders who are not familiar with cryopreservation programs and who own the largest part of the rabbit population with high patrimonial value. Secondly, this project aimed to develop technical and genetic tools for the development of a dynamic and efficient management of all rabbit populations in order to always have stored in liquid nitrogen a representative sample of the rabbit population living in France. During this project, techniques of transfer and recovery of rabbit embryos by the cœlioscopic way were developed and evaluated as alternative solutions to the classical methods that cannot be used for high value rabbit strains. These methods allow cryopreserving, with the same efficiency as classical methods, threatened populations and specific strains with high

* Correspondance et tirés à part : psalvetti@isara.fr

economic and/or scientific value respecting ethical considerations. Finally, this study shows that cryopreservation and particularly embryos freezing is an important tool for the development of dynamic management strategies of cryopreserved genetic resources thanks to the concrete evaluation of the genetic progress in the population. The re-establishment of two populations of rabbit without genetic drift (one commercial strain and one threatened population) after more than ten years of storage in liquid nitrogen highlights the efficiency of embryo freezing and transfer methods. The comparison of these populations with their actual congeners shows that the evolution of phenotypic performances is important whatever the population considered. Further studies are needed to determine the refresh frequencies of the biological samples cryopreserved in the National Cryobank considering the reproductive characteristics of each rabbit population. In conclusion, we can consider that the French National Cryobank is now an efficient tool for the dynamic management of rabbit populations and the long term *ex situ* preservation of the genetic resources in this species.

cryopreservation/ genetic resource/ rabbit/ embryo/ cryobank

Résumé : Afin d'améliorer le fonctionnement de la Cryobanque Nationale lapin, le projet consistait à identifier les points de blocage d'ordre social et à développer des outils visant à surmonter les difficultés rencontrées d'ordre technique et génétique. Le développement des techniques de collecte et de transfert d'embryons sous cœlioscopie permet désormais d'étendre les actions de la Cryobanque Nationale à l'ensemble des populations. De plus, les verrous sociologiques semblent levés et l'ensemble des acteurs de la filière cunicole sont désormais impliqués dans la Cryobanque Nationale, convaincus de l'utilité et de l'efficacité de la cryoconservation des populations. Enfin, le projet a montré que la cryoconservation était un outil important pour la définition de stratégies d'actualisation des stocks de matériels biologiques congelés afin que ceux-ci soient toujours représentatifs des populations vivantes. Désormais, la Cryobanque Nationale constitue un outil efficace pour la gestion dynamique des populations de lapin et est devenue une réalité concrète pour la conservation *ex situ* à long terme des ressources génétiques cunicoles.

cryoconservation/ ressource génétique/ lapin/ embryon/ cryobanque

1. INTRODUCTION

Le lapin n'est pas concerné par la loi sur l'élevage du 31 décembre 1966, tout comme les espèces avicoles et aquicoles. Par conséquent, la filière cunicole n'a pas pu bénéficier d'un système unique d'identification individuel des animaux au niveau national. Ainsi, la quantité d'informations sur les généalogies des reproducteurs et la description des populations est extrêmement variable suivant les acteurs engagés dans la gestion des ressources génétiques cunicoles. Cette situation particulière peut s'expliquer par les caractéristiques zootechniques remarquables du lapin dont on peut souligner la maturité sexuelle précoce (~16 à 17 semaines), durée de

gestation faible (~31 jours), taille de portée importante (6 à 12 lapereaux par mise bas) et un intervalle entre générations court par rapport aux autres espèces domestiques. De plus, l'espèce cunicole se distingue aussi par son statut sanitaire particulier au regard des recommandations sanitaires de l'Office International des Epizooties (OIE) qui ne recense actuellement aucune épizootie majeure à inscrire dans le livre A.

Le lapin est une espèce animale qui présente une très large diversité de populations liée à son histoire et à son statut particulier dans notre société. En effet, le lapin peut être trouvé à l'état sauvage (lapin de garenne) ou domestiqué principalement en tant que lapin d'élevage pour la production de viande ou de fourrure. De plus, le lapin est de plus en plus plébiscité en tant qu'animal de compagnie ou encore comme animal de laboratoire pour la recherche biomédicale et la production de protéines thérapeutiques pour l'industrie pharmaceutique. La conservation des ressources génétiques chez cette espèce concerne donc une diversité expérimentale mais aussi une diversité zootechnique avec une large dimension patrimoniale.

La congélation des embryons au stade morulae compactée est la principale voie de cryoconservation à long terme des ressources génétiques chez le lapin [9], la congélation de la semence n'étant pas totalement maîtrisée chez cette espèce est moins avantageuse. Désormais, cette méthode éprouvée est précisément caractérisée par des protocoles standardisés définis dans le cadre d'une démarche qualité (Action CRB 2001). Actuellement, la Cryobanque Nationale renferme un stock de près de 18 000 embryons congelés issus de 50 populations de lapins différentes. En fonction des objectifs recherchés et de l'originalité génétique de l'échantillon biologique cryoconservé, trois grands types de matériel ont été définis (type I, II et III) selon la classification générale proposée par la Cryobanque Nationale [6]. Les populations de lapins de types I et II sont peu représentées dans la Cryobanque Nationale avec seulement 5 % et 12 % des embryons congelés par rapport à des populations de type III. Depuis leur mise en place, les programmes de sauvegarde des ressources génétiques lapin ont permis de mener des actions ponctuelles de cryoconservation visant à sauvegarder une population à un instant donné. Cependant, une population ne représente pas une diversité génétique figée. Qu'elles soient soumises aux schémas de sélection précis des sélectionneurs ou simplement entretenues chez des éleveurs amateurs, les populations continuent d'évoluer selon un rythme qui est propre à chaque situation. La question qui se pose désormais est : comment suivre la représentativité de ce matériel biologique stocké en regard des populations vivantes ?

L'objectif de ce projet est de faire de la Cryobanque Nationale un outil de gestion dynamique des ressources génétiques lapin (RGL) complémentaire aux opérations de conservation *in situ* menées par les

acteurs de la filière cunicole. Ainsi, ce projet transdisciplinaire a eu pour objectif de mobiliser l'ensemble des gestionnaires de ressources génétiques cunicoles, d'appliquer les derniers outils de génétique et de simplifier les techniques de cryoconservation.

2. LA CRYOBANQUE NATIONALE: DES ACTEURS IMPLIQUÉS

La conservation et le transfert des ressources génétiques sous forme de matériel biologique congelé a posé beaucoup de difficultés dans la définition des cadres juridiques, notamment en ce qui concerne les notions de propriété intellectuelle [14]. La cryoconservation des RGL n'a pas échappé à ce flou juridique qui s'est traduit par un blocage sociologique des différents acteurs de la filière cunicole vis-à-vis de la Cryobanque Nationale. Initialement, les difficultés rencontrées portaient essentiellement sur les incertitudes des déposants par rapport au contrôle des sorties et de l'utilisation des matériels biologiques déposés à la cryobanque. Ce frein socio/économique concernait principalement les sélectionneurs privés de lapins qui restaient circonspects à l'idée de remettre leurs propres ressources génétiques à la Cryobanque Nationale. Afin de lever ces blocages, des enquêtes de terrain ont été menées pour caractériser les motivations de l'ensemble des acteurs de la filière cunicole, pour identifier les freins socio-économiques et culturels des sélectionneurs privés et pour saisir la perception de l'utilité de la cryobanque pour cette espèce. Ces études ont été réalisées auprès d'un échantillon d'acteurs représentatifs (sélectionneurs privés et publics, éleveurs amateurs et professionnels...) en privilégiant la méthode d'entretien semi directif susceptible de dégager avec suffisamment de spontanéité les représentations des différents acteurs. L'échantillon enquêté lors de ce projet s'est voulu suffisamment représentatif de la filière et comprenait deux des trois sélectionneurs privés (Grimaud Frères Sélection SAS, Hycole), l'INRA (centre de Toulouse), des structures développant des lapins transgéniques (Bioprotein, Charles River, INRA de Jouy en Josas) et aussi quelques éleveurs de races anciennes de la Fédération Française de Cuniculture (FFC). Aujourd'hui, l'évolution des statuts de la Cryobanque Nationale, le contexte d'épizooties récurrentes et le développement des collaborations sont autant de facteurs qui ont permis une implication de la quasi-totalité des acteurs dans les activités de la Cryobanque. Ainsi, nous avons pu arriver à une typologie des différents gestionnaires de la Cryobanque présentés selon l'importance de leur implication dans les opérations de sauvegarde du patrimoine génétique lapin et leurs motivations vis-à-vis de la Cryobanque Nationale.

Les acteurs de la recherche publique sont des acteurs impliqués dans le développement de la cryoconservation de la diversité génétique chez le

lapin. Ceux-ci regroupent principalement des structures en lien étroit avec la recherche fondamentale ou l'industrie pharmaceutique dont la principale motivation est de disposer d'outil de gestion de leurs souches. En effet, les populations de lapin concernées par ces acteurs sont plutôt de type II, comprenant les animaux de laboratoire modèles pour la recherche ayant une haute valeur scientifique ou encore les souches transgéniques qui sont en plein essor, notamment pour la production de protéines thérapeutiques par l'industrie pharmaceutique. La cryobanque représente un formidable outil de gestion qui permet la conservation sûre, efficace et à moindre coût de populations précieuses non utilisées à un instant donné. La congélation d'embryons leur permet aussi de garantir la conservation et les échanges de leurs populations en cas de problème sanitaire ou d'opérations commando menées par les groupes de libération des animaux, très en vogue actuellement.

La seconde catégorie correspond aux principaux acteurs de l'amélioration et de la sélection cunicole en France. Elle comprend à la fois des acteurs privés (Grimaud Frères Sélection SAS, Hycôle, Eurolap) responsables de la sélection des voies mâles et femelles et un acteur public, l'INRA (Station d'Amélioration Génétique des Animaux, centre de Toulouse) en charge de la création de lignées originales (lignées de chair ou lignées productrices de phanères). Pour ces acteurs, la conservation de ces ressources génétiques d'une grande valeur économique représente avant tout une sécurité importante pour les sélectionneurs face à un risque sanitaire éventuel qui pourrait entraîner l'abattage de cheptels entiers. De plus, la cryoconservation de certains noyaux de sélection permet aux filières de réagir rapidement face à une réorientation des schémas de sélection selon la demande des consommateurs ; il devient alors possible de faire « marche arrière » pour répondre aux nouveaux besoins entraînant une segmentation des marchés. Aujourd'hui, certaines lignées de lapin sont complètement éteintes et ne subsistent que sous forme d'embryons congelés en attendant une éventuelle perspective de réutilisation (populations Dutch Petit Russe, Chinchilla 1229, INRA 2666...). Enfin, la cryoconservation permet de conserver indéfiniment des populations sélectionnées à un instant donné qui pourront servir de populations témoins lors d'expériences de sélection.

La dernière catégorie d'acteurs recensés représente les éleveurs sélectionneurs amateurs intégrés dans des clubs de races et regroupés au sein de la FFC ainsi que les éleveurs amateurs possédant des élevages familiaux. On y trouve généralement des populations à grands effectifs (Fauve de Bourgogne, Papillon...) et des populations à petit effectif menacées de disparition (Brun Marron de Lorraine...) qui sont toutes bien identifiées dans le Standard Officiel édité par la FFC [8]. Pour ces acteurs, les motivations de la cryoconservation sont principalement de pérenniser la

race sur le long terme tout en assurant la possibilité de reproduire leurs animaux en race pure et de limiter l'accroissement de la consanguinité. La diminution du nombre d'éleveurs amateurs et d'élevages familiaux liée aux nouvelles réglementations ainsi que l'évolution des critères d'évaluation lors des concours de races accentuent l'importance des actions de conservation des populations de cette catégorie. L'implication de la FFC est importante mais repose sur la volonté individuelle des clubs de races à participer aux opérations de la Cryobanque Nationale. Pour cette catégorie d'acteurs, il existe un important besoin de communication de la FFC envers les différents éleveurs ou club de races ayant un fonctionnement indépendant et autonome. Grâce à une collaboration étroite entre la FFC, les éleveurs sélectionneurs amateurs et les acteurs de la Recherche, une opération de communication a pu être mise en place au Salon International de l'Agriculture (Paris, 2007) sur la recréation d'une population de « Brun Marron de Lorraine » à partir d'embryons stockés 14 ans dans l'azote liquide [12]. Au total, 134 embryons Brun Marron de Lorraine qui avaient été congelés en 1992, ont été décongelés et transférés en novembre 2006 dans 14 femelles receveuses pour donner naissance à 69 lapereaux vivants (51,5 % de taux de survie embryonnaire). L'expertise de juges cunicoles officiels sur des critères d'évaluation classiques (évaluation de la morphologie et des phanères optiques) a montré une évolution significative de la race par comparaison avec leurs congénères actuels qui ont continué d'évoluer sous la pression de sélection appliquée par les éleveurs [4].

Actuellement, l'ensemble des acteurs de la filière cunicole s'implique dans les actions de cryoconservation menées par la Cryobanque Nationale avec des motivations diverses et variées. A l'avenir, il est probable que les acteurs de la recherche et les sélectionneurs soient de plus en plus demandeurs d'opérations de cryoconservation tant les enjeux sont importants pour le développement de leurs activités. Cependant, la cryoconservation des populations sauvages ou des races standardisées par la FFC posera plus de problèmes. Le BRG et la FFC devront poursuivre leurs actions de communication envers les éleveurs sur l'importance de la sauvegarde des ressources génétiques lapin. Face aux difficultés de repérer les populations originales, il sera nécessaire de développer des outils concrets pour identifier efficacement les individus de manière complémentaire à la simple évaluation phénotypique.

3. LA CRYOBANQUE NATIONALE : UN OUTIL EFFICACE POUR LA GESTION DE LA DIVERSITÉ GÉNÉTIQUE ET DE LA SÉLECTION DU LAPIN ?

Au cours de ce projet, nous avons essayé de voir si la Cryobanque Nationale était un outil efficace de gestion dynamique de la sélection et de la diversité génétique chez le lapin à travers la comparaison objective de deux populations de lapins issues d'un même génotype. Grâce à la congélation des embryons, une population d'origine a pu être comparée à la même population après 10 générations de sélection en mesurant le progrès génétique réalisé (observations phénotypiques et caractérisations zootechniques). De plus, l'établissement de la carte génétique lapin [5] a permis de développer de nouveaux outils de génomique permettant de mesurer la diversité génétique de chaque population.

Après un échantillonnage représentatif de l'ensemble de la diversité génétique présente, la population Orylag® Castor Rex producteurs de fourrures (type III) a été stockée sous forme d'embryons congelés en 1995. Le stock était constitué de 391 embryons issus de 64 femelles donneuses accouplées à 15 mâles différents appartenant à 10 groupes de reproduction différents. Après décongélation de ce stock en juin 2004, 347 embryons de bonne qualité ont été transférés chez 40 femelles receveuses dont 20 ont donné naissance à un total de 57 lapereaux vivants. Après la phase d'élevage de ces lapereaux jusqu'au stade adulte, la population issue du stock d'embryons congelés en 1995 était constituée de 13 femelles et 18 mâles représentant 7 groupes de reproduction. Ces animaux ont été mis à la reproduction selon le même principe de gestion en groupes de reproduction que celui en vigueur dans la population initiale de 1995 en vue de produire un minimum de 100 descendants dont les performances phénotypiques et la valeur génétique ont été comparées en 2005 à celles d'animaux contemporains de la population en sélection.

La diversité génétique des trois populations (populations de 1995, 2005 et population 1995 reconstituée) a pu être caractérisée par le calcul des probabilités d'origine des gènes et par l'analyse du polymorphisme moléculaire. Les probabilités d'origine des gènes ont été calculées à l'aide du pack pedigree [2], [3], pour chacune de ces 3 populations. De plus, l'analyse du polymorphisme moléculaire a été effectuée pour un échantillon de 74 animaux de la population 1995 reconstituée et 61 animaux de la souche en sélection en 2005. Ces échantillons ont été génotypés pour 35 marqueurs moléculaires microsatellites indépendants choisis parmi le panel disponible identifié lors de l'établissement de la carte génétique du lapin. Les variations de fréquences alléliques entre les 2 populations ont été analysées selon la méthodologie des signatures de sélection proposée par Laval *et al.* [10] chez

la poule consistant à évaluer l'évolution des distances entre les différents marqueurs moléculaires choisis.

Le tableau I résume la probabilité d'origine des gènes dans les populations en sélection en 1995 et 2005 et dans la population issue du stock d'embryons congelés en 1995. Les nombres de fondateurs, fondateurs efficaces et ancêtres majeurs contribuant à l'origine des gènes sont très similaires dans les 3 populations. Ils sont identiques dans les cohortes 1995 et 2005 de la population en sélection : 36 fondateurs dont 16,8 fondateurs efficaces et 12 ancêtres majeurs communs contribuent à 75 % de l'origine des gènes, malgré un accident sanitaire ayant entraîné la perte de 3 groupes de reproduction, survenu en 2002 dans la population en sélection. La méthode de gestion de la population en groupes de reproduction utilisée permet d'expliquer ce maintien de la variabilité génétique dans cette population en sélection. Dans la population issue du stock d'embryons congelés en 1995, une très faible réduction de la diversité génétique est observée : 34 fondateurs dont 15,9 fondateurs efficaces, soit une perte de seulement 0,9 fondateur efficace malgré un faible taux de réussite lors de la décongélation et du transfert d'embryons. Sous condition d'un bon échantillonnage de la population lors de la mise en conservation, ce résultat montre l'efficacité de l'outil cryobanque pour la conservation et le maintien de la diversité génétique en vue de reconstituer des populations en sélection. La constitution d'un stock de 350 embryons congelés est suffisant y compris en cas d'un faible taux de réussite du transfert d'embryons: taux de développement global (nombre de lapereaux nés/nombre d'embryons décongelés) de 16,4 % au lieu de près de 45 % en moyenne [12], [13].

Tableau I : Nombre de fondateurs, fondateurs efficaces et ancêtres majeurs ayant une contribution totale de 75 % des gènes.

		Population 1995 issue des embryons congelés	Cohorte 1995 population sélectionnée	Cohorte 2005 population sélectionnée
Mâles	Nombre de fondateurs	34	36	36
	Dont efficaces	15,9	16,8	16,8
	Ancêtres majeurs contribuant 75 % gènes	10	12	12
Femelles	Nombre de fondateurs	34	36	36
	Dont efficaces	15,7	16,8	16,8
	Ancêtres majeurs contribuant 75 % gènes	8	12	12

L'utilisation de marqueurs moléculaires polymorphes n'a pas permis de mettre en évidence une dérive génétique ni de détecter des signatures de sélection en raison du faible polymorphisme moléculaire observé dans

l'espèce lapin au niveau des marqueurs microsatellites actuellement disponibles. Dans cette étude, l'ensemble des animaux était homozygote pour 20 % des marqueurs et le taux d'homozygotie moyen était de 67 % environ dans chacune des 2 populations étudiées. La mise à disposition prochaine de la séquence haute densité du génome du lapin et d'un important panel de marqueurs SNP devrait permettre de reprendre cette étude.

Le progrès génétique a été mesuré entre les animaux descendants du stock d'embryons congelés en 1995 et décongelés en 2004 et un échantillon d'animaux contemporains de la souche sélectionnée en 2005. Cette comparaison a porté sur le poids à 8 semaines, les différentes caractéristiques de qualité de la fourrure à 16 semaines (jarrosité, maturité et compacité de la fourrure, hauteur du pelage [15]) et d'autres caractéristiques non mesurées en routine dans la population en sélection (densité de follicules pileux, couleur de la fourrure, qualité globale des fourrures). La mesure des performances phénotypiques (tabl. II) dans les deux populations étudiées montre qu'un progrès génétique important a été réalisé dans cette population de lapin Rex sélectionnée pour la production de fourrures entre 1995 et 2005 (environ 10 générations de sélection). Ce résultat était évidemment attendu en ce qui concerne la réponse directe sur les caractères sélectionnés. Mais l'outil cryobanque permet aussi de réaliser ponctuellement des mesures de la réponse indirecte sur des caractères d'intérêt non mesurables directement ou difficilement mesurables en routine en sélection et ainsi apprécier toute évolution favorable ou défavorable de différents caractères. Ainsi nous avons mis en évidence un gain significatif sur la densité de follicules pileux qui détermine la densité de la fourrure, un caractère long et fastidieux à mesurer en routine, ainsi qu'une amélioration sensible de la qualité globale de la fourrure (plus de peaux royales et primes) appréciée exclusivement sur peaux sèches brutes après le sacrifice des animaux. Par contre, cet outil de gestion dynamique de la sélection nous a aussi permis de mettre en évidence l'apparition plus fréquente d'une cause de déclassement des fourrures pour un défaut de maturité particulier.

Les résultats obtenus montrent que la congélation d'embryons représente un outil de gestion dynamique efficace de la diversité génétique et de la sélection chez le lapin. Elle permet la conservation de la diversité génétique sous réserve d'une gestion maîtrisée de la variabilité génétique et d'un bon échantillonnage préalable des populations mises en conservation, y compris en situation d'un faible taux de réussite de la décongélation et du transfert d'embryons. De plus, ces résultats montrent qu'il est nécessaire de développer des stratégies d'actualisation des stocks de matériel génétique afin que ceux-ci soient toujours représentatifs des populations existantes. Cependant, dans notre projet, l'analyse du polymorphisme moléculaire n'a

pas apporté d'éléments nouveaux pour la mesure de la diversité génétique des populations étudiées.

Tableau II : Performances phénotypiques et gain génétique sur les caractères sélectionnés des 2 populations animales (lapin Rex).

		Population issue des embryons congelés en 1995 (n=87)	Cohorte 2005 Population sélectionnée (n=113)	Gain phénotypique (%)	Gain génétique (σ_g)
Caractères sélectionnés	Poids à 8 semaines (g)	1084.0	1304.1	+20.0%	+0.80
	Note de jarrosité	44.0	31.9	+27.5%	+1.94
	Hauteur du pelage	18.9	19.8	+4.4%	+1.00
	Compacité fourrure	36.3	38.7	+6.6%	+0.65
	Maturité fourrure	5.7	3.6	+35.5%	+0.77
Autres caractères mesurés	Densité follicules	32.7	39.8	+21.8%	
	Qualité globale fourrure				
	% peaux Royales	13.3	20.9	+57.1%	
	% peaux Primes	43.4	53.2	+22.5%	
	% peaux Autres	38.8	27.0	-28.6%	

4. ÉVOLUTION DES TECHNIQUES DE SAUVEGARDE POUR LA CRYOBANQUE NATIONALE

La plupart des embryons stockés dans la cryobanque sont issus de populations de type III et ont été collectés après abattage systématique des femelles pour des raisons pratiques et techniques. Dans cette situation, la stratégie d'échantillonnage prend en compte l'ensemble de la diversité de la population et non pas un seul individu, l'échantillon à conserver étant facilement réajusté par abattage d'autres femelles. Cependant, l'euthanasie des femelles donneuses d'embryons constitue un obstacle à la collecte des populations à petit effectif (type I) ou des génotypes particuliers (type II). Aussi les populations de types I et II sont-elles très peu représentées dans la Cryobanque Nationale. Il est donc urgent d'intégrer de nouvelles méthodes de collecte et de transfert d'embryons qui soient non destructrices, peu invasives et répétables. La laparotomie a été développée chez la lapine de manière à éviter l'euthanasie des animaux lors des opérations de collecte et de transfert d'embryons et permet l'accès au tractus génital après incision de l'abdomen de la lapine. Cependant, cette méthode reste invasive et difficilement répétable sur une même femelle à cause du développement post-opératoire d'adhérences fibreuses liées à l'intervention chirurgicale. Plus récemment, la cœlioscopie a été développée pour accéder aux organes

génitaux de la femelle de manière atraumatique permettant aussi bien la collecte que le transfert d'embryons [1].

Nous avons éprouvé concrètement les méthodes de collecte et de transfert d'embryons par cœlioscopie en conditions de terrain afin d'en évaluer les avantages et les limites. D'un point de vue technique, les différentes opérations menées ont montré que la collecte d'embryons par cœlioscopie était facilement réalisable dans les élevages avec des taux de collecte proches de ceux obtenus jusqu'à maintenant après euthanasie des femelles (~70 à 80 %). De la même façon, les résultats de transfert embryonnaire dans les oviductes par cœlioscopie sont équivalents à ceux obtenus habituellement par laparotomie et transfert dans les cornes utérines (taux de survie embryonnaires de l'ordre de 65 % avec des embryons frais et de 40 % avec des embryons congelés). L'avantage majeur de la cœlioscopie est qu'elle est très peu invasive, permettant de s'affranchir de l'abattage des femelles lors des collectes d'embryons mais aussi d'avoir un très bon comportement des femelles transférées (aucun risque d'éventration). De ce fait, ces méthodes offrent une possibilité de répétition des collectes et des transferts sur une même femelle sans aucune complication post-chirurgicale et dans le respect du bien être animal.

La principale limite de ces techniques est leur coût (technique, main d'œuvre, déplacement) principalement pour les opérations de collecte d'embryons. Le tableau III montre une comparaison des coûts de production et de transfert par cœlioscopie et par les méthodes classiques. Le principal inconvénient de la collecte d'embryons par cœlioscopie par rapport à une collecte classique est son coût de main d'œuvre qui est beaucoup plus élevé. De ce fait, cette avancée technologique est pour l'instant réservée aux populations précieuses de types I ou II pour lesquelles la survie de l'animal est primordiale. Contrairement à la collecte, le transfert des embryons sous cœlioscopie est amené à remplacer progressivement la technique de la laparotomie car il présente les avantages d'être plus rapide, moins coûteux et surtout moins invasif (absence de sutures et peu de délabrement tissulaire) réduisant ainsi les risques d'éventration et de perte des embryons transférés.

La Cryobanque Nationale cunicole possède désormais des outils supplémentaires indispensables pour le développement des opérations de cryoconservation principalement pour les populations de types I et II et satisfaisant entièrement aux nouvelles règles de bioéthique.

Tableau III : Comparaison des coûts de production et de transfert hors taxes d'un embryon de lapin par cœlioscopie et par les méthodes classiques.

		Collecte d'embryons		Transfert d'embryons	
		Euthanasie	Cœlioscopie	Laparotomie	Cœlioscopie
Investissement matériel		3000 €	10000 €	3000 €	9000 €
Coûts techniques	Consommables/♀ (hormones, fluides, matériels usage unique)	6,2 €	15,8 €	13,0 €	9,0 €
	Coûts de production / transfert d'un embryon*	0,3 €	0,8 €	1,3 €	0,9 €
Coûts main d'œuvre	Main d'œuvre/♀	5 minutes à 1 personne	30 minutes à 2 personnes	40 minutes à 3 personnes	10 minutes à 3 personnes
	Rendement journalier	60 ♀	10 ♀	12 ♀	48 ♀

*Coûts de production estimés à partir d'une femelle superovulée produisant 20 embryons de bonne qualité hors coûts de congélation et valeur de la lapine donneuse. Coûts de transfert estimés à partir du transfert de 10 embryons congelés/décongelés dans une femelle receveuse hors valeur de la lapine receveuse.

5. CONCLUSION

Ce projet mené conjointement par l'ISARA-Lyon, l'INRA, la FFC et GRIMAUD FRERES SELECTION a permis d'optimiser les actions de conservation *ex situ* des RGL en France. Dorénavant, l'ensemble des acteurs de la filière est impliqué dans la Cryobanque Nationale pour des motivations diverses et variées y compris pour les éleveurs amateurs pour qui d'importants efforts de communication sont à envisager. De plus, la méthode de congélation d'embryons a pu être éprouvée concrètement et avec succès après une longue période de stockage du matériel biologique. Sous réserve d'un bon échantillonnage préalable des populations cryoconservées, cet outil permet une gestion efficace de la diversité génétique cunicole même lorsque les taux de survie des embryons sont faibles. Enfin, les techniques endoscopiques permettent désormais d'étendre la congélation d'embryons aux populations de type I et II pour lesquelles la cryoconservation est une nécessité et qui sont encore trop peu représentées au sein de la Cryobanque Nationale.

Cependant le fonctionnement de la Cryobanque Nationale et les stratégies de sauvegarde des RGL doivent encore être optimisés notamment par le développement de nouvelles voies de cryoconservation des gamètes permettant la sauvegarde de la génétique d'un seul individu notamment dans des situations d'urgence [7], [11]. Enfin, nous avons observé qu'après 10 générations de stockage dans l'azote liquide, les matériels biologiques ne sont plus représentatifs des populations contemporaines. A l'avenir, le défi le plus important reposera sur la définition de stratégies d'actualisation des

stocks du matériel biologique qui seront propres à chaque population, de manière à toujours garantir la représentativité des populations vivantes.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Besenfelder U., Strouhal C., Brem G., 1998. A method for endoscopic embryo collection and transfer in the rabbit. *Zentralbl Veterinarmed A*, 45 (9): 577-579.
- [2] Boichard D., Maignel L., Verrier E., 1995. Estimation of the effective number of founders of a population from pedigree information. *2nd European workshop on Advanced Biometrical Methods in Animal Breeding*, Salzburg, Austria, June 12-20, 1995.
- [3] Boichard D., 2007. PEDIG: A fortran package for pedigree analysis suited for large populations.
- [4] Boucher S., Menigoz J.J., Salvetti P., Arnold J., Hurtaud J., Joly T., 2007. Comparaison phénotypique simultanée *in vivo* d'une souche de lapins de race Brun marron de Lorraine cryopréservés en 1992 et des descendants actuels. *12^{mes} Journées de la Recherche Cunicole*, 27-28 novembre 2007, Le Mans, pp.137-140.
- [5] Chantry-Darmon C., Urien C., de Rochambeau H., Allain D., Peña B., Hayes H., Grohs C., Cribiu E. P., Deretz-Picoulet S., Larzul C., Save J. C., Neau A., Chardon P., Rogel-Gaillard C., 2006. A first-generation microsatellite-based integrated genetic and cytogenetic map for the European rabbit (*Oryctolagus cuniculus*) and localization of angora and albino. *Animal Genetics*, 37(4): 335-341
- [6] Danchin-Burge C., Bibé B., Planchenault D., 2003. La Cryobanque Nationale: une cryo-collection patrimoniale des races animales d'élevage. *10^e Journées de la Recherche Cunicole*, 19-20 novembre 2003, Paris, pp.35-38.
- [7] Daniel N., Chesné P., Baratte M., Renard J.P., 2007. Lapins produits par injection intra-cytoplasmique de spermatozoïdes (ICSI) stockés à température ambiante. *12^e Journées de la Recherche Cunicole*, 27-28 novembre 2007, Le Mans, pp.33-36.
- [8] Fédération Française de Cuniculture, 2000. Les lapins de race: spécificités zoologiques, standards officiels. Fédération Française de Cuniculture, Paris, 288 p.
- [9] Joly T., de Rochambeau H., Renard J.P., 1998. Etablissement d'une cryobanque d'embryons pour la conservation *ex situ* de la diversité génétique chez le lapin: aspects pratiques. *Genet. Sel. Evol.* 30 (Suppl. 1) (1998) S259-S269.
- [10] Laval G., Chevalet C., Pitel F., Amigues Y., Bordas A., Coville J.L., Tixier-Boichard M., 2003. Modélisation et évaluation expérimentale de la diversité génétique neutre ou adaptative dans le cadre de lignées sélectionnées chez la poule. *Les Actes du BRG*, 4: 35-47.
- [11] Neto V., Guérin P., Salvetti P., Lefranc A.C., Buff S., Joly T., 2007. Autogreffé du tissu ovarien cryoconservé : bilan et perspectives d'utilisations. *12^e Journées de la Recherche Cunicole*, 27-28 novembre 2007, Le Mans, pp.29-32.
- [12] Salvetti P., Joly T., Boucher S., Hurtaud J., Renard J.P., 2007. Viability of rabbit embryos after 15 years storage in liquid nitrogen. *44th Annual Meeting of the Society for Cryobiology CRYO 2007*, 28th July to 1st August, Lake Louise (Canada). (Poster).

- [13] Salvetti P., Neto V., Theau-Clément M., Guérin P., Beckers J.F., Joly T., 2007. Effect of LH on embryo production in superovulated rabbit does. *Theriogenology*, 67: 1185-1193.
- [14] Sontot A., 2005. Conservation, utilisation et échange des ressources génétiques: droits et pratiques. *Les Actes du BRG*, 5: 615-627
- [15] Thebault R.G., Allain D., de Rochambeau H., Vrillon J.L., 2000. Selection scheme and genetic improvement of Orylag for fur production. 7th International Congress in Fur Animal Production, Scientifur, Kastoria, Macedonia, Greece, 13-15 September.

Déperdition ancienne et restauration moderne de la diversité génétique du blé

Anne-Céline THUILLET⁽¹⁾, Pierre ROUMET⁽²⁾, Amandine BORDAT⁽²⁾,
Christine TOLLON⁽²⁾, Gérard POUX⁽²⁾, Sylvain SANTONI⁽²⁾,
Jacques DAVID^{(2)*}

⁽¹⁾INRA-IRD-Supagro, UMR Diversité et Adaptation des plantes Cultivées 1097
911 avenue d'Agropolis, BP 64501, 34394 Montpellier Cedex 5, France
⁽²⁾Centre de Montpellier, 2 place Viala, 34070 Montpellier, France

Abstract: Loss and restoration of wheat genetic diversity. Durum wheat and bread wheat are both cultivated species that evolved from an initial domestication event and through several selection steps to fit agricultural requirements. Through their history, they underwent an overall reduction of diversity. The decrease of diversity was more dramatic in genomic regions involved in adaptive traits for agriculture. A good example in wheat is a genomic region corresponding to a QTL around the *Tg2* gene, located on the short arm of the chromosome 2B, and involved in grain threshability. The objective of this study is to evaluate two methods allowing genetic diversity reintroduction. One of them consists in a classical genealogical selection after wide crosses. The other one allows outbreeding through the use of a male sterility gene. Genitors of these two populations include current cultivated wheat varieties and diverse wheat ancestors. We used nine microsatellite loci randomly chosen on the genome and seven microsatellite loci on chromosome 2B short arm, including four loci in the QTL interval, to genotype a representative collection of wheat evolution and a sample from each restoration method. Effective population sizes estimation allows a representation of diversity that reflects only drift and selective effects and insensitive to markers mutation rates. Based on these estimates, we confirm the decrease of genetic diversity through wheat history. We estimate an effective population size of 27,354 in the wild ancestor, we find 3,898 for the domesticated form, 2,514 for landraces of durum wheat, and 917 for current varieties of durum wheat. We also confirm an accentuated reduction of diversity in the 2B genomic region from *T. t. dicoccum* towards all threshable forms. We show that both restoration methods permit a similar increase of genetic diversity globally but that the outbreeding method is more effective in selected genomic regions, allowing a much greater increase of diversity despite selection by introducing effective recombination in the QTL interval.

wheat/ diversity/ selection/ allelic enrichment

* Correspondance et tirés à part : david@supagro.inra.fr

Résumé : Les blés cultivés sont issus d'une domestication initiale suivie de plusieurs épisodes de sélection pour l'agriculture. Leur évolution est marquée par une réduction de diversité globale, et par une réduction supplémentaire dans les régions génomiques porteuses de gènes d'intérêt. C'est le cas par exemple pour une zone du chromosome 2B, impliquée dans la nudité des grains. L'objectif est d'évaluer l'efficacité de deux méthodes de réintroduction de la diversité à partir de la diversité ancestrale dans le blé dur cultivé, en fonction des zones génomiques. L'une de ces méthodes repose sur une sélection généalogique, l'autre sur l'établissement d'un régime allogame dans la population. Nous avons génotypé à l'aide de neuf locus microsatellite génomiques et de sept microsatellites du chromosome 2B une collection représentative de l'histoire des blés cultivés, et un échantillon issu de chacune des deux méthodes de sélection. L'estimation indépendante du taux de mutation a permis d'obtenir les effectifs efficaces de chaque groupe. Nous confirmons la réduction de diversité liée à l'histoire du blé et une sélection sur le chromosome 2B. Nous montrons également que si les deux méthodes de sélection étudiées permettent de restaurer la diversité dans l'ensemble du génome, l'efficacité de la méthode allogame dans les zones soumises à sélection est probante en termes de réintroduction de la diversité.

blé/ diversité/ sélection/ enrichissement allélique

1. INTRODUCTION

Les blés cultivés, représentés essentiellement par le blé dur (*Triticum durum*) et le blé tendre (*Triticum aestivum*), ont traversé au cours de leur histoire des épisodes de sélection importants, qui ont conduit à une bonne adaptation aux exigences de l'agriculture [12]. Pour le blé dur, espèce tétraploïde de génome AB, on identifie la domestication, qui correspond au passage de la forme sauvage *Triticum turgidum dicocoides* à la première forme cultivée *Triticum turgidum dicocum*; la sélection pour les grains nus, qui a donné naissance à la sous-espèce *Triticum turgidum durum* fondamentalement identique à celle qui est cultivée aujourd'hui; et l'amélioration des plantes récente, avec notamment l'introduction de gènes de nanisme, qui est à l'origine des variétés de blé dur utilisées dans l'industrie agro-alimentaire. Le blé tendre, espèce hexaploïde de génomes ABD, n'a pas d'ancêtre sauvage direct et est issu de l'hybridation d'une forme de *Triticum turgidum* domestique, tétraploïde de génome AB, avec l'espèce diploïde sauvage *T. tauschii*, de génome D. Son histoire récente connaît comme le blé dur des épisodes de sélection pour des caractères agronomiques.

Ces différents épisodes historiques ont provoqué une réduction globale significative de la diversité génétique neutre par des effets démographiques dus à l'échantillonnage d'un nombre réduit de plantes [10], [1]. Un effet sélectif a pu accentuer la perte de polymorphisme sur les locus sélectionnés

et sur les locus proches. On observe par exemple sur la base de marqueurs microsatellites que seulement 20 % de la diversité génétique présente au sein de la forme sauvage a été transmise aux variétés actuelles chez le blé dur [10]. Sur la base de polymorphisme de séquences géniques, la diversité maintenue n'est que de 16 % pour les blés durs et 31 % pour les blés tendres [1]. Dans des zones potentiellement soumises à sélection, on peut observer la disparition totale du polymorphisme dans le compartiment cultivé, comme c'est le cas par exemple pour le gène *Gyp1A*, impliqué dans la dureté des grains [1]. Chez le blé dur, un effet sélectif directionnel a aussi été mis en évidence sur le bras court du chromosome 2B, lors du passage des grains vêtus aux grains nus [10]. Cette zone porte le locus *Tg2* (*Tenacious glume*) et correspond à un QTL détecté [6]. Des microsatellites sont cartographiés dans l'intervalle du QTL [7]. Chez le blé tendre, ce QTL n'a pas été détecté sur le chromosome 2B mais sur son chromosome homologue 2D (*Tg1*), éventuellement en raison d'une trop faible densité de marqueurs dans la région homologue 2B [3].

La diversité génétique encore présente dans les compartiments ancestraux est aujourd'hui conservée à l'état statique au sein de collections de ressources génétiques. Il est clair qu'un certain nombre d'allèles de ces formes ancestrales ne sont pas adaptées aux conditions agronomiques actuelles. Toutefois, à d'autres locus, la diversité qu'elles recèlent pourrait se révéler utile. Il est donc important de rendre la diversité ancestrale conservée accessible aux améliorateurs, tout en conservant la valeur agronomique acquise. Ceci correspond à créer un matériel nouveau, au sein duquel les liaisons entre allèles fixés au cours de l'amélioration des plantes auront été remaniées, afin de permettre à nouveau une réponse à la sélection dans les zones du génome n'étant pas impliquées dans les caractères agronomiques, mais accidentellement fixées au sein du matériel moderne par effet d'entraînement.

Plusieurs stratégies sont possibles. Une démarche classique consiste à repérer des génotypes intéressants pour le ou les trait(s) recherché(s), et à introgresser spécifiquement certaines zones du génome *via* la sélection généalogique, ou la sélection assistée par marqueurs. Ces méthodes se révèlent satisfaisantes pour des caractères dont le déterminisme génétique est déjà étudié et simple. Elles sont également efficaces lorsque le matériel « source » n'est pas trop éloigné génétiquement du matériel à améliorer. Il n'y a pas de transfert de caractères indésirables, et la sélection peut se faire sur un nombre de générations raisonnable. Une seconde démarche moins ciblée vise un élargissement massif de la base génétique au sein d'un matériel végétal disponible pour l'amélioration des plantes [8]. Pour les autogames, une des méthodes est de constituer une population à base génétique large, rendue allogame par la présence d'un gène de stérilité mâle qui, en restaurant

une recombinaison efficace, permet un brassage plus important de la diversité génétique.

L'efficacité relative de ces deux méthodes d'élargissement de la base génétique (sélection généalogique *versus* population allogamisée) reste à évaluer. Si on définit cette efficacité comme la quantité de variabilité restaurée sur la durée de l'expérimentation (identique pour les deux méthodes), une façon de l'apprécier est d'observer le polymorphisme neutre comme indicateur de la quantité de diversité d'intérêt potentiellement continue dans le génome. Celle-ci n'est pas attendue identique dans toutes les zones du génome. L'enrichissement allélique est simple à réaliser dans une zone évolutivement neutre en regard des conditions agronomiques actuelles. Au contraire, une zone proche de locus jouant un rôle important dans la valeur agronomique actuelle ne pourra récupérer de la diversité que si elle devient indépendante *via* la recombinaison des sites directement sous sélection. En effet, plus une zone est porteuse de gènes impliqués dans les grandes transitions morphologiques et architecturales (rachis cassant, grains vêtus, taille), plus son polymorphisme aura été réduit et plus la réintroduction de diversité autour de ces gènes sera rendue difficile par la réapparition de caractères défavorables qui devront être éliminés. On s'attend donc dans de telles zones à une efficacité plus grande dans la population allogame.

L'objectif de cette étude est de comparer la réintroduction de diversité génétique au sein du blé dur *via* la sélection généalogique *versus* la constitution d'une population allogamisée, pour des régions chromosomiques neutres (aléatoirement réparties sur le génome) et sélectionnées (autour de la zone QTL 2B responsable de la dureté des glumes). La difficulté est de pouvoir comparer entre locus la perte de diversité subie lors d'épisodes historiques et de quantifier la part de diversité réintroduite dans les populations à base large. En effet, le nombre d'allèles peut varier énormément entre locus à cause de leur taux de mutation. En utilisant un estimateur de la taille génétique efficace comme mesure de diversité, de telles comparaisons deviennent possibles [10]. Nous avons choisi pour cela les marqueurs microsatellites, dont le taux de mutation peut être évalué pour obtenir un estimateur de la taille efficace à partir de la diversité observée au sein d'un échantillon. Une partie des résultats présentés ici se base sur les méthodologies décrites par ailleurs sur un matériel similaire [9], [10]. Nous commençons par confirmer les résultats obtenus dans ces études précédentes sur la déperdition de diversité, puis nous présentons les résultats sur le matériel de régénération de la diversité nouvellement produit.

2. MATÉRIEL ET MÉTHODES

2.1. Matériel végétal

Une collection des principaux représentants de l'histoire des blés cultivés a servi à caractériser la diversité présente au sein du blé dur, du blé tendre et de leurs compartiments ancestraux. Cette collection est constituée de 29 *T. t. dicoccoides*, 11 *T. t. dicoccum*, 17 *T. t. durum* représentant la diversité disponible dans les variétés de pays avant 1950, 146 lignées de *T. t. durum* représentant les variétés inscrites au catalogue français entre 1950 et 2004, 4 *T. t. polonicum*, 4 *T. t. turgidum*, et 15 *T. t. carthlicum*, qui sont toutes trois des sous-espèces apparentées au blé dur. Quarante-sept *T. aestivum*, tirés d'une core collection établie sur plusieurs milliers d'accessions par F. Balfourier (INRA Clermont-Ferrand) à partir de la maximisation de la diversité d'un jeu de 42 microsatellites.

Des croisements entre formes ancestrales et variétés de blé dur élites ont été entrepris dès 1996 au laboratoire. A partir de ce matériel, deux voies de sélection ont été mises en œuvre. La première est une sélection généalogique classique. Durant quelques années, la sélection précoce a porté sur 60000 plantes F2 par an, sélectionnées individuellement dans la pépinière du domaine INRA de Melgueil (Hérault). Les meilleures lignées ont été évaluées et fixées dans un réseau multilocal (7 sites). Un échantillon de 98 lignées à différents stades de fixation (F3 à F8) constitué à partir de ce matériel représente dans cette étude la sélection généalogique. La seconde voie repose sur une population à base large créée à partir de croisements de variétés élites et d'individus représentant la diversité de *T. dicoccoides* et *T. dicoccum*. Cette population est conduite tous les ans sur une parcelle de 10 à 20000 plantes. Elle a été rendue allogame par l'introduction et le maintien d'un gène de stérilité mâle récessif génique. Cette population a été alimentée annuellement par les meilleurs génotypes issus de la sélection généalogique et a subi une sélection massale douce au cours des générations pour le retour à une forme agronomique cultivable. Un échantillon de 45 individus de cette population représente la population allogamisée.

Enfin, le taux de mutation des marqueurs microsatellites a été caractérisé à l'aide de 500 lignées d'accumulation de mutations, issues d'un génotype unique homozygote de la variété ShamI, maintenues pendant 20 générations d'autofécondation stricte [9].

Tous les ADN ont été extraits selon le protocole décrit par Thuillet *et al.* [9] à partir de 0,4 g de matière fraîche provenant de plantes au stade 2-3 feuilles.

2.2. Marqueurs microsatellites

Parmi les 16 marqueurs analysés, 9 sont répartis aléatoirement sur le génome, et 7 sont localisés sur le bras court du chromosome 2B, dont 4 dans l'intervalle du QTL correspondant au caractère d'adhérence de la glume (fig. 1). Les informations concernant l'ensemble des locus sont récapitulées dans le tableau I. Les locus ont été amplifiés par PCR et les allèles visualisés sur séquenceur à capillaire (ABI Prism 3100). Un allèle de la variété ShamI a été séquencé pour chaque locus afin de déterminer la composition en bases du motif répété. Les protocoles utilisés sont décrits par Thuillet *et al.* [9].

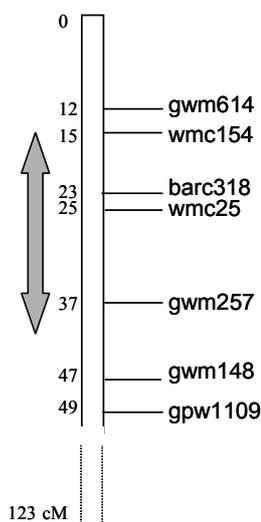


Figure 1 : Positionnement des marqueurs analysés sur la carte génétique du chromosome 2B du blé tendre [7] et positionnement des autres marqueurs relativement à la carte consensus (Wheat Composite 2004-2B, <http://rye.pw.usda.gov>). La flèche indique la zone supposée du QTL liée au gène *Tg2*.

2.3. Estimation des taux de mutation

Le taux de mutation de chaque marqueur est estimé directement sur le dispositif d'accumulation de mutations en comptant le nombre n de mutations apparues parmi les 500 lignées en génération 20 d'autofécondation. À partir de n , on obtient le taux de mutation μ selon l'équation $\mu = n / [L(2 - 2^{l-g} + g)]$ où L est le nombre de lignées analysées et g le nombre de générations d'accumulation [9]. Pour les locus n'ayant pas muté, nous avons calculé un taux de mutation

maximum au risque 0,05 pour qu'aucune mutation ne soit observée au sein du dispositif : $\mu_{\max} = 1 - e^{\ln 0,05 / N}$, où $N = [L(2 - 2^{l-g} + g)]$.

A partir de cette estimation directe, la corrélation attendue entre taux de mutation et longueur des allèles est vérifiée. L'équation établie permet d'estimer le taux de mutation de chaque marqueur à partir de la taille moyenne des allèles pour les différents groupes de blés analysés [10].

2.4. Diversité génétique et effectifs efficaces

Pour chaque locus et au sein de chaque groupe analysé, nous avons calculé la diversité génique de Nei notée H_e [4]. A partir de cet indice, nous avons estimé le paramètre $\theta = 4N_e\mu$ sous l'hypothèse d'un modèle de mutation pas à pas pour les locus analysés tel que : $\hat{\theta}_{H_e} = 1/2[(1 - H_e)^{-2} - 1]$ [5]. Nous avons alors utilisé l'estimateur $\hat{\theta}_F$, qui est une version non biaisée de $\hat{\theta}_{H_e}$ [13], [11].

En nous situant dans l'hypothèse forte que la diversité au sein de chaque échantillon était proche de l'équilibre mutation-dérive, l'effectif efficace N_e à chaque locus et dans chaque groupe de blé étudié a été obtenu en divisant $\hat{\theta}_F$ par quatre fois le taux de mutation correspondant μ . La moyenne des N_e peut alors être utilisée comme estimateur de diversité et permet de comparer les pertes historiques et les gains réalisés à travers les deux méthodes d'élargissement de la base génétique de manière globale et indépendamment des différences de diversité entre locus. Les intervalles de confiance autour des N_e moyens sur le génome, basés sur 9 locus, ont été calculés par bootstrap sur 10000 répétitions. Les N_e moyens sur le chromosome 2B et *a fortiori* dans l'intervalle du QTL autour de *Tg2* étant calculés sur un nombre de locus parfois beaucoup plus faible, allant de 2 à 6, les intervalles de confiance correspondant ne sont pas toujours disponibles et n'ont pas été calculés. Nous avons toutefois vérifié que les différents locus pris indépendamment donnent des résultats cohérents.

3. RÉSULTATS

3.1. Taux de mutation et effet de la longueur des allèles

Les données liées à l'estimation des taux de mutation sont présentées dans le tableau I. Treize mutations ont été détectées sur 6 locus parmi les 16 étudiés. Les taux de mutation estimés varient de 0 à 3.7×10^{-4} mutations par génération. Après séquençage, il apparaît que dix locus ont des motifs

dinucléotides parfaits, trois des trinuécléotides parfaits et trois sont composés de deux types de répétitions, avec pour deux d'entre eux, des imperfections dans la zone répétée (tabl. I). Les mutations ne concernent que les locus à motifs dinucléotides parfaits. Seul le locus *wmc25* n'a pu être séquencé complètement. Pour ce locus, on a pu déterminer que le nombre de répétitions minimum était de 29 unités de répétition.

Tableau I : Caractéristiques et taux de mutation des locus étudiés.

locus	localisation	longueur de l'allèle <i>ShamI</i> (bp)	motif de l'allèle <i>ShamI</i>	nombre de mutations de type pas à pas	taux de mutation	Intervalle de confiance 95%
<i>wmc154</i>	2BS - QTL	143	(GT)35	3	2.77 10 ⁻⁴	[5.70 10-5,8.08 10-4]
<i>wmc25</i>	2BS - QTL	183	(GT)>29	4	3.70 10 ⁻⁴	[1.01 10-4,9.48 10-4]
<i>barc 318</i>	2BS - QTL	266	(TAA)15	0	0	[0,3.07 10-4]
<i>gmw257</i>	2BS - QTL	190	(GT)12	0	0	[0,2.76 10-4]
<i>gmw614</i>	2BS	148	(CT)25	1	9.29 10 ⁻⁵	[2.35 10-6,5.18 10-4]
<i>gmw148</i>	2BS	143	(CA)12(TA)4	0	0	[0,2.77 10-4]
<i>gfw1109</i>	2BS	146	(GT)11	0	0	[0,2.78 10-4]
<i>gfw3142</i>	1AL	208	(GA)23	2	1.84 10 ⁻⁴	[2.23 10-5,6.66 10-5]
<i>gfw2239</i>	4AS	197	(CT)19	1	9.33 10 ⁻⁵	[2.36 10-6,5.20 10-4]
<i>gfw2103</i>	7AL	234	(CT)16	2	1.86 10 ⁻⁴	[2.25 10-5,6.72 10-4]
<i>gfw124</i>	1BL	194	(TC)21imp(TG)16	0	0	[0,2.76 10-4]
<i>gfw7577</i>	2AS	226	(GCA)11	0	0	[0,2.76 10-4]
<i>gfw7101</i>	2AS	140	(TGC)9	0	0	[0,2.79 10-4]
<i>gfw493</i>	3BS	171	(AC)37	0	0	[0,2.78 10-4]
<i>gfw299</i>	3BL	190	(GA)26	0	0	[0,2.76 10-4]
<i>gfw193</i>	6BS	169	(CT)23(CA)10 imp	0	0	[0,2.76 10-4]

Les locus de cette étude ont été ajoutés à 12 locus à motifs dinucléotides parfaits étudiés antérieurement sur le même dispositif [10] pour étudier la corrélation entre nombre d'unités de répétition et taux de mutation (fig. 2). Les taux de mutation estimés dans cette étude sont plus faibles qu'attendus sur la base de l'équation de Thuillet *et al.* [10]. Sur l'ensemble des locus, la variation du nombre d'unités de répétition explique 57 % de celle du taux de mutation.

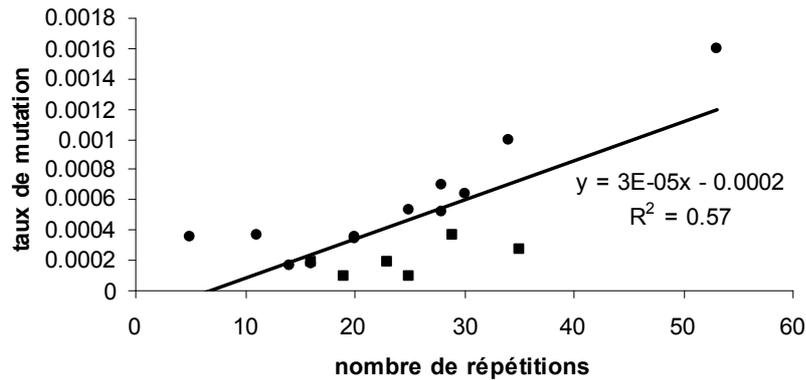


Figure 2 : Corrélation entre la longueur des allèles et le taux de mutation. Les locus de cette étude (carrés) ont été ajoutés aux locus de Thuillet *et al.* [10] (ronds).

L'estimation des taux de mutation μ à partir du nombre d'unités de répétition r au sein des groupes de blé étudiés est basée sur la droite de régression obtenue : $\mu = 3.10^{-5}r - 2.10^{-4}$. Le nombre moyen d'unités de répétition au sein de chaque groupe a été déterminé à partir de la longueur moyenne des allèles dans chaque groupe et de la séquence connue chez la variété ShamI. Le tableau II donne les taux de mutation estimés dans chaque groupe de blé. Le locus *gpn1109* apparaît très instable et particulier dans son type de mutation (9 mutants variant d'une paire de base) et n'a pas été retenu pour les calculs de taille efficace.

Pour un même locus le taux de mutation estimé peut présenter une grande variance (*wcm154*, *wmc25*, *gwm148*, *gwp2103*, tableau II), ce phénomène étant particulièrement marquant pour le locus *gwm148* dont les taux de mutation varient d'un facteur 20, en raison d'une grande différence de taille moyenne des allèles entre les groupes. Le taux de mutation du locus *barc318* pour le groupe *turgidum* n'a pas pu être déterminé, la longueur moyenne des allèles dans ce groupe étant plus petite que la taille de la séquence flanquante chez la variété ShamI.

Tableau II : taux de mutation estimés dans chaque groupe de blé étudié pour chaque locus ($\times 10^4$). ND = non disponible. DD : *T. t. dioicaoides* ; DC : *T. t. dicoccum* ; DR pays : variétés de pays de *T. durum* ; DRElite : variétés élités de *T. t. durum* ; PO : *T. t. polanicum* ; TU : *T. t. turgidum* ; CA : *T. t. cartholicum* ; SG : lignées issues de la sélection généalogique ; Pop : individus issus de la population allogame ; Parents : géniteurs des populations sélectionnées (SG et Pop).

locus	DD	DC	DR pays	DRElites	SG	Pop	PO	TU	CA	parents	variance
<i>nmcl54</i>	8,2	10,6	8,5	9,1	8,8	8,8	8,5	7,6	19,9	8,8	13,0
<i>nmc25</i>	11,8	11,5	7,6	7,6	8,8	8,2	7	6,4	17,8	8,5	11,6
<i>barc318</i>	2,5	2,5	2,5	2,5	2,5	2,5	3,1	ND	2,2	2,8	0,1
<i>gpm257</i>	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,6	0,0
<i>gpm614</i>	4,6	4,6	5,2	5,5	5,5	5,5	4,6	4,9	4,6	5,5	0,2
<i>gpm148</i>	9,7	8,5	6,7	10	11,5	8,8	8,2	13,9	0,7	7	12,0
<i>gpm1109</i>	0,1	0,1	0,4	0,7	0,7	0,4	0,4	0,4	0,1	0,4	0,0
<i>gpm3142</i>	4,3	4,6	4,6	4,3	4,3	4,3	5,5	4	4,3	4	0,2
<i>gpm2239</i>	4	3,1	3,1	3,4	3,4	3,4	3,7	3,4	2,8	3,7	0,1
<i>gpm2103</i>	3,7	4,3	16,3	3,1	2,8	3,1	2,5	2,8	4,6	3,1	17,3
<i>gpm124</i>	5,5	4,3	3,7	4	3,7	3,7	3,4	3,4	3,1	4	0,4
<i>gpm7577</i>	1,3	1,3	1	1,3	1	1	1	1	1	1	0,0
<i>gpm7101</i>	0,4	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,0
<i>gpm493</i>	7,9	5,5	7,3	6,7	7,6	6,7	9,7	6,7	6,4	7,3	1,2
<i>gpm299</i>	6,4	7,6	6,4	6,4	7,9	6,7	6,4	5,8	5,8	7,9	0,6
<i>gpm193</i>	5,5	3,7	4	4	3,7	4,3	3,1	4,3	4,9	4	0,4

3.2. Évolution de la diversité au cours de l'histoire du blé dur

La diversité mesurée par la diversité génique diminue en moyenne sur les 16 locus en passant de 0,87 au sein de la sous-espèce ancestrale *T. t. dicoccoides* à 0,73 pour *T. t. dicoccum*, à 0,64 pour les variétés de pays de blé dur et à 0,45 pour les variétés élites. Les sous-espèces apparentées cultivées présentent des valeurs de diversité proches du blé dur : 0,55 pour *T. t. polonicum*, 0,52 pour *T. t. turgidum* et 0,55 pour *T. t. carthlicum*.

La figure 3 démontre l'importance de l'effet du taux de mutation sur la valeur observée de la diversité génique qui est corrélée au taux de mutation alors que l'effectif efficace ne l'est pas (les données sont présentées au sein du groupe de variétés élites de *T. t. durum* ; le résultat est similaire dans les autres groupes). La réduction de diversité exprimée en terme d'effectifs efficaces montre la même tendance que la diversité génique, mais de façon plus marquée, puisque sur la base de l'effectif efficace les variétés de blé dur élites ne retiennent que 2,3 % de la diversité présente au sein du compartiment sauvage, alors que cette valeur atteint 51,6 % lorsqu'elle est calculée à partir de la diversité génique.

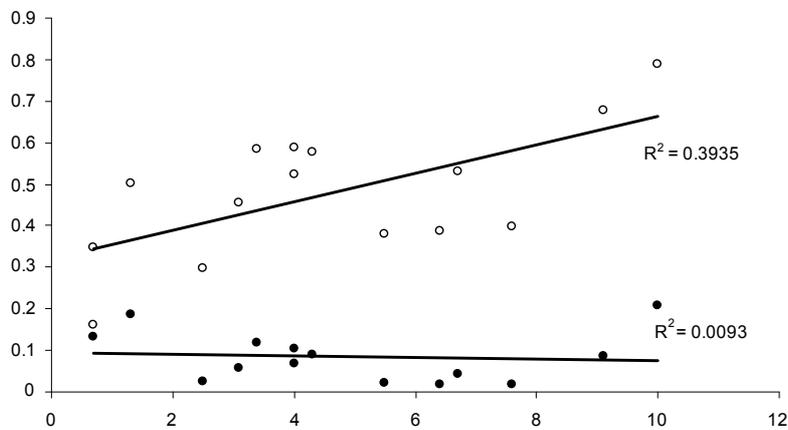


Figure 3 : Relation entre le taux de mutation ($\times 10^{-3}$) et la diversité génique (ronds blancs) versus l'effectif efficace (ronds noirs ; $\times 10^{-4}$) au sein du groupe de variétés élites.

Le tableau III résume les valeurs de N_e calculées indépendamment pour les 9 locus aléatoirement répartis dans le génome, les 7 locus du bras court du chromosome 2B dont les 4 locus situés dans l'intervalle autour de *Tg2*. Le passage de *T. dicoccoides* à *T. dicoccum* ne s'est pas accompagné d'une réduction de diversité particulièrement plus marquée dans la zone de *Tg2* que sur le reste du génome. Par contre, les locus de cette zone perdent comparativement plus

de diversité que le reste du génome lors du passage de *T. t. dicoccum* aux variétés de blé dur et plus encore lors du passage aux variétés élites. La sous-espèce *T. t. polonicum* semble avoir subi cet effet de façon encore plus marquée, puisqu'elle ne contient que 3 % de la diversité de son ancêtre *T. t. dicoccum* sur les locus de l'intervalle du QTL alors que ce chiffre atteint 15 % sur le reste du génome. C'est par ailleurs pour ce groupe et pour le groupe *T. t. durum* que la réduction de diversité spécifique aux locus autour de *Tg2* est la plus convaincante, dans la mesure où elle s'accroît lorsqu'on considère uniquement les locus de l'intervalle du QTL versus l'ensemble des locus du bras chromosomique.

Tableau III : Efficacités estimées sur l'ensemble du génome (Ne genome), sur le bras court du chromosome 2B (Ne2B), et dans l'intervalle du QTL autour de Tg2 (NeQTL). Les pourcentages de diversité maintenue sont calculés par rapport à la sous-espèce précédente pour *dicoccum*, *durum* variétés et *durum* élites, et par rapport à *dicoccum* pour *polonicum*, *turgidum* et *carthlicum* et le blé tendre.

	Ne genome	IC 95%	proportion de diversité retenue	Ne2B	proportion de diversité retenue	NeQTL	proportion de diversité retenue
dicoccoïdes	27354	[16622,39946]		48147		35823	
dicoccum	3898	[1557,7006]	0,14	8012	0,17	8264	0,23
durum variétés	2514	[1553,3583]	0,64	3305	0,41	4016	0,49
durum élites	917	[610,1243]	0,36	712	0,22	428	0,11
polonicum	690	[334,1116]	0,18	934	0,12	221	0,03
turgidum	600	[345,882]	0,15	732	0,09	913	0,11
carthlicum	1649	[343,3409]	0,42	1951	0,24	1122	0,14
blé tendre	9902	[3612,17447]	2,54	7737	0,97	2048	0,25

3.3. Élargissement de la base génétique dans les deux programmes de sélection étudiés

Sur l'ensemble du génome, une partie de la diversité a été réintroduite via les deux programmes de sélection et de façon comparable : les effectifs efficaces de la population allogame et de la population de lignées valent respectivement 1272 et 1282, tandis que celui de l'ensemble des variétés élités ne vaut que 917 (tabl. IV). Toutefois les deux types de population n'ont pas complètement retrouvé le niveau de diversité de leurs géniteurs, qui comprenaient un ensemble de variétés et de formes ancestrales ($N_e = 3725$). Une différence se fait sentir au contraire entre les deux programmes lorsqu'il s'agit de la zone sélectionnée : dans la population allogame, l'enrichissement de diversité par rapport aux variétés de blé dur élités sur le bras court du chromosome 2B, et encore plus dans l'intervalle du QTL, est plus marqué que sur le reste du génome. Cet enrichissement est également plus fort comparé aux lignées sur les même locus : par exemple sur le chromosome 2B, alors que l'effectif efficace augmente à peine des variétés élités aux lignées de sélection généalogique, il double au sein de la population allogame (tabl. IV). De la même façon, l'appauvrissement de diversité de la population allogame par rapport au groupe de géniteurs est moins fort que pour le reste du génome, alors que la perte est accentuée dans la zone du QTL pour les lignées.

Tableau IV : Effectifs efficaces estimés chez le blé dur avant et après les deux programmes de sélection, et chez le blé tendre, sur l'ensemble du génome, sur le chromosome 2B et dans l'intervalle de QTL.

	Ne génome	IC 95%	Ne2B	NeQTL
<i>durum</i> élités	917	[610,1243]	712	428
parents sélection généalogique	3725	[1876,5669]	3092	2195
population allogame	1282	[852,1756]	805	856
	1272	[868,1712]	1419	1530

4. DISCUSSION

Les effectifs efficaces calculés sont basés sur un certain nombre d'hypothèses. Tout d'abord, le modèle de mutation supposé des microsatellites est le modèle de mutation pas à pas également appelé Stepwise Mutation Model (SMM), qui considère que les locus ne mutent que par pas de une unité de répétition en plus ou en moins. Ainsi, parmi les locus étudiés, le locus *gwm1109*, qui a subi au sein du dispositif

d'accumulation de nombreuses mutations d'une base, a été écarté du calcul des effectifs efficaces. Il est possible que les autres locus, bien qu'apparemment conformes à un SMM sur la base de l'observation directe, aient subi au sein des compartiments étudiés des mutations de taille supérieure à une unité de répétition. Mais la méthode de calcul des effectifs efficaces est relativement robuste à de telles déviations [10]. Une seconde réserve à la fiabilité du calcul des tailles efficaces est l'utilisation des tailles d'allèles pour inférer les taux de mutation moyens au sein de chaque compartiment. La corrélation entre le nombre d'unités de répétition et le taux de mutation est bonne, mais les allèles ont pu subir des insertions ou des délétions en dehors de la zone répétée. Dans ce cas, le taux de mutation est surestimé par rapport au nombre réel de répétitions. Dans l'ensemble, les locus étudiés montrent toutefois une variance de leur taux de mutation peu élevée entre les différents compartiments étudiés, ce qui écarte dans ce cas un effet trop gênant des insertions/délétions hors de la zone répétée si elles existent. Les quatre locus pour lesquels ce problème peut être soulevé (*wmc154*, *wmc25*, *gwm148* et *gwp2103*) ne le doivent chaque fois qu'à un ou deux groupes (*T. t. turgidum*, *T. carthlicum* et pour le locus *gwm2103*, *T. t. durum*, variétés de pays), pour lesquels les tailles d'allèle sont particulièrement petites ou particulièrement grandes. Les effectifs efficaces moyens de ces groupes ont été recalculés en excluant les locus concernés, et nous avons vérifié que nos conclusions restaient inchangées.

La réduction d'effectif efficace observée depuis l'ancêtre sauvage du blé dur jusqu'aux variétés élites donne une image un peu différente de celle obtenue à partir des résultats publiés antérieurement par Thuillet *et al.* [10]. Tout d'abord, les effectifs efficaces estimés dans l'étude présente sont notamment tous plus petits que précédemment estimés. Ensuite, si les valeurs restent cohérentes dans les études pour l'ancêtre sauvage, étant donnés les intervalles de confiance, la diminution de N_e qui suit dans les autres compartiments est clairement accentuée ici. En effet, *T. t. dicoccoides* a une taille efficace estimée de 32500 avec un intervalle de confiance de [20739-45991] dans l'étude de Thuillet *et al.* [10] et de 27357 avec un intervalle de confiance de [16622-39946] dans cette étude, alors que par exemple, *T. t. dicoccum* présentait un $N_e = 12000$ ([5790-19300]) contre un $N_e = 3898$ ([1557-7006]) ici. Une explication peut s'articuler autour des taux de mutation estimés. En effet, les taux de mutation des locus observés dans la présente étude sont tous en deçà de la valeur prédite par la régression sur la taille des allèles dans les différents compartiments de blé dur (fig. 2). Cette surestimation du taux de mutation a pu conduire à une sous-estimation globale des effectifs efficaces. Par ailleurs, l'étude de Thuillet *et al.* [10] est basée sur des échantillons d'individus plus importants, notamment pour *T. t. dicoccum*, alors que nous utilisons ici une core collection définie à partir de ces échantillons. L'observation des valeurs de diversité génique appuie cette

interprétation en donnant des valeurs plus comparables entre les deux études sauf pour la sous-espèce *dicoccum*. On a en effet 0,85 – 0,81 – 0,61 – 0,51 et 0,87 – 0,73 – 0,64 – 0,45 respectivement pour les sous-espèces *dicoccides*, *dicoccum*, *durum* variétés de pays et *durum* variétés élites, selon Thuillet *et al.* [10] et dans cette étude.

L'effectif efficace du blé tendre apparaît particulièrement élevé. Il est notamment deux à trois fois plus grand que l'effectif efficace de *T. t. dicoccum*, et quatre fois plus grand que les variétés de pays de blé dur. L'hypothèse de formation de cette espèce est un croisement entre un *T. turgidum* domestique (sous-espèce *dicoccum* ou *durum*) et un diploïde sauvage. On aurait donc pu s'attendre à ce que l'effectif efficace du blé tendre, conformément à l'image obtenue sur la base des valeurs de diversité génique, soit inférieur au moins à celui de la sous-espèce *dicoccum* ($H_e = 0,73$ pour *T. t. dicoccum* et $H_e = 0,68$ pour le blé tendre). Là encore, il est probable qu'une mauvaise représentation de la sous-espèce *dicoccum* reflète une valeur de N_e plus petite qu'en réalité. Ceci n'exclut pas néanmoins une explication biologique. La taille efficace du blé tendre après la domestication a été exponentiellement augmentée par le succès agronomique de cette espèce. Des mutations ont pu s'y accumuler facilement. L'échantillon étant basé sur une core collection mondiale, la différenciation entre accessions est maximisée ainsi que le nombre d'allèles à différents locus. Au contraire, *T. dicoccum* a connu une forte régression depuis l'apparition des formes modernes et n'a pas connu une expansion aussi grande.

Concernant les autres sous-espèces cultivées de l'espèce *turgidum* (*polonicum*, *turgidum*, *carthblicum*), leurs tailles efficaces sont du même ordre que les variétés actuelles de *T. t. durum*, ce qui est cohérent avec leur statut. Seule la sous-espèce *carthblicum* semble avoir une histoire plus particulière, avec un effectif efficace plus élevé. Une hypothèse, peu étayée mais parfois avancée, suggère un lien de parenté entre cette sous-espèce et le blé tendre et pourrait expliquer cette observation [2]. Comme cela était attendu, pour toutes les formes étudiées à grains nus, la proportion de diversité retenue par rapport à leur ancêtre à grains vêtus (*T. t. dicoccum*) est plus faible sur le chromosome 2B que sur la base des locus répartis aléatoirement sur le génome. Cette tendance s'accroît même lorsque l'on considère les locus directement situés dans l'intervalle du QTL autour du gène *Tg2*. Ceci est compatible avec un fort effet sélectif existant dans cette zone chromosomique en relation avec la nudité des grains.

La réintroduction de diversité au sein d'un ensemble de lignées élites de blé dur sur l'ensemble du génome a fonctionné de façon identique pour les deux méthodes de sélection appliquées (sélection généalogique et population allogame), puisque les deux échantillons présentent des effectifs efficaces très similaires entre eux (respectivement 1282 et 1272) et plus élevés que

celui de l'ensemble des variétés élites inscrites au catalogue français depuis 1950 (917). Dans la zone du QTL déterminant la nudité des grains, l'effet des deux méthodes est au contraire différent. Si les lignées généalogiques ont récupéré de la diversité par rapport aux variétés élites de blé dur (805 *versus* 712), cette enrichissement reste plus faible que les autres zones génomiques couvertes par les locus microsatellites ($N_e = 1282$). Cette observation est compatible avec une élimination rapide des individus à grains vêtus lors de la sélection et suggère la persistance dans la sélection généalogique d'un déséquilibre de liaison entre le gène causal du caractère et les locus neutres alentours. Par contre, les individus issus de la population allogame présentent dans cette zone une restauration de diversité du même ordre que pour le reste du génome, et à peu près du niveau existant au sein de leurs géniteurs (2195). Ceci montre que la recombinaison permise sur plusieurs générations dans des populations de grande taille lève la contrainte sélective aux alentours du gène. Le maintien de la pression de sélection sur le caractère vêtu ou non des grains permet de produire du matériel nouveau sans perte de valeur par rapport aux conditions agronomiques actuelles. Dans le même temps, le brassage assuré par davantage de recombinaisons autour de la zone d'intérêt permet la régénération de la diversité dans les zones génomiques adjacentes, initialement impliquées dans l'adaptation de l'espèce aux contraintes de l'agriculture par effet d'entraînement, mais non causal du trait étudié. Les ressources génétiques nouvelles créées par ce type de gestion que l'on pourrait qualifier de néo-domestication sont donc potentiellement porteuses de caractéristiques intéressantes qui avaient disparu du génome des blés cultivés depuis leur domestication et l'apparition des grains nus il y a 8000 ans. La diversité réintroduite, observée ici sur la base de marqueurs neutres comme indicatrice de la diversité présente sur les zones génomiques adjacentes, est non caractérisée, mais présente un intérêt potentiel de réponse à la sélection dans des environnements futurs nouveaux.

REMERCIEMENTS

Les auteurs remercient François Balfourier pour avoir fourni une core collection de 47 blés tendres représentative de la diversité de l'espèce.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Haudry A., A. Cenci, C. Ravel, T. Bataillon, D. Brunel, C. Poncet, I. Hochu, S. Poirier, S. Santoni, S. Glémin, J. David, 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24: 1506–1517.

- [2] Metakovsky E.V., Kudryavtsev A.M., Iakobashvili Z.A., A.Y. Novoselskaya, 1988. Analysis of phylogenetic relations of *durum*, *carthlicum* and common wheats by means of comparison of alleles of gliadin-coding loci. *Theor. Appl. Genet.* 77:881–887.
- [3] Nalam V.J., Vales M.I., Watson C.J.W., E.B. Johnson, O. Riera-Lizarazu, 2007. Map-based analysis of genetic loci on chromosome 2D that affect glume tenacity and threshability, components of the free-threshing habit in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 116:135–145.
- [4] Nei M., 1987 *Molecular evolutionary genetics*. Columbia University Press, New-York.
- [5] Ohta T., M. Kimura, 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* 22: 201–204. 239.
- [6] Simonetti M. C., M. P. Bellomo, G. Laghetti, P. Perrino, R. Simeone *et al.*, 1999. Quantitative trait loci influencing free threshing habit in tetraploid wheats. *Genet. Resour. Crop Evol.* 46: 267–271.
- [7] Somers D.J., Isaac P. and K. Edwards, 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.) *Theor. Appl. Genet.* 109: 1105–1114
- [8] Spoors W., N.W. Simmonds, 2001. Base broadening: introgression and incorporation. In: *Broadening the genetic base of crop production* (eds) HD Cooper, C. Spillane, T. Hodgkin Cabi Pub., p. 71-81.
- [9] Thuillet A.C., D. Bru, J. David, P. Roumet, S. Santoni, P. Sourdille, Bataillon T., 2002. Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Mol. Biol. Evol.* 19: 122–125.
- [10] Thuillet A.C., Bataillon T., Poirier S., Santoni S., David J.L., 2005. Estimation of the effective population size from microsatellite data in different groups representative of key stages in the history of *durum* wheat. *Genetics* 169:1589-1599.
- [11] Xu H., and Y. X. Fu, 2004. Estimating effective population size or mutation rate with microsatellites. *Genetics* 166: 555–563.
- [12] Zohary D., M. Hopf, 2000. *Domestication of plants in the old world*. Oxford University Press, New York.
- [13] Zouros E., 1979. Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics* 92: 623-646.

Interactions génomiques en régime de sélection et taille de population

Michel VEUILLE^{(1)*}, Lionel BRAZIER⁽¹⁾, Florent MELI⁽¹⁾

⁽¹⁾EPHE, UMR 5202 CNRS-MNHN, Muséum, 16 rue Buffon, 75005 Paris, France

Abstract. Genomic interactions under selection regime. Several population genetics models suggest that adaptation is slower in smaller populations. This condition is encountered in endangered species, and when selecting varieties in agricultural research. It thus matters as an important factor in the management of genetical resources. The main hypothesis that we wanted to check was that population size influences associative overdominance, a phenomenon that slows down the selective fixation of alleles located at distinct, but genetically linked, loci. We established a protocol of observation at the molecular level, using loci that our research group had already characterized as having recently undergone selective sweep events. They are located in regions where crossing-over inhibitors are expected to maintain associative overdominance. To this end, we sequenced twenty chromosomes in 10700 nucleotides over 11 loci covering 170 kb in a first candidate region, and in 16500 nucleotides over 18 loci covering 850 kb in a second candidate region. In these two regions, the population sample is divided into two sub-populations which exchange little or no genetic material, since they are located near inversions breakpoints. In each case, the selective sweep was very limited in size, in the center of the surveyed region. In one case, the selective sweep occurred some time ago (probably about 60,000 generations ago). In the other case, the selective sweep is recent, with no evidence of subsequent mutations. The sweep was characterized in the first case, by a lack of high frequency silent nucleotide polymorphism. In the other case, it was characterized by statistically significant linkage disequilibrium, using a haplotype test of selective neutrality. In the two cases, there was evidence that a few haplotypes of short length had invaded one of the sub-populations from the other. In the first genomic region, only gene conversions could exchange genetic material across inversion, and only one of them seems to have occurred. In the second genomic region, there is evidence of three gene conversions and of one recombination event. Our observations thus show evidence of an unexpected phenomenon. Recombination inhibition was circumvented by genic conversion, an as yet poorly understood genetic mechanism, which is difficult to record, and is, for this reason, often overlooked. Virtually no other instance of gene conversion was observed in the rest of the sequenced regions. Clearly, a few unique events of gene conversion had been favoured by natural selection, had increased in frequency, and had eventually fixed. These selectively favoured gene conversion events were circumscribed in the two small regions where we had detected selective sweeps. Our results suggest that

* Correspondance et tirés à part : veuille@mnhn.fr

associative overdominance generates a strong selective pressure in favour of gene conversion events. Since the number of new gene conversion events produced at each generation depends on the population number. They also substantiate our original hypothesis of an effect of population size upon the response to selection, since the success of the migration of a favoured allele from one chromosomal sub-population to the other depends on the number discrete events of gene conversion – not on their frequency. In this study, we used the *Drosophila* genome as a model for technical reasons, but its conclusions should have a broader level of generality. In particular, they imply that the immediate population number (not simply the effective population size) is an important factor to consider in carrying out selection protocols.

selection/ population size/ recombination/ gene conversion

Résumé. Plusieurs modèles de génétique des populations suggèrent que l'adaptation est plus lente dans les petites populations. Cette situation concerne d'une part les espèces menacées, et d'autre part les variétés sélectionnées en agronomie, c'est-à-dire des situations de gestion des ressources génétiques. Notre hypothèse principale est que la taille de la population influence la surdominance associative, phénomène de ralentissement de la fixation sélective d'allèles situés en des locus distincts, mais liés. Nous avons établi un programme d'observations à l'échelle moléculaire, utilisant des gènes précédemment répertoriés par notre équipe comme ayant été récemment soumis à des événements de balayage sélectif, et situés dans des régions où des inhibiteurs de crossing-over sont susceptibles de maintenir de la surdominance associative. Nos observations mettent en évidence un phénomène inattendu. L'inhibition de la recombinaison a été tournée par la conversion génique, un phénomène génétique aux mécanismes encore mal compris, car difficile à étudier, et donc, souvent négligé. Nos résultats suggèrent que la surdominance associative engendre une pression sélective forte en faveur des phénomènes de conversion génique. Le nombre d'évènements géniques, et non pas leur fréquence, détermine le succès du transfert d'un allèle avantageux d'une sous-population chromosomique à l'autre. Nos résultats confortent ainsi notre hypothèse d'une influence de la taille de la population sur la réponse à la sélection. Ils suggèrent que la taille instantanée d'une population expérimentale (et non simplement son effectif efficace) est un important facteur dans le succès des protocoles de sélection.

sélection/ taille de population/ recombinaison/ conversion génique

1. INTRODUCTION

Trois séries d'arguments mènent à prédire un ralentissement de l'évolution dans les petites populations. Un premier argument fait intervenir « l'interaction de Hill et Robertson », ou la « surdominance associative » [14], [15]. Il pose que la sélection à un locus contrarie la sélection à d'autres locus, sauf dans le cas particulier où les allèles favorisés des deux locus sélectionnés

sont en configuration « cis », c'est-à-dire placés sur le même chromosome. Mais deux allèles avantageux situés à des locus distincts ont une faible probabilité d'apparaître sur le même chromosome. S'ils possèdent tous deux un avantage équivalent, leurs fréquences augmenteront, mais ils resteront localisés en configuration « trans », maintenant une sorte de polymorphisme, jusqu'à ce que la recombinaison les associe sur le même haplotype, et permette à un chromosome unique portant les deux adaptations de se fixer. La durée de cette période dépend du temps d'attente jusqu'à la première recombinaison engendrant l'haplotype optimal. En effet, une fois celui-ci créé, il parviendra rapidement à fixation. Ce temps dépend du nombre de recombinaisons par génération, donc de la taille de la population. Si les avantages entre les deux allèles de loci différents sont de valeur différente, celui qui est le plus avantageux peut éliminer l'autre. Cependant, cette issue n'est pas fatale. Dans les populations naturelles, la plupart des chromosomes portent des mutations délétères. Lorsque le chromosome portant un allèle avantageux augmente de fréquence, les mutations délétères qu'il porte tendent à être plus souvent présentes à l'état homozygote, ce qui crée un désavantage qui peut progressivement ralentir, voire annuler, l'augmentation de fréquence de cet allèle. Ainsi, cette interaction de Hill et Robertson peut-elle se généraliser à tout allèle doté d'un avantage, pour lequel le temps d'attente jusqu'à fixation correspond au temps nécessaire pour une « purge » des allèles délétères voisins sur le chromosome.

Le deuxième argument est dérivé de l'effet « d'auto-stop », ou de « balayage sélectif », imaginé par Maynard-Smith et Haigh [16]. Il pose que la fixation d'un allèle avantageux à un locus abaisse la variabilité génétique de la région génomique voisine. En effet, l'allèle avantageux entraîne avec lui les allèles des autres locus avec lesquels il est en déséquilibre de liaison. Cette baisse de variabilité est contrecarrée par la recombinaison. Elle dépend du logarithme de la taille de la population, car la fixation d'un allèle avantageux prend un temps approximativement égal à $t = \log(2N/s)$; avec N , taille efficace, et s , avantage sélectif [1]. La variabilité génétique disponible dans la population étant abaissée, le potentiel évolutif futur de la population sera également affecté. Cette baisse sera d'autant plus sensible dans les petites populations.

Un troisième argument est simplement basé sur le fait que le nombre de mutations qui apparaissent dans les petites populations à chaque génération ($2N\mu$, où μ est le taux de mutation), est plus faible que dans les grandes populations. En outre, ces mutations sont plus rapidement éliminées par dérive. C'est une autre manière de dire que le paramètre mutationnel $\theta = 4N\mu$, dont la valeur prédit celle de la diversité moléculaire, est lui-même plus faible. Sur cette base, Frankham (1995) prédit l'extinction des petites

populations, citant en exemple ses propres résultats sur des populations expérimentales de drosophile.

Dans ce travail, nous avons bâti un protocole d'observation nous permettant de savoir si, lors d'évènements de « balayage sélectif », la dynamique du processus sélectif avait été ralentie par les interactions génomiques créées par le déséquilibre de liaison au sein des haplotypes où était apparu l'allèle favorisé. Notre projet de recherche initial comportait aussi une partie théorique, cependant, nos résultats inattendus en génétique nous ont conduits à documenter complètement cette partie. Le sous-projet théorique initial est désormais moins lié aux hypothèses génétiques, et sera traité indépendamment. Comme matériel d'étude, nous avons utilisé la drosophile, qui a été à l'origine, par le passé, de beaucoup de nos connaissances en génétique quantitative et sélection expérimentale et qui, par sa maniabilité comme matériel, peut encore servir à éclairer certaines questions appliquées de génétique moléculaire de la sélection.

2. MATÉRIEL ET MÉTHODES

2.1. Système génétique

Lors d'études antérieures, notre équipe a révélé de nombreux cas de balayage sélectif dans le génome de *Drosophila melanogaster* [1], [2], [4], [7], [17] et a élaboré des tests statistiques basés sur la théorie de la coalescence, démontrant l'action de la sélection [10], [4], [5], [8], [9], [18]. Nous avons proposé d'utiliser ces situations pour étudier la façon dont la sélection était parvenue au bout du processus de fixation d'un allèle avantageux. Pour cela, nous avons étudié les profils de variabilité moléculaire autour des régions concernées. L'organisme que nous avons utilisé comme modèle ne pose pas en lui-même de question fondamentale en matière de sélection ou de biologie de la conservation, hormis le rôle-clé que lui ont accordé des auteurs comme Frankham ou Parsons comme modèle d'étude en génétique de la conservation. C'est un organisme au génome entièrement séquencé, qui présente l'avantage d'avoir une génétique très bien connue. En particulier, les régions où nous avons trouvé des cas de balayage sélectif portent des « inversions chromosomiques », qui agissent comme des inhibiteurs de crossing-over. Dans la population de Lamto (Côte d'Ivoire) étudiée par notre équipe, l'inversion *In(2L)t* a une fréquence proche de 50 %, ce qui signifie que dans certaines zones du chromosome concerné (le chromosome 2), la population est fractionnée en deux types chromosomiques (*inversé* ou *standard*), presque deux « sous-populations », qui n'échangent pas, ou très peu, de matériel génétique. Comment, dans ces circonstances, un allèle avantageux apparu dans l'une des configurations

chromosomiques (*standard* ou *inversé*), avait-elle pu se propager à l'ensemble de la population ? A ce stade, il faut considérer que, si l'allèle avantage n'a pas la possibilité de s'étendre aux autres chromosomes, son avantage sélectif devrait éliminer à la longue les chromosomes exclus. Si le système d'inversion reste polymorphe, c'est bien plutôt parce que l'allèle avantage a la possibilité de « passer » d'un chromosome à l'autre.

2.2. Région génomique *Vha68-1*

Cette région est ici nommée d'après le locus où un cas de balayage sélectif avait été antérieurement mis en évidence [7]. Ce travail montrait un balayage total très proche du pied proximal de l'inversion chromosomique *In(2L)t*, intérieurement à celui-ci. Cette position interdit tout crossing-over simple dans cette région. Autour de ce locus, nous avons séquencé et aligné des fragments d'environ 1 kb sur 10 loci répartis sur 170 kb, et dans 20 chromosomes issus de lignées naturelles (Côte d'Ivoire) rendues homozygotes. En tout, nous avons donc le polymorphisme de 10 700 nucléotides. La moitié de l'échantillon présente le chromosome standard, et la moitié présente le chromosome inversé. Le travail précédent avait en outre montré que le balayage sélectif était assez ancien pour que quelques mutations soient apparues depuis, permettant de dater l'évènement de quelques dizaines de milliers d'années. Le balayage sélectif étant total, laissait supposer que la cible de la sélection était centrée sur ce gène, ou au voisinage immédiat de celui-ci.

2.3. Région génomique *Su(H)*

Un cas de balayage sélectif centré sur ce gène a été mis en évidence par Depaulis, Brazier et Veuille [6]. Il est très récent, car aucune mutation ultérieure n'a été mise en évidence sur ce gène. Notre étude concluait que ce gène n'était pas la cible de la sélection, celle-ci étant vraisemblablement plus proximale (plus proche du centromère). Ce gène est situé assez loin du pied de l'inversion, entre celle-ci et le centromère. L'effet d'inhibition des crossings-overs s'y fait fortement sentir, bien qu'on ne puisse exclure des crossings-over simples. Dans cette région, en comptant le locus *Su(H)*, nous avons séquencé le polymorphisme de 12 locus sur 180 kb, selon le même échantillonnage que précédemment. En incluant 6 autres loci à longue distance, nous avons séquence au total le polymorphisme de 18 locus sur 850 kb, pour un total de 16 500 nucléotides. Nous avons utilisé les mêmes lignées homozygotes que précédemment.

3. RÉSULTATS

3.1. Région génomique *Vha68-1*

Les résultats bruts sont montrés sur les tableaux I et II. La figure 1 montre la carte génétique de la région étudiée. La figure 2 montre les profils de variation moléculaire dans la région étudiée. L'effet du balayage sélectif est bien visible dans la zone du locus *Vha68-1* (dénommé « F » par simplification sur la figure). En effet, le polymorphisme moléculaire θ a la valeur attendue dans ce génome sous un modèle neutre pour les autres gènes (1-2 %), mais s'abaisse brusquement sur le locus F et un locus voisin, E (*Vha68-2*). Le résultat donné par le F_{ST} est frappant et concordant avec le reste. Le F_{ST} est une mesure classique de structuration d'une population subdivisée en sous-populations. Dans ce cas, on a mesuré la différenciation génétique entre les chromosomes *inversé* et *standard*. Le F_{ST} est significatif pour l'ensemble des gènes, sauf pour le locus F. A ce niveau, sa valeur s'abaisse d'une moyenne de 10-15 % à zéro. Simultanément, le F_{ST} n'est plus significatif au niveau de ce gène.

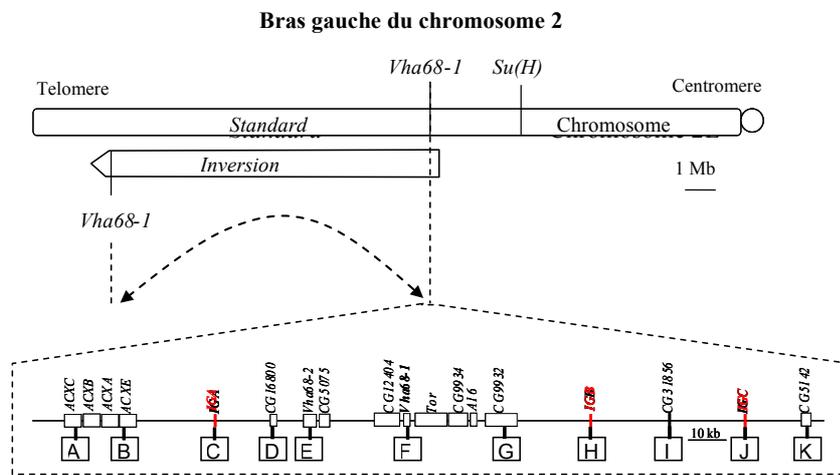


Figure 1 : Région *Vha68-1* : carte génétique de la région entourant le locus. Les symboles {A, .., K} représentent les marqueurs moléculaires utilisés. Les boîtes figurent l'ensemble des régions codantes identifiées. La carte chromosomique montre les deux emplacements où se trouve le locus *Vha68-1*, selon le variant chromosomique (standard ou inversé), où il se trouve. Le trait d'échelle représente 1 megabase.

Tableau 1: Statistiques descriptives pour la région *Vha68-1*.

Gene region	Marker locus	#	Length	Length (silent)	Polymorphic positions	Polymorphic positions (silent)	Mutations	Singlerons	Haplotypes	Haplotype diversity
					<i>S</i>	(silent)	<i>η</i>	<i>η_e</i>	<i>K</i>	<i>H</i>
<i>ACXC</i>	A	20	926	433	20	20	20	6	8	0,795 *
<i>ACXE</i>	B	19	894	374	37	36	37	12	13	0,891
<i>IGA</i>	C	18	758	758	26	26	26	5	11	0,901
<i>CG16800</i>	D	20	942	349	58	26	60	23	14	0,915
<i>Vha68-2</i>	E	20	952	924	18	18	18	9	12	0,89
<i>Vha68-1</i>	F	20	1 063	524	14	13	14	10	10	0,72 *
<i>CG932</i>	G	20	843	769	31	31	31	9	9	0,825 *
<i>IGB</i>	H	20	902	902	30	30	30	8	11	0,87
<i>CG31856</i>	I	20	1 344	1 344	53	53	54	25	13	0,905
<i>IGC</i>	J	20	999	999	23	23	23	6	11	0,88
<i>CG5142</i>	K	20	1 039	780	44	44	44	11	10	0,86 *
Total			10 662	8 156	354	320	357	124		

*: $p < 0,05$.

Notre interprétation est la suivante : au cours du processus sélectif qui a vu sa fréquence passer d'une valeur $1/2N$ à la valeur 1 (fixation), une mutation avantageuse est d'abord apparue dans un type de chromosome (*standard* ou *inversé*, nous n'avons pas actuellement d'hypothèse à ce propos), puis a « migré » dans l'autre configuration. Le processus sélectif s'est donc développé en deux temps.

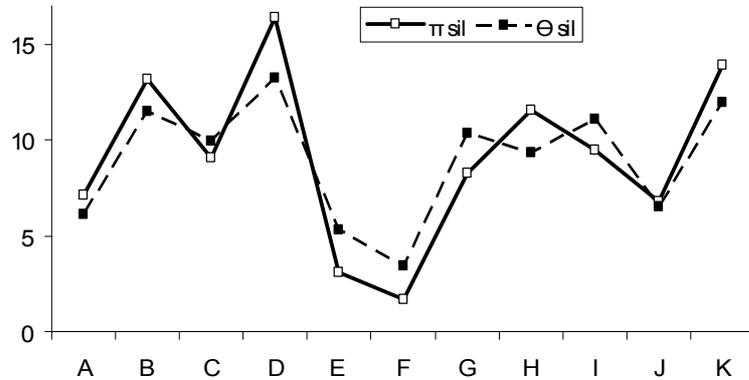


Figure 2: Région *Vha68-1* : valeur observée du polymorphisme nucléotidique silencieux le long de la carte chromosomique, dans la région de *Vha68-1*. π : estimateur de la diversité nucléotidique selon Nei ; θ : estimateur du paramètre mutationnel selon Watterson.

La manière dont l'allèle avantageux est passé d'une configuration à l'autre est plus visible encore sur la figure 4. Cette figure résume l'information portant sur 354 positions polymorphes réparties sur onze gènes. Cette figure fait apparaître de longs haplotypes s'étendant sur l'ensemble de la zone étudiée, donnant naissance à un important « déséquilibre de liaison ». Cependant, au niveau du locus F, cette figure est interrompue par l'apparition de séquences communes à l'ensemble du jeu de données, y compris entre les chromosomes *standard* et *inversé*. Cela montre que la migration du balayage sélectif consécutif à un conversion génique a permis à ces chromosomes de partager l'haplotype avantageux. On peut exclure un cas de double crossing-over sur une région courte, car de tels mécanismes n'existent pas chez la drosophile, par suite du phénomène « d'interférence » : tout événement de crossing-over supprime la possibilité d'un deuxième événement sur une portion chromosomique très longue, de l'ordre d'une mégabase. Ici, la conversion génique porte sur quelques *kb*.

Interactions génomiques et sélection

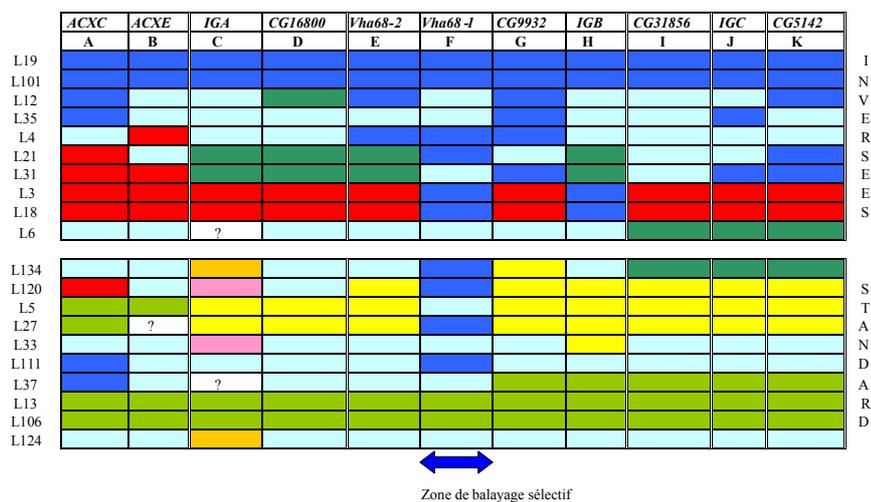


Figure 4 : Région *Vha68-1* : représentation symbolique de la continuité des haplotypes des différents locus marqueurs séquencés. Noter l'importance de la conversion génique trouvée au locus E. Des cas de conversions géniques s'observent aussi aux locus A, et I-K.

Tableau II : Variabilité silencieuse, test de Tajima et F_{ST} pour la région *Vha68-1*.

Gene region	Marker locus	π Silent**	Θ Silent**	Tajima's D	F_{ST}	p -value
<i>ACXC</i>	A	7,13	6,09	0,64883	0,1035	0,0592
<i>ACXE</i>	B	13,19	11,52	0,50054	0,1201	0,0104
<i>IGA</i>	C	9,06	9,97	-0,36317	0,0972	0,0054
<i>CG16800</i>	D	16,4	13,25	0,14083	0,0778	0,0125
<i>Vha68-2</i>	E	3,12	5,33	-1,56106 *	0,1333	0,0027
<i>Vha68-1</i>	F	1,7	3,44	-1,88349 *	0,0058	0,3294
<i>CG9932</i>	G	8,28	10,37	-0,79069	0,2778	0,0002
<i>IGB</i>	H	11,57	9,37	0,91814	0,1778	0,0010
<i>CG31856</i>	I	9,46	11,12	-0,66572	0,0797	0,0216
<i>IGC</i>	J	6,81	6,49	0,19115	0,1358	0,0034
<i>CG5142</i>	K	13,9	11,94	0,65824	0,1807	0,0015

* : $p < 0,05$.

** ($\times 10^3$)

3.2. Région génomique *Su(H)*

Les résultats bruts sont montrés sur le tableau III. La figure 3 montre la carte de la région centrale (séquencée sur 180 kb). La figure 4 montre l'interprétation des profils de polymorphisme de séquence en haplotypes sur 850 kb. On note sur la figure 5 que la région est composée de très longs haplotypes, indiquant le très petit nombre de recombinaisons.

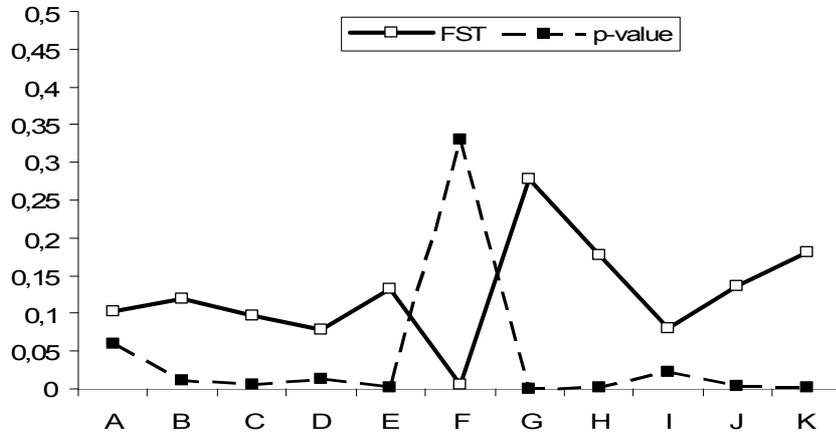


Figure 3 : Région *Vha68-1* : différenciation de l'échantillon en deux sous-échantillons de 10 séquences chacun, sur la base de l'appartenance à l'une des deux catégories chromosomiques, standard ou inversé. En trait continu : valeur du F_{ST} . En tireté : valeur de la probabilité. Le pointillé représente le seuil $p = 0,05$.

Les résultats du tableau III permettent de retrouver la zone de balayage sélectif signalée distalement par rapport à *Su(H)* (vers la droite sur la figure 6). Plusieurs gènes semblent affectés par le balayage sélectif. Un test haplotypique [5] permet de mettre en évidence un écart significatif à la neutralité sélective sur plusieurs gènes entre *Su(H)* et *IG2* (zone intergénique n° 2, fig. 7). C'est aussi dans cette région que sont observées trois conversions géniques et un événement de recombinaison qui semble être un crossing-over. Ces résultats sont très similaires à ceux de la région *Vha-68-1*, sauf pour le fait que l'évènement de balayage sélectif de cette région semble très récent, car les haplotypes issus de la conversion génique sont identiques, sans accumulation de mutations. De ce fait, l'identification de la région sélectionnée semble plus difficile, en raison de l'absence prévisible de polymorphisme à ce niveau.

Tableau 3. Statistiques descriptives, variabilité silencieuse, test de Tajima, et test haplotypique K pour la région $\Delta\mu(H)$.

Gene region	Marker locus	Position kb	interval kb	Lbp	θ using π	θ using S	π	π silent	Div mel /strn	Tajima's D	Test K $r = 0$	id, $r = 32 \times 10^{-10}$
<i>Adh</i>	A	0	0	998	0,0056	0,0056	0,0072	0,0131	0,0238	1,088	0,9426	0,6237
<i>CG18636</i>	B	164	164	-	0,0092	0,0093	0,0071	0,0098	0,0619	-0,841	0,6094	0,2207
<i>Mtr35Bb*</i>	C	262	98	959	0,0081	0,0081	0,0086	0,0086	0,0867	0,221	0,6102	0,1902
<i>Mel</i>	D	336	94	799	0,0100	0,0100	0,0084	0,0090	0,0476	-0,621	0,4007	0,1
<i>G4B.A</i>	E	408	52	-	0,0199	0,0196	0,0158	0,0182	0,0760	-0,836	0,4063	0,0061*
<i>Yellow-C</i>	F	411,6	3,6	749	0,0173	0,0173	0,0190	0,0190	0,0772	0,391	0,1401	0,12
<i>Sr(H)</i>	G	415	3,4	989	0,0128	0,0130	0,0154	0,0201	0,0716	0,872	0,0341*	0,0012*
<i>Sr(H)-CK</i>	H	416,6	1,6	1175	0,0047	0,0047	0,0053	0,0053	0,0346	0,507	0,3373	0,0218*
<i>CK3'</i>	I	418,2	1,3	890	0,0107	0,0104	0,0130	0,0372	0,0396	0,855	0,3	0,0448*
<i>CK3'</i>	J	423,2	5	531	0,0129	0,0129	0,0148	0,0144	0,0756	0,576	0,1813	0,0384*
<i>V43A</i>	K	437,2	14	-	0,0044	0,0044	0,0045	0,0045	0,0814	0,078	0,304	0,185
<i>CG15270</i>	L	461,2	24	-	0,0084	0,0084	0,0090	0,0098	0,0914	0,316	0,5181	0,14
<i>IG1</i>	M	480,7	19,5	970	0,0042	0,0042	0,0050	0,0050	0,0994	0,678	0,1737	0,0061
<i>IG2</i>	N	499,7	19	912	0,0043	0,0043	0,0046	0,0046	0,0457	0,031	0,98	0,7453*
<i>IG3</i>	O	519,2	19,5	1038	0,0060	0,0060	0,0065	0,0065	0,0310	0,373	0,6816	0,2112
<i>CGH168</i>	P	535,7	16,5	1072	0,0140	0,0137	0,0123	0,0145	0,0628	-0,438	0,3668	0,0858
<i>nkl</i>	Q	681,7	146	1005	0,0128	0,0130	0,0094	0,0099	0,0654	-1,021	0,2828	0,055
<i>Tam17b2</i>	R	846,7	165	841	0,0061	0,0061	0,0068	0,00685	0,0377	0,414	0,75	0,31

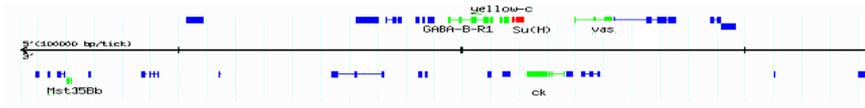


Figure 5 : Région *Su(H)* : carte génétique de la région entourant le locus.

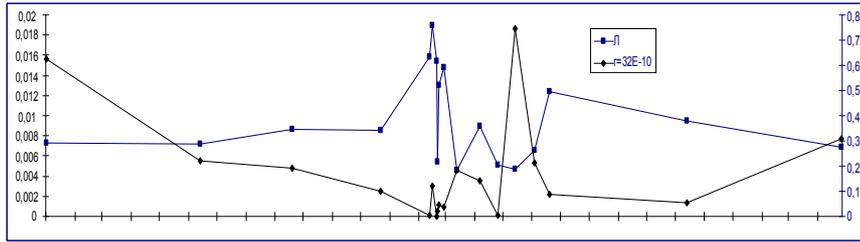


Figure 6 : Profils de variabilité génétique (π) et valeur du test haplotypique K sur l'ensemble de la région étudiée autour de *Su(H)*.

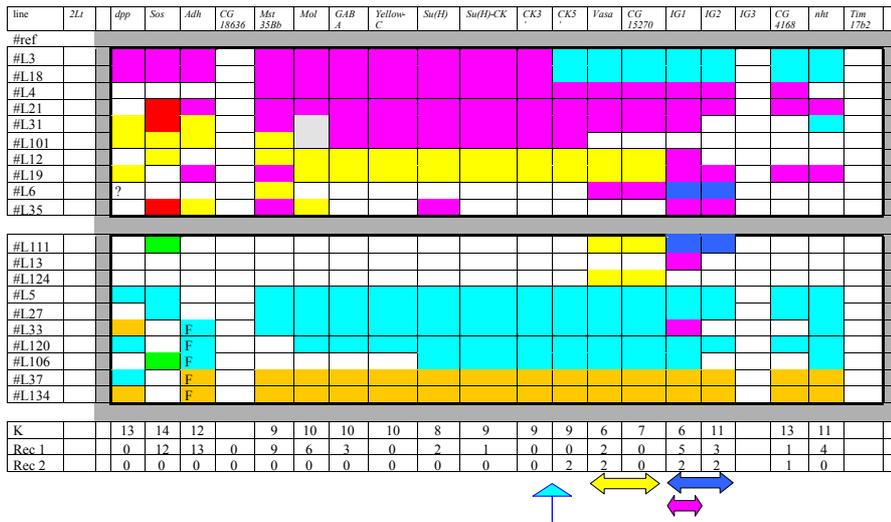


Figure 7 : Distribution des haplotypes sur la région étudiée autour de *Su(H)*.

4. CONCLUSION

Dans de nombreux groupes d'organismes, le génome est polymorphe pour des inversions chromosomiques. Le fait est bien documenté chez les drosophiles [11] et les primates [12], où les données de comparaison génomique ou caryotypique montrent que de très nombreuses inversions se sont produites au cours de l'évolution des espèces. Ces données permettent de supposer que les populations ont présenté des polymorphismes d'inversions au cours de l'évolution, et que de nombreuses espèces

présentent actuellement de tels polymorphismes. Ces polymorphismes ne sont généralement détectables que par des études longues et difficiles. Ils restent donc mal connus. De même reste mal connu le phénomène de conversion génique, au contraire de l'autre mécanisme de recombinaison, qu'est le phénomène de crossing-over. Pourtant, des données de la littérature suggèrent que la quantité de recombinaisons occasionnées par les deux mécanismes est du même ordre de grandeur. *D. melanogaster* s'est révélée un remarquable matériel pour étudier les mécanismes de conversion génique [3], [19]. Ces phénomènes ont également été étudiés chez *D. pseudoobscura* en tirant avantage des inversions naturellement polymorphes [20]. Dans un cas au moins, une conversion génique s'est produite à l'intérieur d'une zone de balayage sélectif [13], bien que dans un contexte différent du cas présent, puisque l'évènement s'est produit à la fin du processus, et non pas pour le déclencher dans un arrangement chromosomique.

Dans ce travail, nous avons montré que la conversion génique est un moyen efficace de résoudre les phénomènes de surdominance associative, en présence de polymorphisme d'inversion. En l'absence de recombinaison, ce phénomène est susceptible d'entraîner de très fortes pressions de balayage sélectif sur de larges portions du génome, et d'occasionner une sorte de « conflit » entre des locus situés sur le même chromosome et polymorphes pour des allèles donnant prise à la sélection.

Le phénomène observé dans notre étude devrait avoir deux conséquences sur l'évolution génétique des populations. En premier lieu, il confère un avantage aux génomes qui ont un taux significatif de conversion génique. Il pourrait expliquer l'évolution de cette propriété universelle des systèmes génétiques. En second lieu, il introduit un temps de latence supplémentaire dans la fixation d'un allèle avantageux. Rappelons que, selon la génétique des populations théorique, un couple d'allèles suivant un modèle de sélection génique avec un avantage sélectif s par allèle avantageux, ce dernier a une probabilité $k = 2s$ de survivre, et que sa fixation se fait en principe au terme d'un délai $t = \log(2N/s)$. Cependant, le processus sélectif peut être suspendu dans le cas d'un polymorphisme d'inversion. Dans ce cas, il faut encore attendre un temps $t = [2c(N-I)]^{-1}$. Dans cette formule, c est la probabilité qu'un nucléotide fasse l'objet d'une conversion génique, et I est la fréquence de l'inversion.

Nos résultats sont une bonne nouvelle pour les programmes de sélection dans les espèces pouvant comporter des polymorphismes d'inversion, puisqu'ils montrent que certains mécanismes de recombinaison permettent de contrecarrer le phénomène de surdominance associative, dont l'effet est de diminuer la réponse à la sélection. Ils montrent cependant qu'un tel phénomène est très dépendant de la taille de la population utilisée. Il

pourrait être l'un des facteurs expliquant le phénomène encore paradoxal rencontré en sélection expérimentale, montrant que la sélection est plus efficace dans les populations de grande taille [21]. Une conséquence nous amenant à faire un distinguo par rapport à nos hypothèses de départ est que, dans le cas considéré, ce n'est pas seulement la variabilité génétique mais aussi la taille absolue de la population qui joue sur la dynamique de la sélection expérimentale, et cela est assez inattendu en génétique des populations. Dans la pratique de la sélection expérimentale, la conséquence pourrait en être de travailler systématiquement sur des populations de grande taille.

Nous remercions les rapporteurs anonymes de ce travail pour leurs commentaires judicieux.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Barton, N. H. 2000. Genetic hitch-hiking. *Phil Trans. R. Soc. London B.* 355: 1553-1562
- [2] Bénassi, V., F. Depaulis, G. K. Meghlaoui, M. Veuille. 1999. Partial sweeping of variation at the *Flp2* locus, in a highly recombining region of the *Drosophila melanogaster* genome. *Molecular Biology and Evolution* 16: 347-353.
- [3] Chovnick A., 1973 Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. *Genetics* 75: 123-31
- [4] Depaulis, F., S. Mousset, M. Veuille. 2001. Haplotype tests using coalescent simulations conditional on the number of segregating sites. *Molecular Biology and Evolution* 18: 1136-1138.
- [5] Depaulis, F., M. Veuille. 1998. Neutrality tests based on the distribution of haplotypes under an infinite site model. *Molecular Biology and Evolution* 15: 1788-1790.
- [6] Depaulis, F., L. Brazier, M. Veuille. 1999. Selective sweep at the *Drosophila melanogaster* *Suppressor of Hairless* locus and its association with the *In(2L)t* inversion polymorphism. *Genetics* 152 : 1017-1024.
- [7] Depaulis, F., L. Brazier, S. Mousset, A. Turbé, M. Veuille. 2000. Selective sweep near the *In(2L)t* inversion breakpoint in an African population of *Drosophila melanogaster*. *Genetical Research* 76: 149-158.
- [8] Depaulis, F., S. Mousset, M. Veuille. 2003 Power of neutrality tests to detect bottlenecks and hitchhiking. *Journal of Molecular Evolution* 57: S190-S200.
- [9] Depaulis, F., S. Mousset, and M. Veuille. 2004 Detecting selective sweeps with haplotype tests. In: Nurminsky (eds). *Selective sweeps*, Landes bioscience.
- [10] Derome, N., K. Métayer, C. Montchamp-Moreau, M. Veuille. 2004 Signature of selective sweep associated with the evolution of *sex-ratio* drive in *Drosophila simulans*. *Genetics*, 166: 1357-1366.
- [11] Drosophila 12 genomes consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203-218.

- [12] Dutrillaux B. 1979 Chromosomal evolution in primates: tentative phylogeny from *Microcebus murinus* (Prosimian) to man. Hum. Genet. 48: 251-314.
- [13] Glinka S, De Lorenzo D, Stephan W. 2006 Evidence of gene conversion associated with a selective sweep in *Drosophila melanogaster*. Mol Biol Evol. 23:1869-1878.
- [14] Hill, W. G., A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genetical Research. 8: 269-274.
- [15] Kimura, M., Ohta, T. 1971. Theoretical aspects of population genetics. Princeton University Press, Princeton NJ.
- [16] Maynard-Smith, J., Haigh, J., 1974. The Hitch-hiking effect of a favorable gene. Genet. Res. 23: 23-35
- [17] Mousset, S., L. Brazier, M-L Cariou, F. Chartois, F. Depaulis, M. Veuille. 2003 Evidence of a high rate of selective sweeps in African *Drosophila melanogaster*. Genetics. 163: 599-609.
- [18] Mousset, S., N. Derome, M. Veuille. 2004. A test of neutrality and constant population size based on the mismatch distribution. Molecular Biology and Evolution 21: 724-731.
- [19] Navarro A, Betrán E, Barbadilla A, Ruiz A. 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics 146: 695-709.
- [20] Schaeffer SW, Anderson WW. 2005 Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics 171: 1729-1739
- [21] Weber K.E., Diggins LT. 1990. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in *Drosophila melanogaster* at two population sizes. Genetics. 125: 585-597.

Contribution des éléments transposables à la création de diversité et à l'adaptation du maïs - une étude pilote

Tatiana ZERJAL⁽¹⁾, Christine DILLMANN⁽¹⁾, Alain CHARCOSSET⁽¹⁾,
Karine ALIX⁽¹⁾, Corinne MHIRI⁽²⁾, Marie-Angèle GRANDBASTIEN⁽²⁾,
Maud TENAILLON^{(1)*}

⁽¹⁾INRA/Univ. Paris-Sud/CNRS/AgroParisTech, UMR8120 de Génétique Végétale
91190 Gif-sur-Yvette, France

⁽²⁾INRA, Laboratoire de Biologie Cellulaire, Institut Jean-Pierre Bourgin
78026 Versailles cedex, France

Abstract. How did transposable elements contribute to maize diversity and adaptation? – a pilot study. In order to evaluate the contribution of transposable elements (TEs), more specifically MITEs (Miniature Inverted Transposable Element), to the diversity and evolutionary history of maize, we developed SSAP markers based on 6 MITE families. These markers reveal polymorphism of insertion/deletion associated with the MITE families understudied and were applied to a collection of maize inbred lines and teosintes (wild relatives of maize). Because of their recent activity, as evidenced by our bioinformatic survey, their insertion near genic regions and their high level of polymorphism, MITE-based markers appear as excellent candidates to screen genetic resources collections for both neutral and adaptive variability. Although our results are somehow preliminary, we did not find any evidence for a burst of amplification associated with the domestication process for any of the MITE families. After cloning and sequencing many of these markers, we made two observations: first most of the polymorphism is generated by the insertion/deletion of elements; second, a co-migration of bands of similar size amplifying different loci is commonly observed. Ideally, each band should be converted into a specific-locus PCR marker. We are in the process of converting our markers and we will use them for screening a broad collection of maize landraces in order to depict the evolutionary processes that accompanied maize diffusion from its centre of origin. A second goal of this project was to investigate the role of TEs in the creation of variability using a material from a divergent selection experiment. This experiment started from two maize inbred (F252 and MBS847) lines. Within each line, two populations, late and early flowering were derived by selection on the phenotype at each generation. After six years of divergent selection on supposedly initial homozygotes individuals (maize inbreds), a strong response to selection was observed. Our working hypothesis was that TEs might have contributed to the generation of variability and allowed a rapid response to selection. Originally, we were planning on using SSAP markers to screen this material. However,

* Correspondance et tirés à part : tenailon@moulon.inra.fr

because of technical problems encountered during the development of SSAP markers, we undertook an RFLP assay by using candidate genes for flowering time as probes. Our results reveal RFLP polymorphism for one of the candidate gene in the late population derived from F252. The polymorphism did not occur during the time of experiment (6 generations) but instead pre-existed in F252 as residual heterozygosity. Using gene dropping simulations in an association mapping framework, we showed that this polymorphism associates significantly with variation in flowering time. We are now in the process of validating this association.

SSAP/ diversity/ maize/ flowering time/ selection

Résumé : Afin d'évaluer le rôle des éléments transposables dans la diversité génétique du maïs et dans son évolution, nous avons développé des marqueurs (SSAP) permettant de révéler du polymorphisme d'insertion de MITE (Miniature Inverted Transposable Element), sur une collection de lignées de maïs et de téosintes. Nos résultats montrent que de par leur activité récente, leur préférence d'insertion dans les régions géniques, et leur niveau de polymorphisme très élevé, les MITE sont d'excellents candidats pour décrire la variabilité neutre et adaptative de collections de ressources génétiques. Nous n'avons pas pu mettre en évidence de phénomènes d'amplification massive d'éléments au cours de la domestication pour les 6 familles de MITE étudiées. Enfin, nous soulignons l'importance de vérifier l'allélisme des bandes révélées. Le rôle des éléments transposables dans la création de variabilité a également été évalué sur un matériel de sélection divergente pour le caractère de précocité de floraison. Nous avons employé une approche de type RFLP en utilisant des sondes de gènes candidats à la précocité. Nos résultats montrent l'existence d'hétérozygotie résiduelle dans une lignée de départ pour l'une des sondes testées. Ce polymorphisme ségrège dans une des populations issues de la sélection divergente et nous avons montré, par simulations, qu'il est associé à la variation de précocité de floraison.

transposon display/ diversité/ maïs/ précocité de floraison/ sélection

1. INTRODUCTION

Connaître les mécanismes générateurs de diversité génétique chez une espèce et appréhender comment la sélection naturelle ou artificielle façonne cette diversité présente un intérêt majeur pour la gestion et la valorisation des ressources génétiques. Dans ce contexte, la domestication des plantes cultivées apparaît comme un modèle de choix car il s'agit d'une expérience *in vivo* d'évolution au cours de laquelle l'adaptation en réponse à une pression de sélection humaine a conduit à l'apparition rapide (quelques centaines ou milliers d'années) de nouvelles formes. Par ailleurs les apparentées sauvages dont dérivent les formes cultivées sont encore trouvées en habitat naturel.

La domestication de l'espèce modèle maïs (*Zea mays* ssp. *mays*) à partir de téosintes annuelles s'est accompagnée d'une complète modification de l'architecture de la plante. Sa domestication a eu lieu au Mexique il y a environ 7 500 ans à partir de téosintes annuelles de la sous-espèce *parviglumis* [15]. La sélection du maïs s'est poursuivie par l'adaptation de populations traditionnelles à des conditions climatiques variées, et son introduction en Europe s'est faite à partir de plusieurs origines américaines [20]. L'utilisation de marqueurs moléculaires (microsatellites et SNP) a permis de quantifier la perte globale de diversité liée à la domestication. Cette perte de diversité résulte de la sélection de gènes du syndrome de domestication [6], [31], [26], [32], et plus généralement, à l'échelle du génome, d'un effet fondateur ou goulot d'étranglement [27]. Enfin, la mise en évidence d'associations entre le polymorphisme moléculaire de certains gènes et la variation phénotypique [21], [28], permet de décrypter les mécanismes génétiques qui sous-tendent l'adaptation à une pression de sélection humaine.

Jusqu'à présent, très peu d'études ont été entreprises pour évaluer la contribution des éléments transposables (ET) à la diversité génétique des maïs actuels. De la même manière, le rôle qu'ils ont pu jouer dans le processus de domestication ou plus généralement dans la création de variabilité n'a encore jamais été étudié. Les ET sont pourtant les constituants majeurs du génome du maïs puisque des études ont estimé à plus de 75 % [23], [29], leur contribution génomique. De par leur propriété de mobilité ils sont une source de diversité génétique importante. Des études basées sur le séquençage de BAC de lignées de maïs, ont montré un polymorphisme d'insertion d'ET extrêmement élevé [7]. Cette diversité peut être neutre ou bien sélectionnée. En effet, si la plupart des éléments sont inactifs et localisés dans des régions hyperméthylées, l'insertion de copies dans des gènes ou dans des régions régulatrices peut engendrer des modifications de l'expression des gènes et des modifications de fonction [2]. Ainsi, certaines copies sont régulées par des mécanismes épigénétiques (méthylation) et pourraient influencer l'expression des gènes adjacents [13]. Certains travaux suggèrent leur rôle dans les processus adaptatifs tels que l'adaptation aux conditions climatiques [11]. Enfin, la transposition des ET peut être activée en réponse à des stress [9], à des événements d'allopolyploidie [12] ou même au cours de la domestication, comme cela a été montré récemment chez le riz pour l'élément *mPing* [17].

On distingue plusieurs classes d'éléments transposables, en fonction de leur mode de transposition [33]. Les éléments de classe I, ou rétrotransposons, se répliquent par l'intermédiaire d'un ARN. Cette classe d'élément est notamment représentée par les rétrotransposons à LTR (Long Terminal Repeat) et les rétrotransposons sans LTR. Chez le maïs, la grande majorité de l'ADN répété est constituée de rétrotransposons. En effet, du

fait du mécanisme de rétrotransposition qui permet de conserver la copie-mère, les retrotransposons se transposent par un mécanisme de type copier-coller, évoluent souvent par ‘bursts’ d’amplification (*e.g.* chez le maïs [24] ; chez le riz, [30]) et sont en général présents en de très nombreuses copies dans le génome des plantes. Chez le maïs, ils ont tendance à s’insérer de façon emboîtée les uns dans les autres dans les régions intergéniques hétérochromatiques [29], [7], [16], créant des sortes d’îlots d’éléments transposables dont la position peut varier fortement d’une lignée de maïs à l’autre [7]. Les éléments de classe II assurent leur cycle de transposition par excision directe de leur ADN à l’aide d’une transposase, sans l’intermédiaire d’un ARN. Cette classe d’éléments comprend notamment les transposons (comme par exemple les éléments *Mutator* ou *Mu-like* du maïs), les MITEs et les héliçons [33]. Les transposons, à l’inverse des rétrotransposons, transposent donc par un mécanisme de type couper-coller, et sont en général en nombre réduit, moins de 100 copies par génome. Les MITEs sont pour cela une exception, puisqu’ils sont présents en un grand nombre de copies (de 500 à 10 000 par génome haploïde). Ce sont de petits éléments (<500 pb) non-autonomes ; l’hypothèse la plus répandue de leur origine, supportée par diverses études moléculaires, serait qu’ils dériveraient par mutation d’éléments transposables autonomes [4]. Ils se rencontrent préférentiellement dans des régions géniques [29] ; de façon intéressante, des études récentes ont montré que certains MITEs étaient encore actifs [35]. Ces caractéristiques font des MITEs de très bons candidats pour la réponse adaptative, puisqu’ils peuvent potentiellement, du fait de leur association avec des gènes, avoir des conséquences directes sur le phénotype et donc la valeur sélective des individus.

Le projet en cours est le fruit d’une collaboration entre (1) le laboratoire de Biologie Cellulaire de l’INRA-Versailles, spécialisé dans l’étude de l’impact des éléments transposables sur l’évolution des génomes végétaux (Solanacées en particulier), et (2) l’équipe GEAR de l’UMR de Génétique Végétale du Moulon, qui travaille sur l’impact des effets sélectifs et démographiques associés au processus de domestication et d’adaptation du maïs. Ce projet constitue une étude pilote visant à développer une technique de détection de polymorphisme d’insertions d’éléments transposables (SSAP) de type MITE chez le maïs afin de mieux appréhender la contribution des ET au potentiel adaptatif de cette espèce. Dans un premier temps, la SSAP a été appliquée à un échantillon restreint de maïs cultivés comprenant des lignées élites d’origine américaine et européenne et des téosintes, afin d’examiner si la domestication et les étapes ultérieures de sélection sont associées à une modification spécifique de certaines familles d’ET qui auraient joué un rôle dans l’adaptation du maïs à de nouveaux milieux (volet 1 du projet). Par ailleurs, nous avons aussi cherché à voir si les ET sont responsables de la création de nouvelle variabilité. Pour cela, nous

avons étudié une descendance issue d'une expérience de sélection divergente mise en place par C. Dillmann et A. Charcosset (équipes GQF et GQMS de l'UMR GV du Moulon) sur le caractère de précocité de floraison. A partir d'une lignée homozygote de départ et après 6 générations de sélection, deux populations ont été obtenues, l'une précoce et l'autre tardive décalées de 10 jours à la floraison. Afin d'expliquer cette réponse rapide à la sélection (6 générations) d'un matériel fixé au départ, l'hypothèse que nous avons examinée concerne la mobilisation d'ET et leur insertion en de nouveaux points du génome (volet 2 du projet).

Cette étude pilote donne des résultats tout a fait encourageants quant à l'utilisation de ces marqueurs sur un échantillon étendu de maïs cultivés permettant d'évaluer une large collection de ressources génétiques; les marqueurs développés sont également en cours d'utilisation sur le matériel de sélection divergente. Ce travail ouvre de nouvelles voies d'exploration pour l'utilisation de ces marqueurs en génétique d'association.

2. MATÉRIEL ET MÉTHODES

2.1. Matériel Végétal

Volet 1. Un panel de 26 lignées de maïs, dont 13 d'origine européenne et 13 d'origine américaine, ainsi que 15 téosintes a été utilisé.

Volet 2. Afin d'explorer le rôle des éléments transposables dans la création de variabilité génétique, nous étudions deux descendance issues de deux expériences de sélection divergente menées en parallèle depuis 1996 sur le caractère de précocité de floraison. Deux lignées de départ, l'une précoce (F252) et l'autre tardive (MBS847), ont été choisies. A partir de chacune de ces lignées, deux populations l'une précoce et l'autre tardive ont été sélectionnées pendant 6 générations. A chaque génération, le matériel a été évalué dans un dispositif de blocs randomisés, contenant les 10 familles de 100 individus correspondant à 10 individus sélectionnés. Par ailleurs, au cours des étés 2004 et 2005, une expérience d'évaluation de l'ensemble du matériel (toutes les générations) a été mise en place par C. Dillmann. Nous avons représenté dans la figure 1, les pedigrees des individus de la génération 2002 pour les populations tardives et précoces issues des deux lignées de départ. Des graines issues de l'autofécondation de chaque plante sélectionnée au cours de l'expérience sont stockées en chambre froide et sont disponibles pour les analyses génétiques.

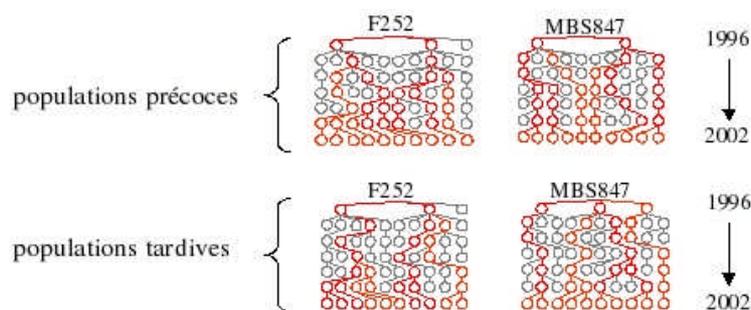


Figure 1 : Pedigrees des individus issus de l'expérience de sélection divergente sur le caractère de précocité de floraison. L'expérience a été conduite à partir de deux lignées (F252 et MBS847) dont sont dérivées des populations précoces et tardives. Les individus représentés par des cercles rouges sont ceux qui, dans la généalogie, ont contribué à la génération 2002, autrement dit, ceux qui ont laissé des descendants. Les autres sont représentés par des cercles gris.

2.2. Mise au point des marqueurs SSAP

Afin de révéler du polymorphisme d'insertion/délétion, nous avons développé la technique SSAP (fig. 2). Cette technique consiste en une étape de double digestion du génome (ici *Eco*RI and *Msp*I notés R1 et R2 dans la figure 2), suivie d'une étape de ligation d'adaptateurs complémentaires des sites de restriction. Nous procédons ensuite à une pré-amplification avec des amorces complémentaires des adaptateurs, puis à une amplification dite sélective à l'aide d'une amorce marquée, définie spécifiquement dans une famille d'éléments transposables (ET) et d'une amorce anonyme ancrée dans l'un des adaptateurs. Les amorces choisies pour la pré-amplification et l'amplification sélective peuvent contenir 1 à 3 bases sélectives (notées +N) qui permettent de révéler un nombre de produits d'amplification compatible avec la lecture du gel d'acrylamide sur lequel les produits de PCR sont déposés et séparés par électrophorèse. L'utilisation de plusieurs combinaisons de bases sélectives permet d'avoir accès à l'ensemble de l'information. Par ailleurs le marquage en fluorescence ou en radioactivité de l'amorce spécifique de la famille d'éléments transposables assure une révélation de polymorphisme associé soit à l'insertion/délétion d'un élément de la famille (①), soit à une mutation ponctuelle dans le site de restriction (②), soit à une insertion/délétion entre le site d'insertion de l'ET et le site de restriction (③). Les fragments provenant de l'amplification entre 2 sites de restriction ne sont pas révélés.

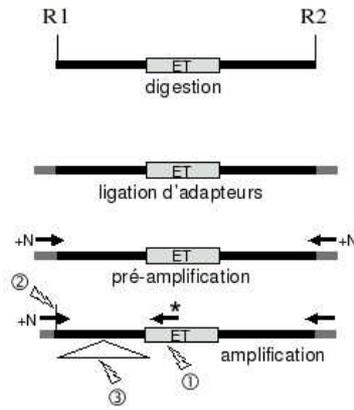


Figure 2 : Schéma de la technique SSAP employée. Trois sources de polymorphisme sont à considérer : ① insertion/délétion d'un ET, ② mutation ponctuelle dans le site de restriction, ③ insertion/délétion entre le site d'insertion de l'ET et le site de restriction.

Grâce à une étude bibliographique sur la datation de l'activité de transposition et les régions préférentielles d'insertion des ET, nous avons retenu pour notre étude 6 familles de MITEs, potentiellement actifs, qui s'insèrent à proximité de régions géniques, et dont certains semblent impliqués dans la variation phénotypique, et un rétrotransposon à LTR dont l'insertion dans le génome est très ancienne. Les familles d'ET étudiées sont: (i) une famille de rétrotransposons à LTR non fonctionnels, *Grande*, présentant beaucoup de polymorphisme inter-lignées chez le maïs [8] (notée *Grande*); (ii) une famille de MITEs, *Heartbreaker*, dont la transposition dans le génome du maïs est récente [34] (notée *Hbr*); (iii) une famille de MITEs dont plusieurs éléments sont trouvés dans des gènes *b1*, *glu1* codant la β -D glucosidase, *mxy* et *su1*, impliqués dans la chaîne de biosynthèse de l'amidon ainsi que dans un gène codant une peroxydase (*ZmPox3*). Dans ce dernier, la présence de l'élément est associée de façon significative avec le caractère de digestibilité [10] (notée *U19*); (iv) une famille de MITEs, dont l'un des éléments présente une association significative avec la teneur en protéines et la teneur en amidon du grain de maïs (L. Camus, com. pers.). Des éléments de cette famille sont également trouvés aux gènes *b ζ* (élément nommé *Ins2* par Ralston *et al.*, [19]) et *c1*, impliqués dans la voie de biosynthèse de l'anthocyanine, et *y1*, un gène codant la phytoène synthase [18]; (v) la famille des éléments *mPIF* (*miniature P Instability Factor*) dont la transposition est encore active chez le maïs [35] (notée *mPIF*); (vi) une famille de MITEs dont un élément a été trouvé dans le gène *D8*, impliqué dans le déterminisme de la précocité de floraison chez le maïs. La présence du MITE a notamment

été associée à l'haplotype précoce [28], [1] (notée *D8*); (vii) une famille de MITEs dont un élément a été trouvé dans le gène *Vgt1*, impliqué dans le déterminisme de la précocité de floraison chez le maïs. Une association significative entre le polymorphisme de présence/absence du MITE et la variation de la précocité de floraison a été mise en évidence par Salvi *et al.*, [22] (notée *Vgt1*).

Pour certaines des familles retenues, des protocoles SSAP incluant les amorces nécessaires à l'amplification étaient déjà disponibles. Il s'agit de *Grande* [8], *Hbr* [3]. Pour les autres, nous avons réalisé une recherche systématique dans des banques de données (TIGR, maizeGDB, GenBank) afin d'aligner le plus grand nombre possible d'éléments et de dessiner des amorces consensus permettant d'amplifier l'ensemble de la famille. La mise au point de la technique a tout d'abord été effectuée en fluorescence. Cependant, une étape préalable nécessaire à la vérification de la nature du polymorphisme des marqueurs (cf. fig. 2) et à l'allélisme des bandes, consiste à récupérer l'ensemble des marqueurs que nous souhaitons génotyper afin de procéder à leur séquençage. Ceci permet en outre la conversion de ces marqueurs en marqueurs PCR, beaucoup plus faciles d'utilisation et qui pourront alors être utilisés sur un panel de lignées beaucoup plus vaste. Suite à des difficultés techniques, nous avons eu recours à la radioactivité pour le prélèvement de ces marqueurs sur gel d'acrylamide. Ce prélèvement est suivi d'une étape de réamplification (similaire à l'étape d'amplification sélective), puis de clonage et de séquençage. Etant donné la lourdeur de cette démarche, nous ne sommes pas en mesure de présenter les résultats issus du génotypage de l'ensemble des marqueurs sur le panel de lignées. Cependant, nous avons déjà mis au point 50 marqueurs PCR qui seront génotypés et souhaitons, dans un futur proche, arriver à un total de 100 marqueurs.

2.3. Analyse bioinformatique

Afin de connaître plus en détail les caractéristiques des familles de MITEs étudiées, nous avons procédé à une recherche systématique des séquences homologues à la copie initialement caractérisée pour chacune des familles de MITEs retenues. Cette recherche a été effectuée à partir d'une banque de données (maizesequence.org) regroupant l'ensemble des données de séquences obtenues pour la lignée de référence (B73), soit 14 000 séquences de BAC. Nous avons employé un critère d'homologie de 85 % entre les séquences récupérées avec une variation dans la longueur de 10 % maximum. A partir des séquences ainsi recueillies, nous avons utilisé MEGA [25] afin d'obtenir une matrice de distance (distance Kimura-2 paramètres) et construire un arbre de distance selon la méthode UPGMA (fig. 3). L'objectif n'est pas de faire des inférences phylogénétiques mais de visualiser les distances entre les différentes copies d'une même famille de MITEs et

éventuellement de repérer des groupes de séquences identiques provenant d'un événement d'amplification récent.

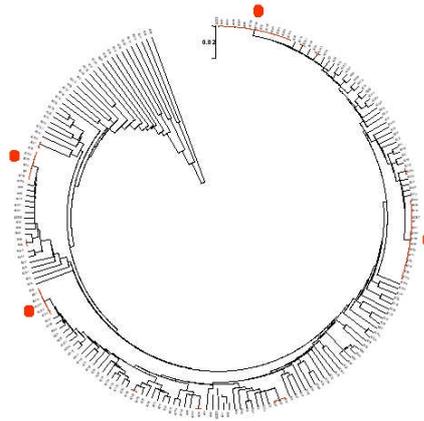


Figure 3 : Arbre de distance obtenu avec 205 éléments de la famille *Vgt1*. Sont signalés par un point rouge, plusieurs groupes de séquences identiques provenant d'événements d'amplification récents.

2.4. Sélection divergente et marqueurs RFLP

Pour ce qui concerne la récupération de bandes SSAP, nous avons été confrontés à des problèmes techniques importants, ce qui ne nous a pas permis d'appliquer cette technique au matériel de sélection divergente dans le temps imparti. Nous avons donc entrepris, en première approche, d'hybrider sur le matériel de sélection divergente, des sondes de gènes candidats à la précocité de floraison. Huit sondes cDNA préalablement cartographiées et identifiées comme de bons candidats [5] ont été utilisées. Chacune des sondes a été hybridée sur l'ADN génomique digéré de 36 individus : 1 individu de F252, 1 de MBS847, 4 individus de première génération correspondant aux individus initiaux de la population tardive et précoce dérivée des lignées F252 et MBS847 et 30 individus de la génération 2002 appartenant à chacune des 4 populations. L'ADN a été digéré par 2 enzymes de restriction, *EcoRI* (site de coupure à 6 pb) et *MboI* (site de coupure à 4 pb). Au cours de cette première expérience, nous avons pu mettre en évidence du polymorphisme de longueur des fragments de restriction (RFLP) pour l'un des gènes candidats à la précocité dans la population tardive issue de F252. Il s'agit du locus révélé par la sonde QCK5e06. Ce polymorphisme peut être dû soit à une mutation ponctuelle dans un site de restriction, soit à l'insertion d'un élément transposable. Cette première phase a été complétée par le génotypage (RFLP) des individus

appartenant aux générations intermédiaires ainsi qu'à la lignée d'origine pour cette population (cf. fig. 1). Le génotypage de la lignée d'origine a été effectué sur 30 individus provenant du sac de graines d'origine de cette lignée.

Afin de tester l'association entre les génotypes et les valeurs génotypiques obtenues pour chaque individu, nous avons utilisé un modèle de génétique d'association pour estimer les valeurs observées des effets d'additivité (a_{obs}) et de dominance (d_{obs}) :

$$T_{ig} = \mu_g + ax_i + dy_i + e_i$$

où T_{ig} est la valeur du caractère pour le génotype i de la génération g ; μ_g est la moyenne de la date de floraison de tous les individus d'une génération, g , auquel le génotype i appartient; a est l'additivité et d , la dominance; x et y sont des variables indicatives; pour un individu homozygote, i , $x_i = 1$ ou -1 , et $y_i = 0$, pour un individu hétérozygote, $x_i = 0$ et $y_i = 1$; e_i est la résiduelle.

L'effet du locus QCK5e06 sur la variation quantitative de la date de floraison a été testé de la manière suivante: nous avons simulé des distributions nulles pour l'additivité (a) et la dominance (d) selon l'hypothèse nulle, H_0 , d'absence d'association après la prise en compte la dérive engendrée par la généalogie. En pratique, nous avons tout d'abord simulé une généalogie d'individus identique à celle de la population F252 tardive, en considérant l'individu F252 initial comme hétérozygote à un locus biallélique. A chaque génération, g , nous avons tiré les génotypes pour chaque individu, comme suit: nous avons considéré 3 classes de génotype, le génotype 0 pour les individus homozygotes pour l'allèle 1, le génotype 2 pour les individus homozygotes pour l'allèle 2, et le génotype 1 pour les hétérozygotes. Pour un individu hétérozygote, les génotypes des descendants ont été tirés avec une probabilité de 0,25 pour le génotype 0, de 0,5 pour le génotype 1, et de 0,25 pour le génotype 2. Pour un individu de génotype 0, l'ensemble des descendants a un génotype 0 et pour un individu de génotype 2, l'ensemble des descendants a un génotype 2. Nous avons donc ignoré les révertants. Ensuite, nous avons estimé les paramètres a et d , en utilisant la matrice des génotypes simulés et celle des phénotypes observés dans le modèle de régression linéaire décrit ci-dessus. Par l'acquisition de milliers de simulations de ce type, il est possible d'obtenir une distribution de a et d et d'attribuer à a_{obs} et d_{obs} une probabilité d'être observé sous l'hypothèse H_0 .

3. RÉSULTATS ET DISCUSSION

3.1. Caractérisation des familles de MITEs

L'analyse bioinformatique de 5 familles de MITEs parmi les 6 étudiées révèle l'existence dans chaque famille de groupes de séquences très divergentes, qui résultent probablement d'événements d'amplification plutôt anciens. En ce qui concerne la famille *Vgt1*, le patron est différent : plusieurs éléments de cette famille sont parfaitement identiques et la topologie de l'arbre obtenu, donné en figure 3, nous laisse penser que cette famille a évolué par des événements d'amplification successifs conduisant à l'émergence d'un nombre limité de copies dans chaque groupe. Certains de ces événements sont récents (copies identiques), étant donné la longueur de ce MITE (133 pb) et le taux de substitution estimé du maïs, la divergence entre ces copies date de moins de 100 000 ans.

3.2. Polymorphisme d'insertion d'éléments transposables

3.2.1. Analyse des profils SSAP

La mise au point de la technique SSAP nous a permis d'obtenir des résultats satisfaisants pour l'ensemble des 7 familles étudiées. Un gel d'acrylamide obtenu en fluorescence pour la famille *Ins2* est présenté dans la figure 4.

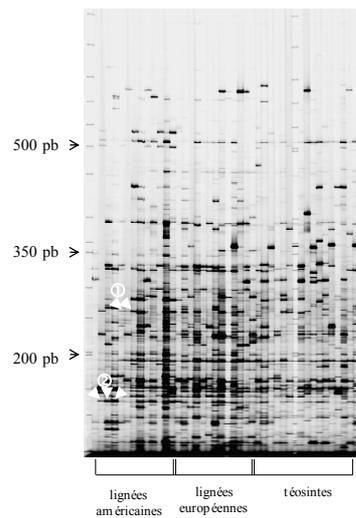


Figure 4 : Photo d'un gel d'acrylamide obtenu en SSAP avec la famille *Ins2*.

En vue d'estimer la structuration du panel de maïs et téosintes étudié sur la base des profils SSAP développés sur un rétrotransposon à LTR *ns*, un

MITE, nous avons exploité 78 niveaux de bandes amplifiées pour le système *Grande* et 68 pour *Ins2*. Le système *Grande* donne en moyenne des profils de 69,5 bandes par génotype alors que celui développé avec *Ins2* en produit 39,4 : le fait que le LTR donne plus de bandes que le MITE est cohérent avec son nombre de copies plus élevé dans le génome (5 100 copies [8] *vs.* quelques centaines). Concernant les sous-espèces, les téosintes présentent significativement plus de bandes en moyenne par génotype (*Grande* : 78,93 ; *Ins2* : 45,51) que les lignées de maïs (*Grande* : 63,94 ; *Ins2* : 35,72). En revanche, l'origine des lignées de maïs, Europe *vs* Amérique, n'influe pas significativement le nombre de bandes SSAP générées. Le gel de la figure 4 illustre la complexité des profils SSAP obtenus, en lien avec le grand nombre de bandes amplifiées. De plus, comme pour tout profil multibande, nous avons voulu vérifier que l'ensemble des bandes amplifiées à une taille donnée (fig. 4 ① et ②, par exemple) correspondait à un même site d'insertion de l'ET étudié, ou en d'autres termes, si ces bandes étaient bien alléliques. En effet, cela pourrait alors conduire à des conclusions fausses quant à la structuration de la diversité génétique générée par les ET si on se basait uniquement sur les données de profils SSAP.

3.2.2. Caractérisation des sites d'insertion des éléments transposables étudiés

En vue d'analyser l'allélisme des bandes SSAP amplifiées, nous nous sommes intéressés à leur séquence nucléotidique. Afin de faciliter la récupération des bandes en vue de leur clonage et séquençage, nous avons utilisé un système de détection radioactif. La récupération de bandes en radioactivité a, pour le moment, porté sur les familles *Grande*, *Vgt1*, *Hbr* et *Ins2*. Nous avons ainsi récupéré 153 bandes et pu en réamplifier 122 ; au final, nous avons réussi le clonage et séquençage pour cent d'entre elles. L'analyse des séquences obtenues révèle que 30 % des bandes sont en réalité constituées d'une population de séquences de même longueur appartenant à des régions différentes du génome, quel que soit le système d'ET étudié. En d'autres termes, il n'est pas rare d'observer des insertions d'ET à différents locus dans le génome qui sont à des distances quasi-identiques de sites de restriction (à quelques paires de bases près) et qui engendrent donc des bandes de longueur similaire. Cette co-migration de bandes de même taille amplifiant des locus différents souligne à quel point il est essentiel de transformer ces marqueurs en marqueurs PCR avant de procéder à une analyse du polymorphisme. Dans ce contexte, 50 % des séquences obtenues ont été converties en marqueurs PCR.

Parmi les 100 bandes séquencées, seules deux présentaient du polymorphisme lié à l'insertion d'un autre élément transposable dans la région amplifiée et une autre présentait du polymorphisme lié à une délétion de 7pb dans la région amplifiée (polymorphisme noté ③ dans la figure 1).

Aucun polymorphisme de site de restriction (noté © sur la figure 1) n'a été trouvé. Les 50 marqueurs PCR développés sont en cours de génotypage sur un panel de 26 populations américaines comprenant 12 individus par population. Ceci nous permettra de valider l'intérêt des marqueurs SSAP pour la caractérisation de collections de ressources génétiques, de caractériser la diversité au niveau intra-populations et inter-populations. Etant donné le niveau de polymorphisme élevé de ces marqueurs, nous pensons que cette analyse sera très utile pour inférer la structure des populations et pour décrypter l'histoire évolutive qui a accompagné la diffusion du maïs à partir de son centre d'origine (migration, effets fondateurs, adaptation locale).

Notre analyse SSAP nous laisse penser qu'aucune des familles étudiées n'a connu d'amplification massive au cours de la domestication, mais celle-ci a quand même pu engendrer des phénomènes d'insertion de MITEs à pouvoir adaptatif intéressant. La variabilité génétique que nous avons mise en évidence pour les MITEs étudiés (au niveau des séquences et des insertions) peut représenter une base de diversité sur laquelle la sélection aurait joué au cours de la domestication puis après, durant les phases d'adaptation. Actuellement, nous cherchons à identifier les quelques insertions de MITEs ayant été l'objet de sélection (avantage sélectif et rôle au niveau de l'expression génique), par rapport aux autres qui seront des marqueurs moléculaires neutres intéressants pour des études de diversité.

3.3. Un candidat pour la précocité de floraison chez le maïs ?

Le génotypage RFLP de la sonde QCK5e06 sur la population F252 tardive (fig. 5), nous a permis de mettre en évidence que le polymorphisme révélé préexiste de manière cryptique (faible fréquence) dans la lignée F252 et que c'est cette hétérozygotie résiduelle de départ qui a pu être sélectionnée. Grâce aux évaluations phénotypiques 2004 et 2005, nous avons pu confronter ces résultats aux données de précocité. Nous avons estimé par un modèle de génétique d'association, les valeurs des coefficients a et d observés, à -1,96 et -0,12. Ce modèle explique 35 % de la variation phénotypique. Les résultats de simulations indiquent que la probabilité d'obtenir dans nos simulations une valeur du coefficient d'additivité égale à celle observée (-1,96) est très faible ($p < 0,001$). Il existe donc une association significative entre ce locus et la variation de la précocité de floraison, et cette association ne résulte pas uniquement de la dérive engendrée par le schéma expérimental (généalogie). Ces résultats sont associés à l'ensemble de l'analyse des données phénotypiques dans une publication en préparation (Dillmann *et al.* en préparation).

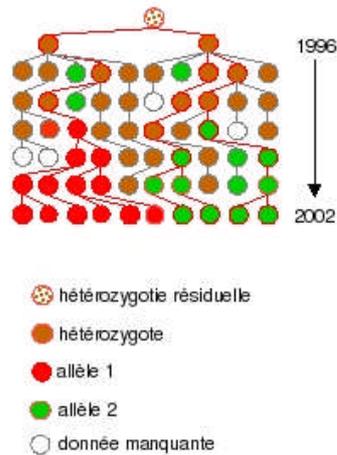


Figure 5. Génotypage RFLP résultant de l'hybridation de la sonde QCK5e06 sur les individus appartenant à la généalogie de la population F252 tardive (cf. Figure 1). De l'hétérozygotie résiduelle a été observée pour la lignée F252 : un individu sur 30 porte un allèle et 29 individus l'autre. Les individus non génotypés ont été traités comme données manquantes dans les simulations.

Afin de valider cette association, nous employons actuellement une démarche classique de détection de QTL. Nous avons ainsi mis en place au champ au cours de l'été 2007, des populations F2 issues de l'autofécondation d'individus hétérozygotes à ce locus et poursuivrons l'analyse par la mise en place des familles F3 durant l'été 2008. Par ailleurs, nous cherchons à identifier la mutation à l'origine du polymorphisme (mutation ponctuelle dans un site de restriction ou insertion/délétion d'un élément transposable). Une étude d'évolution moléculaire pourra également être entreprise pour préciser les patrons de sélection à ce locus. Enfin, l'application de la technique SSAP sur le matériel de sélection divergente est en cours (thèse Eléonore Durand).

REMERCIEMENTS

Cette étude a été financée par le Bureau des Ressources Génétiques (AP 2005-2007) et par l'ANR Jeunes-chercheurs ANR-05-JCJC-0067-01. Nous remercions Johann Joets pour son aide dans l'étude de bioinformatique, ainsi que Céline Ridet pour son aide en hybridation RFLP. Aurélie Bourgeois et Fabrice Dumas ont participé aux extractions d'ADN. Deux stagiaires,

Caroline Esnault et Estelle Bonleu, ont contribué aux aspects techniques de ce projet (hybridation RFLP et mise au point de la SSAP).

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Andersen J.R., Schrag T., Melchinger A.E., Zein I., Lubberstedt T., 2005 - Validation of Dwarf8 polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). *Theor. Appl. Genet.*, 111:206-217.
- [2] Bennetzen J.L., 2000. - Transposable element contribution to plant gene and genome evolution. *Plant Mol. Biol.*, 42:251-269.
- [3] Casa A.M., Mitchell S.E., Smith O.S., Register J.C.I., Wessler S.R., Kresovich S., 2002 - Evaluation of Hbr (MITE) markers for assessment of genetic relationship among maize (*Zea mays* L.) inbred lines. *Theor. Appl. Genet.*, 104:104-110.
- [4] Casacuberta, J. M., Santiago N., 2003 - Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311: 1-11.
- [5] Chardon F., Virlon B., Moreau L., Falque M., Joets J., Decousset L., Murigneux A., Charcosset A., 2004 - Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics*, 168:2169-2185.
- [6] Doebley J., Stec A., Hubbard L., 1997. - The evolution of apical dominance in maize. *Nature*, 386:485-488.
- [7] Fu H., Dooner H.K., 2002 - Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl. Acad. Sci. USA.*, 99:9573-9578.
- [8] Garcia-Martinez J., Martinez-Izquierdo J.A., 2003 - Study of the evolution of the *Grande* retrotransposon in the *Zea* genus. *Mol. Biol. Evol.*, 20:831-841.
- [9] Grandbastien M-A., 1998 - Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.*, 3:181-187.
- [10] Guillet-Claude C., Birolleau-Touchard C., Manicacci D., Rogowsky P.M., Rigau J., Murigneux A., Martinant J-P., Barrière Y., 2004 - Nucleotide diversity of the *ZmPox3* maize peroxidase gene: relationships between a MITE insertion in exon 2 and variation in forage maize digestibility. *BMC Genetics*, 5.
- [11] Kalendar R., Tanskanen J., Immonen S., Nevo E., Schulman A., 2000 - Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA.*, 97:6603-6607.
- [12] Kashkush K., Feldman M., Levy A.A., 2002 - Gene loss, silencing and activation in a newly synthesised wheat allotetraploid. *Genetics*, 160:1651-1659.
- [13] Kashkush K., Khasdan V., 2007 - Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics*, 177:1975-1985.
- [14] Loot C., Santiago N., Sanz A., Casacuberta J.M., 2006 - The proteins encoded by the *pogo*-like *Lem1* element bind the TIRs and subterminal repeat motifs of the *Arabidopsis* Emigrant MITE: consequences for the transposition mechanism of MITEs. *Nucleic Acids Res.*, 34:5238-5246.

- [15] Matsuoka Y., Vigouroux Y., Goodman M.M., Sanchez G. J., Buckler E., Doebley J., 2002. - A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA*, 99:6080-6084.
- [16] Meyers B.C., Tingey S.V., Morgante M., 2001 - Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research*, 11:1660-1676.
- [17] Naito K., Cho E., Yang G., Campbell M.A., Yano K., Okumoto Y., Tanisaka T., Wessler S.R., 2006 - Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl. Acad. Sci USA*, 103:17620-17625.
- [18] Palaisa K.A., Morgante M., Williams M., Rafalski A., 2003 - Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell*, 15:1795-1806.
- [19] Ralston E.J., English J.J., Dooner H.K., 1988 - Sequence of three bronze alleles of maize and correlation with the genetic fine structure. *Genetics*, 119:185-197.
- [20] Rebourg C., Chastanet M., Gouesnard B., Welcker C., Dubreuil P., Charcosset A., 2003. - Maize introduction into Europe: the history reviewed in the light of molecular data. *Theor. Appl. Genet.*, 106:895-903.
- [21] Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S., 2001. - Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA*, 98:11479-11484.
- [22] Salvi S., Sponza G., Morgante M., Tomes D., Niu X., 2007 - Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. USA*, 104:11376-11381.
- [23] SanMiguel P., Tikhonov A., Jin Y-K., Melake-Berhan A., Springer P.S., Edwards K.J., Avramova Z., Bennetzen J.L., 1996. - Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 274:765-768.
- [24] SanMiguel P., Gaut B. S., Tikhonov A., Nakajima Y., Bennetzen J. L. 1998 - The paleontology of intergene retrotransposons of maize. *Nat. Genet.*, 20: 43-45
- [25] Tamura K., Dudley J., Nei M., Kumar S., 2007 - MEGA 4: molecular evolutionary genetics analysis [MEGA] software v4.0. *Mol. Biol. Evol.*, 24:1596-1599.
- [26] Tenaillon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S., 2001. - Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). *Proc. Natl. Acad. Sci. USA*, 98:9161-9166.
- [27] Tenaillon M.I., U'Ren J., Tenaillon O., Gaut B.S., 2004. - Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol. Biol. Evol.*, 21:1214-1225.
- [28] Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S., 2001. - Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.*, 28:286-289.
- [29] Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z., 1999. - Colinearity and its exceptions in orthologous adh regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA*, 96:7409-7414.
- [30] Vitte C., Panaud O., Hadi Quesneville H., 2007 - LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomic* 8:218

- [31] Wang R.L., Stec A., Hey J., Lukens L., Doebley J., 1999. - The limits of selection during maize domestication. *Nature*, 398:236-239.
- [32] Whitt S.R., Wilson L.M., Tenailon M.I., Gaut B.S., Buckler IV E.S., 2002. - Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci. USA.*, 99:12959-12962.
- [33] Wicker T., Sabot F., Hua-Van A., Bennetzen J. L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A., 2007 - A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.*, 8:973-982
- [34] Zhang Q., Arbuckle J., Wessler S.R., 2000 - Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. *Proc Natl Acad Sci USA*, 97:1160-1165.
- [35] Zhang X., Feschotte C., Zhang Q., Jiang N., Eggleston W.B., Wessler S.R., 2001 - P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc. Natl. Acad. Sci. USA*, 98:12572-12577.

II

COMMUNICATIONS EN SÉANCES PLÉNIÈRES

Le séquençage du génome de la vigne dans un projet public franco-italien

The French-Italian Public Consortium for the Sequencing of the Grapevine Nuclear Genome¹

A-F Adam-Blondon²

¹ <http://www.genoscope.cns.fr/spip/Vitis-vinifera-e.html>

² UMR INRA-CNRS-Université d'Evry de Recherches en Génomique Végétale, 2 rue Gaston Crémieux, BP5708, 91057 Evry cedex, France

Malgré l'importance culturelle et économique de la viticulture en France et dans le monde, jusqu'en 2001, les connaissances sur le déterminisme génétique des caractères agronomiques importants étaient extrêmement partielles alors qu'elles avaient explosé durant les 15 années précédentes chez d'autres espèces cultivées comme par exemple la tomate, le maïs ou le riz. En effet, la communauté scientifique était handicapée par la longueur du cycle biologique (au moins trois ans sont nécessaires entre le semis d'un pépin et l'obtention de baies) et surtout ne disposait pas des outils moléculaires nécessaires tels que des cartes génétiques avec des marqueurs transférables, des collections d'ESTs... (Adam-Blondon *et al* 2004).

En 1998, une première étape a été franchie avec la décision d'une vingtaine de laboratoires à travers le monde de s'associer pour développer des marqueurs moléculaires de type microsatellites (SSR) afin de faciliter les études génétiques. Trois ans plus tard, en 2001, les mêmes équipes décidaient de former un consortium international (International Grape Genome Program ou IGGP ; www.vitaceae.org) dont l'objectif était de promouvoir le développement de ressources et d'outils publics pour le décryptage et la compréhension du fonctionnement du génome de cette espèce. Durant l'été 2005, un accord a été signé sous l'égide de l'IGGP entre la France et l'Italie pour joindre les efforts des deux pays dans cet objectif. L'IGGP a donc été très efficace puisqu'en 6 ans la vigne est passée du statut d'espèce sous-étudiée à la quatrième espèce végétale dont le génome a été entièrement séquencé (Jaillon *et al* 2007).

La vigne cultivée, *Vitis vinifera*, comprend 19 paires de chromosomes pour un génome dont la taille avait été estimée à 475 Mb (Lodhi et Reisch 1995). De nombreuses études ont montré que la vigne est une espèce très

hétérozygote (Adam-Blondon *et al* 2004 pour revue) et le consortium a donc décidé de séquencer un individu quasi-homozygote (PN40024) dérivé du Pinot Noir par au moins 6 générations d'autofécondations (Bronner et Oliveira 1990 ; Jaillon *et al* 2007) pour produire une séquence de référence de qualité par la stratégie Whole Genome Shotgun. Les séquences correspondent alors à des séquences d'extrémité de clones provenant de différentes banques (BAC, Fosmides et plasmides) en utilisant la technologie Sanger.

Dans le but d'obtenir une séquence de référence de haute qualité, l'objectif du consortium franco-italien est d'obtenir un ensemble de séquences correspondant à 12 fois (12X) la taille du génome de la vigne (objectif atteint en juillet 2007) et un premier brouillon de qualité a été publié au deux tiers du parcours (8.4X). Cette version de la séquence du génome comprenant 3514 *super-contigs* de séquences assemblés à partir de 6.2 millions de séquences de longueur moyenne de 700 bases obtenues par 3 plates-formes (le Génoscope en France, l'IGA et l'Université de Vérone en Italie) et assemblées à l'aide d'ARACHNE au Génoscope (Jaillon *et al* 2007). La moitié des *super-contigs* a une taille supérieure à 2.07 Mb. Cette première version de la séquence du génome couvre 487Mb et permet déjà à la communauté scientifique de commencer à exploiter la séquence du génome grâce à sa mise à disposition dans les bases de données internationales et les deux sites web du projet (www.genoscope.cns.fr/vitis et www.vitisgenome.it). Pour l'assemblage de la version finale 12X, deux logiciels d'assemblages ont été comparés et combinés (PCAP et ARACHNE).

Ces *supers-contigs* ont été ordonnés et orientés le long des chromosomes de la vigne par l'intermédiaire d'une carte génétique déjà existante (Doligez *et al* 2006) pour la première version de la séquence du génome. 409 marqueurs de cette carte ont permis d'ancrer 69% de la séquence du génome le long des chromosomes (61% en *super-contigs* ancrés et orientés). Cette étape a été largement améliorée dans la version 12X du génome avec la construction de deux cartes génétiques totalisant 527 marqueurs uniques avec un ordre solide. Ces cartes permettent d'ancrer plus de 93% du génome séquencé dans la version 12X du génome.

Une annotation automatique a été publiée pour la version 8X du génome en intégrant différentes couches d'information grâce au logiciel GAZE (Jaillon *et al* 2007): alignement d'ESTs *Vitis* ou plantes (environ 190 000 ESTs publics *Vitis* étaient disponibles), de 48 000 flicDNAs (plus de 100 000 ont été produits et partiellement séquencés dans le cadre du consortium à ce jour et les séquences complètes d'un set unigene sont en cours d'obtention), de protéines disponibles dans les bases de données publiques, recherche de régions du génome conservées entre le génome de la vigne et les génomes d'*Arabidopsis*, du peuplier et du riz, recherche de régions codantes *de novo*. L'ensemble de ces informations a permis de proposer

30 434 modèles de gènes sur cette première version du génome de la vigne. Pour la version 12X, à ces différentes couches d'information vont être rajoutés des résultats de séquençage profonds de cDNA avec la technique Solexa.

Cette annotation automatique sera ensuite corrigée ou validée par la communauté scientifique au cours des années (il y a toujours des nouvelles versions améliorées de l'annotation de la séquence du génome de la plante modèle *Arabidopsis thaliana* qui sont proposées, 8 ans après la publication de la première version; Arabidopsis Genome Initiative 2000). Il est donc important d'organiser cette annotation autour d'une base de données centralisée et de façon coordonnée. En ce qui concerne le génome de la vigne, c'est la plateforme de bioinformatique végétale nationale de l'INRA (URGI ; <http://urgi.versailles.inra.fr>) sous l'égide de l'IGGP qui sera en charge de faciliter et maintenir cette expertise manuelle et de diffuser régulièrement des versions de l'annotation dans les bases de données internationales.

L'annotation de la première version du génome de la vigne a permis de mettre en évidence le fait que certaines familles de gènes avaient subi une expansion chez la vigne. C'est le cas de la famille des gènes codant pour des stilbènes synthases et de celle codant pour les terpènes synthases. Dans les deux cas, les familles sont respectivement dix et deux fois plus importantes chez la vigne que chez *Arabidopsis*. Il va être très intéressant maintenant d'étudier l'expression de l'ensemble de ces gènes dans différents organes, à différents stades de développement, sous différents stress pour comprendre quel est leur rôle respectif. Cette étude sera ensuite complétée par l'étude des allèles existant dans les ressources génétiques pour chacun des gènes et leur association avec la variation d'un caractère agronomique.

Un autre résultat majeur de l'étude de la séquence du génome de la vigne est la découverte d'un ancêtre hexaploïde commun à la vigne, au peuplier et à *Arabidopsis* mais pas au riz. Le génome de la vigne ayant subi peu de remaniements depuis la formation de cet hexaploïde et n'ayant pas subi de duplications supplémentaires, il s'est avéré un très bon révélateur de cet événement ancien à la base de l'évolution du génome des dicotylédones. Une séquence des événements a été proposée par Jaillon et al (2007) et sera validée avec la version finale de la séquence et en utilisant d'éventuelles nouvelles séquences de génome disponibles.

En conclusion, la communauté scientifique devrait bientôt disposer d'une version finale de la séquence de référence du génome de la vigne de très bonne qualité en terme d'assemblage mais également d'annotation automatique grâce à la très grande quantité de séquences ESTs et flicDNA disponibles. Sur cette fondation solide, des projets de génomique

fonctionnelle et de génétique sont en co-construction avec la profession viticole pour répondre aux grands enjeux de cette culture : la réduction des intrants et l'adaptation au changement climatique.

Mais le goulet d'étranglement de la détermination de la fonction des gènes chez la vigne reste lié à son cycle biologique. Comme chez les autres plantes pérennes, les vignes transgéniques restent longues à obtenir (1 an environ) et si le gène d'intérêt est exprimé dans les baies, il faut encore attendre 2 ou 3 ans avant de pouvoir observer le résultat. Une autre solution consiste à créer de grandes collections de plusieurs milliers de mutants EMS pour des criblages TILLING (Dalmais et al 2008). Là encore, la création de telles collections chez des plantes pérennes hétérozygotes, si elle n'est pas impossible, reste beaucoup plus lourde que chez les plantes annuelles. Ces travaux indispensables sont néanmoins engagés en même temps que des recherches pour tenter d'en simplifier les étapes comme par exemple la création d'une vigne naine à cycle biologique plus court (6 mois entre le semi du pépin et l'obtention de baies) et manipulable dans des expériences de transformation (Boss et Thomas 2002). Enfin, le développement d'outils de génotypage à haut débit est également en cours dans le but de mettre en place des approches de génétique d'association chez la vigne.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Adam-Blondon A-F, Romieu C, Bouquet A (2004) Genomics as a road for grapevine improvement. *In* Fonctional Plant Genomics, Morot-Gaudry J-F, Lea P, Briat J-F (eds), INRA publisher, 463-480
- [2] Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408 : 796-815
- [3] Boss PK, Thomas MR (2002) Association of dwarfism and floral induction with a grape 'green revolution' mutation. *Nature*, 416 : 847-849
- [4] Bronner A and Oliveira J (1990) Creation and study of the Pinot Noir variety lineage. *In*: Proc. 5th Intern. Symp. Grape Breeding St Martin/Pflaz, Germany, Sept 1989, *Vitis*, special issue, pp69-80
- [5] Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savoie V, Aubert G, Brunaud V, de Oliveira Y, Guichard C, Thompson R, Bendahmane A (2008) UTILLdb, a *Pisum sativum* *in silico* forward and reverse genetics tool. *Genome Biol*, 9:R43
- [6] Jaillon O, Aury J-M, Noel B, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449 : 463-468
- [7] Lodhi MA, Reisch BI (1995) Nuclear DNA content of *Vitis* species, cultivars and other genera of the *Vitaceae*. *Theor Appl Genet*, 90 : 11-16

Diversité et homogénéité génétique de l'espèce canine : potentiel en génétique médicale

ANDRÉ Catherine, HITTE Christophe

Institut de Génétique et Développement, CNRS UMR6061, Université de Rennes1,
2 Av du Pr. Léon Bernard, 35043 Rennes, France

1. INTRODUCTION

L'espèce canine comporte plus de 350 races, qui diffèrent par leur phénotype, leur comportement et leur susceptibilité aux maladies. En revanche chaque race canine est homogène sur le plan phénotypique et génotypique et correspond à un véritable isolat génétique. Ainsi, bien souvent à partir de petits effectifs à la création des races, et pour se conformer à un strict "standard", les pratiques d'élevage, alliant sélection, consanguinité et "étalon champion" ont conduit à des groupes très homogènes, homozygotes à de nombreux loci et avec une distribution des allèles propre à chaque race. Plus de 350 races sont actuellement recensées, chacune possédant son propre historique mais toutes répondant à des critères morphologiques, comportementaux bien spécifiques, le généticien a à sa disposition autant de combinaisons alléliques uniques que de races. Cette sélection, basée au départ sur des critères morphologiques et comportementaux s'est bien souvent accompagnée de la co-sélection d'allèles indésirables. Les effets fondateurs et les pratiques d'élevage font qu'une entité clinique qui ségrège dans une race donnée sera très probablement la résultante d'une seule origine génétique. Il devient alors plus facile chez le chien que chez l'Homme d'établir la relation phénotype / génotype [1, 2].

Les maladies génétiques qui affectent le chien sont spontanées et non induites comme pour les modèles "rongeurs". La plupart des affections canines existent aussi chez l'Homme constituant ainsi de bons modèles de pathologie comparée. Elles ségrégent chez le chien avec de très fortes fréquences jusqu'à 20% et sont spécifiques de races ; on a donc accès en pratique vétérinaire à de nombreux cas, liés familialement au sein de chaque race, facilitant ainsi les études génétiques. De plus, le chien partage avec l'Homme un environnement commun, il est bien souvent exposé aux mêmes toxiques environnementaux, aux mêmes stress et il partage parfois la même alimentation. De plus, la physiologie et les réponses thérapeutiques de cette espèce sont très proches de celles qui prévalent chez l'Homme.

Face à la difficulté de recruter chez l'Homme des familles informatives ou des cas/contrôles *ad hoc* pour l'étude génétique de nombreuses maladies qu'elles soient multifactorielles ou même monogéniques, le chien offre une alternative intéressante. Les spécificités génétiques du chien font de cette espèce un bon modèle génétique, pour la recherche des bases moléculaires d'affections génétiques, particulièrement les maladies complexes humaines, telles que les cancers, par exemple. La capacité de collecter des ADN de chiens appartenant à des races fortement prédisposées a déjà permis l'identification de gènes impliqués chez le chien et chez l'Homme, pour un bénéfice mutuel. Par exemple, le Berger allemand est sujet à un cancer rénal (RCND "Renal Cystadenocarcinoma and Nodular Dermatofibrosis") dont le gène a été identifié grâce à une analyse de liaison génétique sur une grande famille; il s'agit du gène BHD dont les mutations chez l'Homme sont également responsables de cancers du rein, ce type de cancer nodulaire étant l'une des composantes du syndrome de Birt-Hogg-Dube [3]. Ainsi, la ségrégation d'une entité clinique unique dans une race canine, a permis de mettre en évidence la cause génétique de l'une des affections composant le syndrome humain de Birt-Hogg-Dube.

L'épilepsie est aussi un bon exemple du potentiel de ce modèle. Cette affection, dans sa forme primaire, représente le désordre neurologique le plus répandu dans l'espèce canine avec 5% de chiens atteints, toutes races confondues. De plus, ce qui est particulièrement intéressant, est que plus de cinquante races présentent des risques élevés de développer une forme d'épilepsie primaire. Dans ces races, la forte prévalence de formes d'épilepsies différentes, atteignant des fréquences supérieures à 20%, avec des modes de transmission monogénique, laisse supposer une cause génétique unique dans chaque race et différente d'une race à l'autre. L'hypothèse sous-jacente est que chaque race ou petit groupe de races développe une forme d'épilepsie liée à une cause génétique spécifique provenant d'une mutation fondatrice au moment de la création de la race. Ainsi, l'étude des bases moléculaires de l'épilepsie dans différentes races canines permettrait d'enrichir le nombre de gènes candidats pour des épilepsies humaines dont les bases génétiques sont encore inconnues. La découverte récente de l'implication du gène EPM2B dans la maladie de Lafora chez l'Homme (forme d'épilepsie myoclonique progressive) et dans l'épilepsie affectant le Teckel miniature à poil dur [4] démontre tout l'intérêt de ce modèle.

Le grand nombre de races, chacune avec des prédispositions spécifiques, constitue donc une richesse pour l'identification des bases moléculaires de maladies génétiques, ou encore de divers phénotypes, de comportement. Étudiées seules (comme par ex. le teckel et la maladie de Lafora), ou en différentes combinaisons, leur analyse génétique permet de décrypter les

bases moléculaires des traits étudiés, qu'ils soient pathologiques ou non. À partir de différents projets de recherche, en cours au laboratoire, nous illustrerons la puissance de ce modèle et surtout la capacité de travailler avec différentes combinaisons de races pour augmenter la puissance des analyses de liaison ou d'association, couramment utilisées en génétique humaine.

2. MÉTHODOLOGIE

Différentes situations illustrent l'intérêt de travailler avec plusieurs races, en des combinaisons différentes. La constitution de ces combinaisons demande de bien caractériser les maladies ou traits d'intérêt sur le plan clinique et épidémiologique dans plusieurs races, de bien connaître l'histoire de ces races et de traiter les résultats de génotypage sur le plan statistique, en tenant compte de la stratification entre et au sein des races étudiées :

Utilisation de 2 races ou plus, très éloignées phylogénétiquement mais partageant le même trait phénotypique, comme des traits sélectionnés volontairement par l'homme depuis très longtemps (ex: la taille des membres [5]). On recherche une région du génome commune à plusieurs races différentes sur le plan phénotypique, sauf pour le trait considéré. Par soustraction, les endroits du génome communs à toutes les races présentant ce phénotype (grands chiens), et non partagés par des races ne présentant pas ces phénotypes (petits chiens) correspondront aux locus et aux haplotypes recherchés.

Utilisation de 2 races très proches phylogénétiquement et historiquement et ne différant que très peu phénotypiquement, mais étant différemment affectées par deux maladies différentes (ex : l'épilepsie chez le Grand Bouvier Suisse et un cancer des histiocytes chez le Bouvier Bernois). Ces deux races originaires des Alpes suisses, ont évolué séparément récemment et ne se distinguent, du point de vue phénotypique que par une légère différence de taille et par des poils courts dans la première race et des poils longs pour la seconde. La comparaison de leur génome devrait montrer une grande similarité d'allèles, excepté au niveau des locus responsables de ces phénotypes " taille " et " longueur du poil " et devrait permettre de cibler les locus des deux maladies génétiques, l'une totalement absente chez les Grands Bouviers Suisses, l'autre totalement absente des Bouviers Bernois. Cet exemple étudié au laboratoire sera plus amplement exposé.

Utilisation de 2 races différentes ou de races " croisés " présentant le même phénotype (cliniquement la même maladie) permet de restreindre la zone de recherche à l'intérieur du locus identifié à partir d'une seule race. En effet, faire appel à une ou plusieurs autres races augmente les chances de

trouver des animaux recombinants et donc de réduire le locus déjà identifié. Les pré-requis sont que la maladie doit être la même sur le plan clinique, cela demande un diagnostic clinique extrêmement précis et qu'elle soit due à la même cause génétique (ceci ne se découvre en général qu'à postériori !)

Les analyses génétiques réalisées dans ces différentes races reposent sur les méthodologies classiques de génotypages, suivies d'analyses statistiques de liaison génétique, d'association, ou des trios. Ces méthodes demandent donc de disposer de grandes familles avec des fratries complètes comportant des chiens sains et atteints, ou bien de disposer de chiens non apparentés et de recruter des cas et des contrôles. L'utilisation de l'une ou l'autre des méthodes dépend des effectifs des races étudiées, de leur historique, de leur niveau de consanguinité et du mode de transmission des maladies étudiées.

3. RÉSULTATS

De telles recherches ont été entreprises pour deux affections d'origine génétique différentes, dans deux races proches historiquement et avec des stratégies complémentaires. Une analyse de liaison génétique a été réalisée sur une famille de 200 Grands Bouviers Suisse, comportant 13 chiens atteints d'épilepsie, avec 300 marqueurs microsatellites, répartis sur la totalité du génome. En parallèle, une seconde analyse a été effectuée sur une famille de 300 Bouviers Bernois, comportant 200 chiens atteints d'histiocytose maligne (un cancer très agressif des cellules histiocytaïres) avec les 300 mêmes marqueurs microsatellites. L'analyse statistique de ces deux expériences de génotypage, séparément, pointe sur des locus dans les deux races, mais ces régions sont encore de grande taille (20 Mb). Afin de confirmer ces régions génomiques et de réduire leur taille, l'analyse comparée des génotypages dans les deux races est en cours pour identifier les gènes responsables. Nous recherchons par soustraction les régions dont les haplotypes sont significativement différents entre les deux races. Nous nous attendons à retrouver les locus contenant des gènes impliqués dans la détermination de la taille des chiens et dans la taille du poil (court/long). Les autres locus identifiés seront potentiellement ceux recherchés. En fonction de la précision des résultats, l'utilisation de puces SNP (Single Nucleotide Polymorphism) permettant simultanément l'analyse allélique d'approximativement 60000 marqueurs canins sera effectuée sur les deux familles de chiens.

4. CONCLUSIONS

L'accès à une infinie collection d'allèles répartis dans 350 races canines, permet à l'évidence de décrypter les relations phénotype/génotype pour un grand nombre de traits et de maladies génétiques. L'utilisation appropriée des races canines en fonction de leur phénotypes, de leur historique et de leurs effectifs permet d'identifier les gènes et leurs allèles responsables. Cet exposé présentera certaines combinaisons de races, dont le génotypage complet du génome permet d'identifier de nouveaux gènes responsables d'affections génétiques, simples ou complexes ou de phénotypes, d'intérêt chez l'Homme.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Galibert F, Andre C, Hitte C: Dog as a mammalian genetic model. *Med Sci (Paris)* 2004, 20(8-9):761-766.
- [2] Ostrander EA, Comstock KE: The domestic dog genome. *Curr Biol* 2004, 14(3):R98-99.
- [3] Ostrander EA, Wayne RK: The canine genome. *Genome Res* 2005, 15(12):1706-1716.
- [4] Lohi H, Young EJ, Fitzmaurice SN, Rusbridge C, Chan EM, Vervoort M, Turnbull J, Zhao XC, Ianzano L, Paterson AD *et al*: Expanded repeat in canine epilepsy. *Science* 2005, 307(5706):81.
- [5] Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasa-hasram B, Karlins E, Davis S, Jones PG *et al*: A single IGF1 allele is a major determinant of small size in dogs. *Science* 2007, 316(5821):112-115.

Analyse fonctionnelle de micro-organismes et de communautés complexes par des approches (méta)protéomiques

Florence ARSÈNE-PLOETZE^{(1)*}, Florence GOULHEN-CHOLLET⁽¹⁾,
Bertrand CHAUMANDE⁽¹⁾, Stéphanie WEISS⁽¹⁾, Christine CARAPITO⁽²⁾,
Jessica CLEISS⁽¹⁾, Sébastien GALLIEN⁽²⁾, Christine SCHAEFFER⁽²⁾, Alain
Van DORSSELAER⁽²⁾, Philippe N. BERTIN⁽¹⁾

⁽¹⁾Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS et
Université Louis Pasteur, 28 rue Goethe, 67000 Strasbourg, France

⁽²⁾Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, ULP, CNRS,
25 rue Becquerel, 67087 Strasbourg, France

Résumé : L'effet d'éléments métalliques toxiques, comme l'arsenic, sur la physiologie et le métabolisme des micro-organismes a été étudié en utilisant des approches protéomiques chez des isolats naturels, ou métabolomiques à l'échelle de communautés microbiennes présentes sur des sites contaminés. Ces approches permettront d'avoir une image globale et intégrée des mécanismes d'adaptation des micro-organismes à des environnements extrêmes.

1. INTRODUCTION

Un des aspects les plus remarquables des micro-organismes est leur capacité à s'adapter aux conditions changeantes de leur environnement, que les facteurs impliqués soient de nature biotique (interaction avec d'autres organismes) ou abiotique (variation des facteurs physico-chimiques). L'étude physiologique de différents isolats naturels métabolisant l'arsenic (As), notamment en oxydant une forme toxique, l'As(III) en une forme moins toxique, l'As(V), constitue un exemple intéressant pour l'analyse de l'adaptation des micro-organismes à leur environnement. L'arsenic est un métalloïde retrouvé dans plus de 200 minéraux différents. Les formes inorganiques de l'arsenic sont très toxiques. Cette toxicité est due, au moins en partie, à sa fixation sur les groupements sulfhydriles de certaines protéines empêchant leur bon fonctionnement dans la cellule et pouvant induire un stress oxydatif. Une intoxication aiguë se caractérise par des douleurs œsophagiennes et abdominales et des diarrhées hémorragiques alors qu'une exposition prolongée peut être à l'origine de divers cancers.

* Correspondance et tirés à part : florence.ploetze@gem.u-strasbg.fr

L'arsenic est très soluble dans les eaux dont le pH est proche de la neutralité. Aussi, une contamination des eaux de distribution est fréquemment rapportée et ce composé toxique a été identifié comme un risque majeur pour la santé humaine en divers endroits du monde.

Deux bactéries isolées de deux sites contaminés par l'arsenic ont été étudiées dans le cadre des activités du GDR2909 (<http://gdr2909.u-strasbg.fr>): *Herminiimonas arsenicoxydans* et *Thiomonas* sp. Une étude protéomique a été réalisée pour chacune de ces deux bactéries, qui présentent des capacités d'adaptation différentes. En parallèle, nous avons étudié l'adaptation de bactéries non-cultivables en laboratoire, en réalisant une étude fonctionnelle de la communauté microbienne d'un site fortement contaminé par les métaux, le site minier de Carnoulès (France).

2. ANALYSE GÉNOMIQUE COMPARATIVE ET FONCTIONNELLE DE L'ADAPTATION DE DIVERS ISOLATS NATURELS

Herminiimonas arsenicoxydans est une β -protéobactérie chimiohétérotrophe, issue d'une station d'épuration (Muller *et al.*, 2006; Weeger *et al.*, 1999) et dont le génome a été séquencé en collaboration avec le Génoscope (Centre National de séquençage à Evry) (Muller *et al.*, 2007). C'est un bacille à Gram-négatif, légèrement incurvé doté d'un flagelle polaire unique. Cette bactérie est capable d'oxyder l'As(III) en As(V), grâce à une enzyme, l'arsénite oxydase (Lebrun *et al.*, 2003; Muller *et al.*, 2003). La réponse pléiotrope à l'arsenic, qui met en jeu des systèmes répondant globalement à un environnement, ont été évaluées en réalisant une étude protéomique différentielle. Cette analyse a permis d'identifier 22 protéines dont l'expression est régulée en réponse à l'As (Carapito *et al.*, 2006; Muller *et al.*, 2007). Ensuite, un inventaire de l'ensemble des protéines exprimées (carte protéomique) lorsque la bactérie est cultivée en présence d'arsenic a été établi (Weiss *et al.*, submitted). Les résultats obtenus suggèrent que cette bactérie possède un large panel de stratégies pour résister et métaboliser les différentes formes inorganiques de l'arsenic, avec notamment, en plus d'une arsénite oxydase, plusieurs arsénates réductases et pompes d'extrusion de l'arsénite. De plus cette bactérie exprime en présence d'arsenic, de nombreuses protéines impliquées dans la réponse aux stress, dans le transport du phosphate, dans le métabolisme énergétique et dans la motilité.

Les bactéries du genre *Thiomonas* sont aérobies strictes, mésophiles ou thermophiles modérées (50°C) et acidotolérantes (pH optimal de 6). Ce sont des chimiolithoautotrophes facultatifs, mixotrophes. La souche *Thiomonas* 3As a été isolée d'un ruisseau, le Reigous dont la source traverse les déchets

miniers du site de Carnoulès dans le Gard, fortement contaminé par l'arsenic (Duquesne *et al.*, 2008). Tout comme *Hermiinimonas arsenicoxydans*, certaines souches de *Thiomonas* isolées de différents sites contaminés par l'arsenic sont capables d'oxyder l'As(III) en As(V) (Battaglia-Brunet *et al.*, 2006; Bruneel *et al.*, 2003; Duquesne *et al.*, 2008). A nouveau, le séquençage du génome de la souche 3As a été réalisé en collaboration avec le Génoscope. Parallèlement, nous avons entrepris sur 3 souches de *Thiomonas*, présentant des capacités physiologiques différentes, une analyse comparative des réponses adaptatives par approche protéomique. En particulier, nous avons pu observer que chez *Thiomonas* 3As, comme chez *H. arsenicoxydans*, des protéines impliquées dans la résistance à l'arsenic (arséniate reductase) sont exprimées plus fortement en présence d'arsenic (fig. 1)

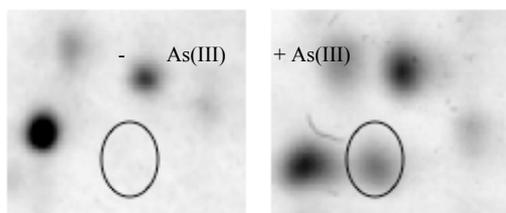


Figure 1 : Augmentation de l'expression de l'arséniate reductase (ArsC) en présence d'arsenic observée en électrophorèse bidimensionnelles sur des extraits protéiques de la souche *Thiomonas* 3As.

Cette étude permettra d'apporter des éléments pour comprendre le fonctionnement de l'écosystème dont ces bactéries sont issues, le site contaminé de Carnoulès. Cependant, l'étude de ce site ne peut se résumer à l'étude de ces seules souches, même si leur métabolisme semble y jouer un rôle important. De ce fait, nous avons envisagé, en parallèle à l'étude d'isolats naturels, l'étude fonctionnelle de la communauté microbienne de Carnoulès, par une approche métaprotéomique.

3. ANALYSE FONCTIONNELLE DE LA COMMUNAUTÉ MICROBIENNE D'UN SITE FORTEMENT CONTAMINÉ PAR LES MÉTAUX, LE SITE MINIER DE CARNOULÈS

L'application à l'étude des écosystèmes, des techniques récentes dérivant de la génomique descriptive (organisation et composition des génomes) et fonctionnelle (transcriptome, protéome, interactome) peut s'avérer très utile pour étudier une communauté dont la majorité des micro-organismes sont non cultivables (Bertin *et al.*, 2008). Récemment, la technique de métagénomique (techniques d'analyse globale des génomes d'organismes d'un environnement sans culture préalable) a été utilisée mais ne permet pas

toujours, à elle seule, d'étudier la fonctionnalité d'une communauté, au contraire peut-être de l'analyse du métaprotéome (analyse des protéines synthétisées par une communauté sans culture préalable).

La mine de Carnoulès dans le Gard a été exploitée jusqu'en 1962 pour le plomb, le zinc et l'argent (Morin *et al.*, 2003). Les déchets miniers sulfurés, composés d'arsénopyrite contenant 0,7 % de plomb, 10 % de fer et 0,2 % d'arsenic ont été stockés au fond d'un vallon. Un petit cours d'eau, le Reigous, prend naissance au niveau de ces déchets. Les conditions de ce ruisseau sont non seulement extrêmes par sa concentration en arsenic et en fer, mais aussi par son pH (environ 3). Le Reigous est un bon exemple d'atténuation naturelle, puisque la concentration en arsenic soluble chute de 56% dans les premiers 150 mètres en aval de la source (fig. 2). L'étude des populations bactériennes (cultivables et non cultivables) a montré que plusieurs micro-organismes eucaryotes (*Euglena*) ou procaryotes (*Thiomonas*, *Acidithiobacillus ferrooxidans*) pourraient jouer un rôle important dans ce processus.

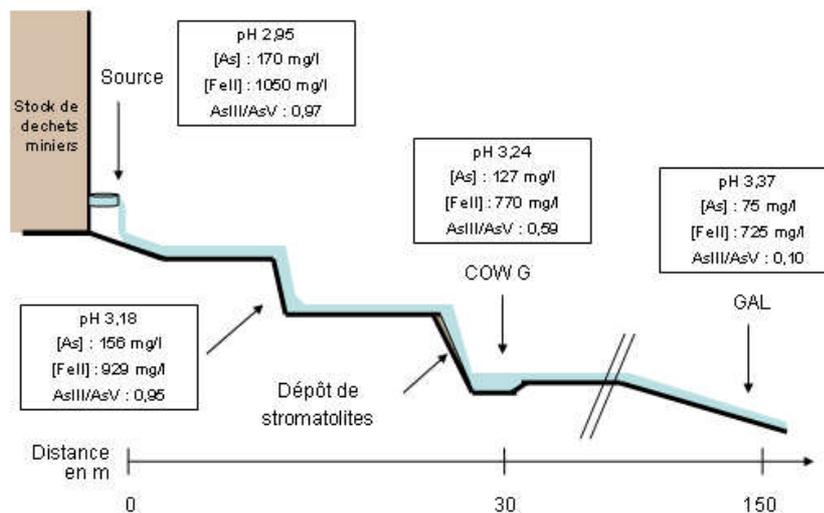


Figure 2 : Vue schématique latérale du site de Carnoulès. Dans les 30 premiers mètres, une précipitation de l'arsénite avec le Fe(III) est observée ; puis c'est la forme arséniate, As(V), qui précipite avec les ions ferriques. Les mesures ont été réalisées en mars 2001 (Casiot *et al.*, 2003).

Pour entreprendre une étude fonctionnelle de la communauté de ce site impliquée dans ce processus d'atténuation naturelle, nous avons développé l'analyse métaprotéomique. 63 protéines exprimées par des bactéries, des micro-organismes eucaryotes ou des végétaux ont été identifiées. Nous avons notamment observé que des micro-organismes majoritaires trouvés

sur ce site (*Euglena* et *Thiomonas*), mais également d'autres micro-organismes minoritaires, jouent un rôle important en exprimant des protéines impliqués dans plusieurs processus métaboliques (Arsène-Ploetze *et al.*, soumis).

Dans le futur, disposer de techniques d'analyses fonctionnelles globales comme la métaprotéomique permettra une étude différentielle en fonction du point de prélèvement ou encore de la saison. De plus, la possibilité de comparer les métaprotéomes d'environnements variés aux séquences de métagénome de ces mêmes environnements, permettra d'avoir une vision globale de la communauté, à la fois descriptive et fonctionnelle, et de suivre plus précisément la dynamique spatio-temporelle qui l'affecte.

4. CONCLUSION

En combinant une approche protéomique d'isolats naturels et des approches métaprotéomiques (et métagénomiques) de communautés microbiennes issues de sites contaminés par l'arsenic, notre objectif est d'identifier les protéines impliquées dans les processus d'adaptation de ces micro-organismes à de tels écosystèmes. Ces protéines présentent potentiellement un intérêt pour la compréhension de ces mécanismes d'adaptation. De plus, la caractérisation de leurs propriétés biochimiques et structurales devrait permettre leur utilisation à des fins d'observation des écosystèmes (bio-indicateurs). Enfin, les résultats obtenus fourniront une image intégrée de la diversité des mécanismes impliqués et devraient permettre ainsi de mieux appréhender la complexité de ces écosystèmes et donc d'en prévoir plus aisément les perturbations en réponse à des variations des facteurs environnementaux.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Arsène-Ploetze, F., Bruneel, O., Gallien, S., Weiss, S., Chaumande, B., Morin, G *et al.* (submitted). Characterization of an ecosystem, the acid mine drainage of Carnoulès, France, by an integrated approach.
- [2] Battaglia-Brunet, F., Joulain, C., Garrido, F., Dictor, M. C., Morin, D., Coupland, K., Barrie Johnson, D., Hallberg, K. B. & Baranger, P. (2006). Oxidation of arsenite by *Thiomonas* strains and characterization of *Thiomonas arsenivorans* sp. nov. *Antonie van Leeuwenhoek* 89: 99-108.
- [3] Bertin, P. N., Medigue, C. & Normand, P. (2008). Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiology* 154: 347-359.
- [4] Bruneel, O., Personné, J. C., Casiot, C., Leblanc, M., Elbaz-Poulichet, F., Mahler, B. J., Le Fleche, A. & Grimont, P. A. (2003). Mediation of arsenic oxi-

- dation by *Thiomonas* sp. in acid-mine drainage (Carnoules, France). *J Appl Microbiol* 95: 492-499.
- [5] Carapito, C., Muller, D., Turlin, E., Koechler, S., Danchin, A., Van Dorselaer, A., Leize-Wagner, E., Bertin, P. N. & Lett, M. C. (2006). Identification of genes and proteins involved in the pleiotropic response to arsenic stress in *Caenibacter arsenicoxydans*, a metalloresistant beta-proteobacterium with an unsequenced genome. *Biochimie* 88: 595-606.
- [6] Casiot, C., Morin, G., Juillot, F., Bruneel, O., Personné, J. C., Leblanc, M., Duquesne, K., Bonnefoy, V. & Elbaz-Poulichet, F. (2003). Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoules creek, France). *Water Res* 37: 2929-2936.
- [7] Duquesne, K., Lieutaud, A., Ratouchniak, J., Muller, D., Lett, M. C. & Bonnefoy, V. (2008). Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ Microbiol* 10: 228-237.
- [8] Lebrun, E., Brugna, M., Baymann, F., Muller, D., Lievreumont, D., Lett, M. C. & Nitschke, W. (2003). Arsenite oxidase, an ancient bioenergetic enzyme. *Mol Biol Evol* 20: 686-693.
- [9] Morin, G., Juillot, F., Casiot, C., Bruneel, O., Personne, J. C., Elbaz-Poulichet, F., Leblanc, M., Ildefonse, P. & Calas, G. (2003). Bacterial formation of tooeelite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoules acid mine drainage, France. A XANES, XRD, and SEM study. *Environ Sci Technol* 37: 1705-1712.
- [10] Muller, D., Lievreumont, D., Simeonova, D. D., Hubert, J. C. & Lett, M. C. (2003). Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium. *J Bacteriol* 185: 135-141.
- [11] Muller, D., Simeonova, D. D., Riegel, P., Mangenot, S., Koechler, S., Lievreumont, D., Bertin, P. N. & Lett, M. C. (2006). *Herminiimonas arsenicoxydans* sp. nov., a metalloresistant bacterium. *Int J Syst Evol Microbiol* 56: 1765-1769.
- [12] Muller, D., Medigue, C., Koechler, S., Barbe, V., Barakat, M., Talla, E. *et al.* (2007). A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genet* 3: e53.
- [13] Weeger, W., Lièvreumont, D., Perret, M., Lagarde, F., Hubert, J.-C., Leroy, M. & Lett, M.-C. (1999). Oxidation of arsenite to arsenate by a bacterium isolated from an aquatic environment. *BioMetals* 12: 141-149.
- [14] Weiss, S., Carapito, C., Cleiss, J., Koechler, S., Turlin, E., Coppee, J.-Y. *et al.* (submitted). Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomic data.

Multiplicité des facteurs conduisant à un choix de conservation

BOULINEAU François

GEVES – 49250 BRION

La conservation de gènes d'intérêt ou de combinaisons de ces mêmes gènes est bien sûr très importante pour les générations futures mais avec la mise en place de nombreux réseaux de conservation qui évaluent, multiplient et conservent les génotypes existants, l'érosion constatée ces dernières années semble fortement ralentie, voire stoppée du moins d'un point de vue génétique.

Il en est tout autrement au niveau de la perte de pratiques agricoles, de diversité d'approvisionnement pour le consommateur ou de traditions riches d'histoire et il est primordial d'assurer une complémentarité entre demande scientifique et demande sociétale (*exemple des Johnnies de Roscoff*). De nombreuses « Confréries » créées autour de variétés de légumes perpétuent des traditions locales pleines d'enseignement pour les prochaines générations. Que ces variétés disparaissent et c'est tout ce patrimoine qui se désagrègera de façon irréversible. Il est donc particulièrement important de prendre en compte cet impact culturel et patrimonial (au même titre que les distances génétiques), lors du choix des espèces à conserver prioritairement ou des variétés à introduire en Collection nationale. Les espèces oubliées ou délaissées sont particulièrement sensibles à l'érosion génétique car ne présentant pas d'intérêt génétique immédiat (pas de programme de sélection), ni d'attractivité culturelle suffisante (*Lentille de St Flour ou de Cilaos, Panais Demi-long de Guernesey, Cardon Vert de Vaulx en Velin*). Une concertation doit se mettre en place entre scientifiques « chercheurs de gènes » et acteurs concernés par la diversité d'utilisation des variétés. La conservation de ces variétés « patrimoniales » peut se faire « *in situ* », dans la région d'origine si la variété a encore suffisamment d'attractivité ou « *ex situ* » dans les réseaux constitués.

Dans tous les cas, conservation « *in situ* » ou « *ex situ* », il faut contrôler :

- que les variétés concernées appartiennent bien à l'espèce annoncée, (*Haricot de Soissons, nombreuses cucurbitacées ...*) ;

- que les variétés utilisées ou conservées correspondent au type d'origine (cas de la *carotte De Tilques* ou des *Maugettes de Vendée* résistantes à l'anthracnose) ;
- que ces mêmes variétés sont bien liées à des pratiques agricoles spécifiques ou appartiennent au patrimoine culturel et historique de certaines régions (variétés d'endives avec terre de couverture dans le nord de la France, séchage de haricots en perroquet dans de nombreuses régions productrices, association maïs/haricot à proximité de Tarbes ...).

Dans le cas de la conservation « *in situ* », il faut de plus s'assurer :

- que la maintenance des variétés est assurée de façon satisfaisante (*Chicorée de Chioggia*).
- Trop souvent les variétés sont maintenues et multipliées sans réels isolements, en fécondation libre ou sur la base d'effectifs trop faibles pour maintenir le potentiel génétique de départ ;
- qu'il n'y a pas tromperie pour l'utilisateur (*Tomate cœur de bœuf*) ;
- que les réglementations en vigueur permettent la commercialisation de ces variétés. La nouvelle directive européenne qui se met en place (Directive UE 2008/62) concernant les « variétés de conservation » (« Landraces ») doit permettre cette conservation « *in situ* » en autorisant la commercialisation de variétés « naturellement adaptées aux conditions locales et régionales et menacées d'érosion génétique » sans toutefois définir ces termes avec plus de précision.

Pour ne pas interagir avec la réglementation catalogue déjà en place, la directive prévoit de limiter les quantités de semences produites et d'informer l'utilisateur par la mention « Variété de conservation ». La maintenance de ces variétés devra se faire obligatoirement dans la région d'origine définie au moment de l'admission et pourra être contrôlée par sondages. La commercialisation des semences est elle aussi limitée à la région d'origine.

Il est préférable à mon sens de favoriser la conservation « *in situ* » des variétés car elle permet à moindre coût de conserver le potentiel génétique d'origine tout en perpétuant des traditions locales mais dans certains cas extrêmes, la conservation « *ex situ* » doit être envisagée pour faciliter un éventuel regain d'intérêt de types variétaux oubliés (les variétés de carottes jaunes ou blanches remises au goût du jour sont pratiquement toutes issues de réseaux de conservation).

Dans tous les cas des échanges doivent être recherchés et une synergie doit se mettre en place entre tous les acteurs de la conservation.

Qui était LUCA? Portrait du Last Universal Common Ancessor

Céline BROCHIER

Les travaux réalisés en biologie moléculaire au cours du XX^e siècle ont révélé l'unicité du monde vivant. En effet, tous les êtres vivants partagent un certain nombre de caractères, comme par exemple l'ADN, support de leur information génétique (associé à un code génétique universel); des mécanismes informationnels de transmission et d'expression de l'information génétique similaires (réplication, transcription, traduction); la possession d'une membrane plasmique constituée de lipides et de protéines membranaires assurant un rôle de barrière sélective et interaction avec le milieu extérieur; l'utilisation de petites molécules organiques (ATP, acides aminés, nucléotides...).

L'unicité du monde vivant a également été démontrée par l'analyse de l'information génétique portée par les organismes, issue tout d'abord du séquençage des protéines dans les années 50 puis à la fin des années 70 de celui de l'ADN. Si les travaux réalisés par Carl Woese sur l'ARN de la petite sous-unité du ribosome (PSU ARN_r) et les premières phylogénies universelles basées sur ce marqueur publiées à la fin des années 70 et au début des années 80 allaient dans ce sens, ils ont par ailleurs bouleversé la vision du monde vivant de l'époque. Les principales conséquences furent la mise en évidence de l'incroyable diversité génétique des microorganismes (bien supérieure à celle des macroorganismes) et la remise en cause de la classification dichotomique procaryotes/eucaryotes proposée par Roger Stanier et Cornelius Van Niels dans les années 60 [1] au profit d'une classification tripartite du vivant en trois domaines distincts: les Bacteria, les Archaea, et les Eucarya [2]. En effet, sur la base de l'analyse des séquences de l'ARN de la PSU ARN_r Carl Woese et George Fox ont montré que les procaryotes ne formaient pas un ensemble homogène, mais deux ensembles aussi distincts l'un de l'autre qu'ils ne le sont des eucaryotes (pour une rétrospective sur la découverte des Archaea, voir [3]). Ils ont baptisé le premier groupe de procaryotes Eubacteria (Eu = vrai) et le second Archaeobacteria (Archaea = ancien). Lorsque dans les années 90 les analyses phylogénétiques suggérèrent que les Archeobacteria pourraient être plus apparentées aux Eucarya qu'aux Eubacteria, Woese proposa de renommer les trois domaines du vivant Eucarya, Bacteria et Archaea, évitant l'emploi du suffixe "bacteria" qui pouvait suggérer que les Archaea sont proches des Bacteria [4].

L'unicité du monde vivant implique la présence d'un ancêtre commun aux trois domaines du vivant qui a été popularisé par l'acronyme LUCA (Last Universal Common Ancestor). Depuis que le concept de LUCA a été formalisé, de nombreuses recherches visant à caractériser cet ancêtre ont été entreprises [5]. Ces travaux sont très importants car LUCA est un organisme clé dans l'évolution de la vie sur terre. Pouvoir le caractériser avec précision permettrait de poser un jalon; jalon à partir duquel il serait possible de remonter encore plus loin dans l'étude de l'évolution ancienne et de se rapprocher des origines de la vie.

Une des conséquences du dynamisme de la recherche autour de LUCA a été l'introduction dans la littérature scientifique d'un certain nombre de concepts qu'il est important de rappeler avant de pouvoir esquisser le portrait de LUCA. Tout d'abord, LUCA, qui est, par définition, l'ancêtre de tous les êtres vivants actuels le plus récent dans le temps (i.e. le plus proche des êtres vivants actuels), est l'organisme à partir duquel les lignées évolutives conduisant aux organismes actuels se sont différenciées (Cellule rouge, Figure 1). Il est important de noter que LUCA possédait lui-même des ancêtres, qui sont également des ancêtres communs de tous les êtres vivants, mais qui ne peuvent pas prétendre au titre de LUCA car ils sont plus anciens. Donc, contrairement à une opinion relativement répandue, LUCA ne représente pas la première cellule. Au contraire, la plupart des évolutionnistes s'accorde sur le fait que LUCA était un organisme cellulaire relativement complexe (voir ci-après et [5, 6]). Ceci implique que les caractéristiques de LUCA (comme par exemple le fait qui aurait pu être hyperthermophile) ne renseignent donc pas directement sur l'origine de la vie (et inversement), ni sur les environnements où elle est apparue et s'est développée. (Figure 1). Bien qu'il soit impossible de dater l'époque à laquelle vivait LUCA, il est possible de le situer par rapport à d'autres événements pour lesquels des dates sont connues. Par exemple l'oxygénisation des océans et de l'atmosphère commencée il y a ~2,4 milliards d'années est attribuée à l'apparition de la photosynthèse oxygénique dans la lignée des Cyanobacteria. Ces organismes, appartenant au domaine Bacteria, sont donc des descendants de LUCA, ceci permet de postuler que LUCA vivait à une époque antérieure, mais postérieure à l'apparition de la vie généralement située entre -3,4 et -3,0 milliards d'années. Ceci implique que LUCA aurait pu vivre à une époque située entre -3,4 et -2,4 milliards d'années. Allant de pair avec ces idées, LUCA n'était probablement pas un organisme isolé. Il était très certainement entouré de nombreux contemporains, qui n'ont pas, à la différence de LUCA, laissé de descendants vivant actuellement. Il est probable que la majorité des lignées évolutives ayant existé sur terre ont disparu sans laisser de descendants (Lignes pointillées, figure 1), ni de traces au niveau du registre fossile. Ainsi, s'il est impossible de dire à quoi ces lignées ressemblaient, il est cependant possible qu'elles aient joué un rôle

dans l'évolution des écosystèmes. Enfin, il est important de souligner que d'un point de vue évolutif les organismes vivants actuels sont tous aussi éloignés de LUCA. Ceci implique qu'il n'y a pas d'organismes plus apparentés à LUCA que d'autres et donc que les notions d'organismes supérieurs ou inférieurs, primitifs ou évolués ne devraient pas être employées pour qualifier des organismes modernes. Si la quantité d'évolution qui sépare tous les organismes actuels de LUCA est identique, elle n'est pas comparable d'un point de vue qualitatif d'une lignée à l'autre (i.e. l'évolution a pu prendre différentes formes chez les Archaea, les Eucarya et Bacteria par exemple).

Connaître LUCA implique que la plupart des relations de parenté qui unissent les êtres vivants (en particulier les plus anciennes) ont été élucidées. Ceci permettrait de réaliser le rêve de Darwin, c'est-à-dire l'élaboration d'une classification phylogénétique universelle, c'est-à-dire qui reflète les relations de parenté entre tous les organismes. Par ailleurs, à partir de cette classification il serait possible de comprendre comment l'évolution, à partir de LUCA, a abouti à l'incroyable diversité des êtres vivants actuels. Elle permettrait également de distinguer les mécanismes/processus/structures biologiques ayant une origine ancienne des innovations plus récentes.

L'existence de trois domaines du vivant pose la question de leur ordre d'apparition. Dans une vision classique de l'évolution ancienne les trois domaines du vivant (Archaea, Bacteria et Eucarya) sont issus de deux événements de spéciation (*i.e.* événement évolutif au cours duquel une lignée évolutive va donner naissance à deux lignées). Selon ce schéma, LUCA est l'organisme qui a donné naissance à la lignée évolutive correspondant à l'un des trois domaines (Lignée évolutive bleue, Figure 1) et à une lignée évolutive (Lignée évolutive orange, Figure 1) qui va se diviser plus tard en deux lignées qui seront à l'origine des deux autres domaines (Lignées évolutives rose et verte, Figure 1). Ceci implique également que deux des trois domaines partagent un ancêtre exclusif qui n'est pas un ancêtre du troisième domaine (Cellule orange Figure 1). Les événements de spéciation mentionnés précédemment (Cellules rouge et orange, Figure 1) correspondent aux points de divergence des domaines. Chacune des lignées résultantes (rose, verte et bleue) va se diversifier indépendamment des autres. Au sein de chacune d'entre-elles, un organisme occupe une place privilégiée, c'est l'organisme à partir duquel les lignées évolutives menant aux bactéries, archées et eucaryotes actuels vont se différencier (Cellules rose, bleue, et verte, Figure 1). Ces organismes représentent les ancêtres communs les plus récents de chacun des trois domaines: Bacteria (Last Bacterial Common Ancessor), des Eucarya (Last Eucarya Common Ancessor) et LACA (Last Archaeal Common Ancessor). Il est important de

noter que ces organismes ne vivaient très certainement pas à la même époque ni dans le même lieu.

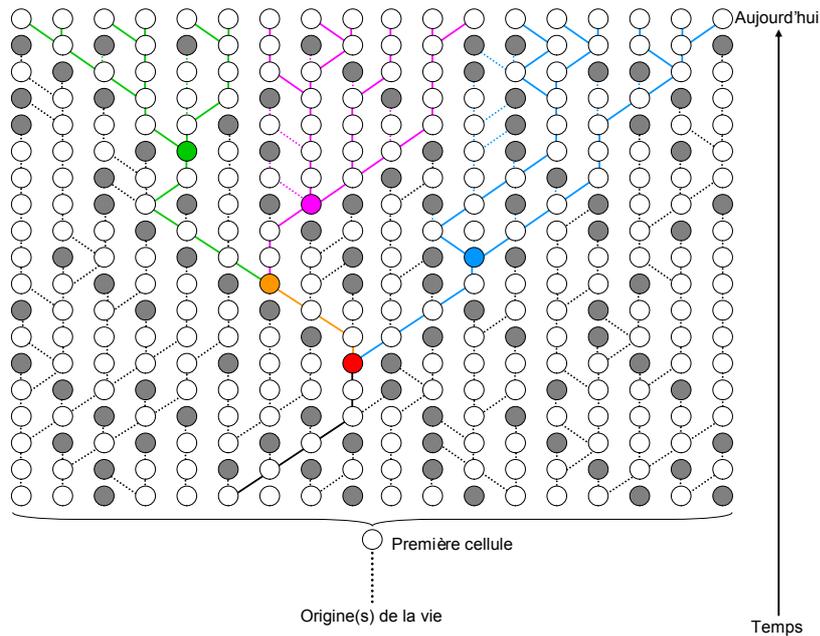


Figure 1 : Schéma représentant l'évolution de la vie sur terre depuis l'apparition de la première cellule. Par souci de clarté les étapes ayant conduit de l'apparition de la vie à la première cellule et de la première cellule aux premières communautés de microorganismes n'ont pas été représentées (pour plus de détails voir [5], [6]). En accord avec le mode de reproduction cellulaire (une cellule donne naissance à deux cellules), il est en théorie possible de retracer la généalogie de toutes les cellules peuplant la terre actuellement. A cause du phénomène de coalescence, les lignées évolutives ayant donné naissance aux cellules actuelles, convergent jusqu'à une cellule unique. Cette cellule représente LUCA, l'ancêtre commun à tous les êtres vivant actuellement. Les couleurs rose, bleue et verte représentent les lignées évolutives correspondant aux trois domaines du vivant. Les cellules bleue, verte et rose représentent les ancêtres communs les plus récents de chacun des trois domaines: Last Bacterial Common Ancestor (LBCA); Last Eukaryotic Common Ancestor (LECA) et Last Archaeal Common Ancestor (LACA); c'est-à-dire les organismes à partir desquels les lignées actuelles de bactéries, d'archées et d'eucaryotes ont commencé à se diversifier. La vision classique de l'évolution qui suppose que les trois domaines sont issus de deux événements de spéciation implique, d'une part l'existence d'une lignée représentée ici en orange qui a pour origine LUCA et qui va donner naissance à deux des trois domaines, et d'autre part l'existence d'un ancêtre commun exclusif à deux des trois domaines.

Dans les années 90, la question de l'ordre d'apparition des domaines a passionné la communauté des évolutionnistes. En effet, la réponse est la clé indispensable pour comprendre l'origine des caractères partagés par les domaines deux à deux. Par exemple, Archaea et Eucarya possèdent des machineries informationnelles composées de protéines similaires, mais très différentes de celles des Bacteria; il serait intéressant de pouvoir dire laquelle des deux machineries était présente chez LUCA et laquelle est apparue secondairement. De même les Bacteria et Eucarya possèdent des lipides similaires, différents de ceux des Archaea; ce qui pose la question de la nature des lipides présents chez LUCA. Enfin, Bacteria et Archaea se caractérisent par une organisation cellulaire (procaryote); différente de celle des eucaryotes. Ces quelques exemples illustrent l'importance à résoudre les relations de parenté entre les trois domaines.

L'approche proposée pour y parvenir est basée sur la phylogénie moléculaire et consiste à analyser des protéines paralogues anciennes (i.e. issues d'un événement de duplication antérieur à LUCA) et qui ont été conservées dans les trois domaines du vivant. De telles protéines sont rares (moins d'une dizaine) et leur analyse très délicate, car il s'agit d'étudier un événement évolutif très ancien. L'analyse de ces protéines réalisée dans les années 90 suggère un événement de spéciation initial qui a donné naissance à la lignée évolutive à l'origine des Bacteria d'une part et à une lignée qui donnera naissance, suite à un deuxième événement de spéciation aux Archaea et Eucarya. Suivant ce résultat, dans la Figure 1, la cellule orange correspond à l'ancêtre commun des Archaea et des Eucarya, la lignée rose et la lignée verte correspondent aux Archaea et aux Eucarya, respectivement, alors que la bleue correspond aux Bacteria. Ceci signifie qu'en dépit de similitudes au niveau de leur organisation cellulaire, les Archaea sont plus apparentées aux Eucarya qu'aux Bacteria. Bien que cette polarisation de l'ordre d'apparition des trois domaines soit toujours très controversée [7, 8], elle a été largement admise dans la communauté scientifique. Elle implique par exemple, que les caractères partagés par les Archaea et les Bacteria (organisation cellulaire de type procaryote) d'une part, et Bacteria et Eucarya d'autre part (lipides...) étaient probablement déjà présents chez LUCA; alors qu'une incertitude demeure quant aux caractères communs aux Archaea et Eucarya versus les caractères spécifiques des Bacteria. Les uns comme les autres pourraient avoir été présents chez LUCA puis remplacés dans la lignée évolutive menant aux Bacteria ou dans celle commune aux Archaea et Eucarya.

Au-delà de la vision classique de l'arbre universel du vivant, certains chercheurs postulent que les Eucarya résultent d'une association entre Archaea et Bacteria. Ces hypothèses, regroupées sous l'appellation hypothèses chimériques, impliquent que LUCA est en fait l'ancêtre direct

des Archaea et des Bacteria. Bien que le nombre et la nature des partenaires diffèrent [9], ces hypothèses cherchent à expliquer la forte ressemblance d'une partie des gènes eucaryotes avec des gènes d'archées et de l'autre avec les bactéries.

Si la réflexion autour de LUCA a vu le jour au cours du XX^e siècle, l'obtention des premières séquences de génomes complets et la naissance de la génomique comparative à la fin des années 90 ont stimulé la recherche autour de LUCA en fournissant les données nécessaires à la mise en œuvre d'études à grande échelle. En effet, l'obtention de séquences de génomes d'organismes appartenant aux trois domaines du vivant a permis de comparer l'information génétique qu'ils portent. Grâce à ces données, il est désormais possible de distinguer les gènes d'origine ancienne, qui étaient probablement présents chez LUCA de ceux qui sont apparus plus tardivement au cours de l'évolution. L'étude qualitative de ces gènes permet d'inférer les structures, processus, métabolismes... ancestraux et d'identifier ceux qui sont plus récents car apparus dans l'un ou l'autre des trois domaines du vivant. En apparence simple, ce type d'approche est compliqué car il est possible que des gènes présents chez LUCA aient été perdus au cours de l'évolution dans un ou plusieurs des trois domaines du vivant (voire même dans les trois domaines); aient divergé de manière trop importante pour être identifiés comme descendant d'un même gène ancestral ayant été présents chez LUCA; et/ou avoir été remplacés par des gènes d'origine évolutive plus récente. Inversement certains gènes peuvent être inférés présents chez LUCA alors qu'ils sont apparus plus tardivement mais ont été transférés entre les domaines; enfin des gènes d'origine évolutive différente, non présents chez LUCA, peuvent avoir été recrutés indépendamment pour remplir la même fonction. Malgré ces difficultés, les différentes études s'accordent sur le fait que LUCA possédait plusieurs centaines de gènes dans son génome (de 320 gènes [10] à plus d'un millier de gènes [11]). En se basant sur les estimations minimales, il est inféré que LUCA était une cellule relativement complexe possédant des machineries de traduction et de transcription complètes, mais également capable de synthétiser *de novo* ses acides aminés, ses nucléotides, ses lipides, des carbohydrates complexes, quelques coenzymes et possédait quelques capacités métaboliques.

Les grandes catégories fonctionnelles présentes ou absentes du génome supposé de LUCA ne sont pas de peu d'importance. Par exemple, l'inférence de gènes codant pour des protéines membranaires dans le génome de LUCA (systèmes d'exportation des protéines, ATPases, cytochromes oxydases...) est un argument fort en faveur de l'hypothèse que LUCA était un organisme cellulaire délimité par une membrane constituée de lipides et non un organisme acellulaire comme cela a été récemment

proposé [12]. La présence de lipides dans les membranes de LUCA a également été confirmée par l'analyse phylogénétique de leur voie de biosynthèse [13]. Inversement, les analyses de génomique comparative ont révélé que plusieurs protéines impliquées dans la réplication de l'ADN ne sont pas homologues entre Bacteria d'une part et Eucarya/Archaea d'autre part. Plusieurs interprétations de ce résultat ont été proposées, chacune ayant des implications fortes sur la vision de LUCA [6]. Tout d'abord, ceci peut refléter le remplacement de la machinerie de réplication initialement présente chez LUCA par un système non homologue dans la branche conduisant aux Bacteria ou dans celle commune aux Archaea et Eucarya. Une autre interprétation est que LUCA ne possédait pas de machinerie de réplication de l'ADN et donc qu'il ne possédait pas d'ADN (cela aurait donc été une cellule avec un génome à ARN). Ceci implique que le passage d'un génome ARN vers un génome ADN se serait produit deux fois indépendamment, une fois chez un ancêtre des Bacteria et une fois chez un ancêtre commun aux Archaea et Eucarya [14].

L'étude de LUCA est un exercice comparable à reconstruire un puzzle dont la plus grande partie des pièces serait manquante. Par conséquent, de nombreuses questions relatives à LUCA restent actuellement sans réponse. Néanmoins, un certain nombre de consensus émergent et permettent de préciser progressivement l'esquisse de LUCA commencée il y a quelques dizaines d'années. Il est vraisemblable que l'accroissement des connaissances de la diversité de l'information génétique portée par les êtres vivants actuels et de la biodiversité va apporter de nouvelles pièces qui permettront de poursuivre l'aventure fascinante qu'est la reconstruction de l'évolution de la vie sur notre planète.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Stanier RY, Van Niels CB: The Concept of a Bacterium. *Arch Mikrobiol* 1962, 42:17-35.
- [2] Woese CR, Fox GE: Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 1977, 74:5088-5090.
- [3] Woese CR: The birth of the Archaea: a personal retrospective. In *Archaea*. Edited by Garrett RA, Klenk HP. Oxford: Blakwell publishing; 2007: 1-15.
- [4] Woese CR, Kandler O, Wheelis ML: Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990, 87:4576-4579.
- [5] Brochier-Armanet C: Origine et évolution des Micro-organismes. *Bull Soc Fr Microbiol* 2007, 22:217-226.
- [6] Forterre P, Gribaldo S, Brochier C: Luca: the last universal common ancestor. *Med Sci (Paris)* 2005, 21:860-865.

- [7] Gribaldo S, Philippe H: Ancient phylogenetic relationships. *Theor Popul Biol* 2002, 61:391-408.
- [8] Baptiste E, Brochier C: On the conceptual difficulties in rooting the tree of life. *Trends Microbiol* 2004, 12:9-13.
- [9] Martin W: Archaeobacteria (Archaea) and the origin of the eukaryotic nucleus. *Curr Opin Microbiol* 2005, 8:630-637.
- [10] Tatusov RL, Koonin EV, Lipman DJ: A genomic perspective on protein families. *Science* 1997, 278:631-637.
- [11] Ouzounis CA, Kunin V, Darzentas N, Goldovsky L: A minimal estimate for the gene content of the last universal common ancestor--exobiology from a terrestrial perspective. *Res Microbiol* 2006, 157:57-68.
- [12] Martin W, Russell MJ: On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* 2003, 358:59-83; discussion 83-85.
- [13] Pereto J, Lopez-Garcia P, Moreira D: Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* 2004, 29:469-477.
- [14] Mushegian A: The minimal genome concept. *Curr Opin Genet Dev* 1999, 9:709-714.

Le mildiou de la vigne : diversité génétique, introductions et durabilité des résistances

DELMOTTE François⁽¹⁾, MESTRE Pere⁽²⁾

⁽¹⁾INRA, Institut des Sciences de la Vigne et du Vin de Bordeaux, UMR 1065 Santé
Végétale, BP81 - 33883 Villenave d'Ornon cedex France

⁽²⁾INRA, UMR 1131 Santé de la vigne et qualité du vin, 6802 Colmar, France

Le mildiou de la vigne est une maladie causée par *Plasmopara viticola* (Oomycète) qui infecte la plupart des espèces du genre *Vitis* (vigne). *P. viticola* a été introduit en Europe en 1878 probablement avec les porte-greffes américains utilisés pour la reconstruction du vignoble français après la crise phylloxérique. Les variétés de vignes cultivées européennes (*Vitis vinifera*) étant toutes sensibles au mildiou, la lutte contre ce pathogène est essentiellement chimique. La façon la plus efficace de réduire l'utilisation des fongicides sur vigne est de créer de nouvelles variétés résistantes au mildiou en utilisant les sources de résistance présentes dans les espèces sauvages américaines (*Vitis* spp.). Le travail entrepris depuis plusieurs années à l'INRA de Colmar aboutit aujourd'hui à l'obtention de variétés combinant un ou plusieurs gènes de résistance au mildiou. Une parcelle de vigne étant plantée pour plusieurs décennies, la durabilité de ces gènes de résistance nécessite maintenant d'être évaluée. Dans le but d'anticiper l'adaptation du mildiou aux nouveaux gènes de résistance de la vigne, nous étudions la diversité génétique des populations de *P. viticola* à l'aide de marqueurs nucléaires et mitochondriaux. Les populations européennes présentent une faible diversité allélique, une faible structuration géographique et une forte capacité à la recombinaison génétique. Parallèlement, nos données indiquent qu'une petite population de souches a été à l'origine de l'introduction du mildiou en Europe. Ce dernier résultat suggère que la diversité génétique du mildiou doit être appréhendée sur des populations issues de leur centre d'origine et qui ont co-évolué avec les résistances trouvées dans les plantes hôtes. C'est pourquoi nous avons entrepris d'étudier le compartiment sauvage de *P. viticola* en retournant en Amérique du Nord. Là-bas, nous avons découvert une importante variabilité génétique qui laisse penser que nous sommes peut-être en présence de plusieurs espèces cryptiques, plus ou moins spécialisées sur leur plante hôte. La découverte de cette variabilité revêt une importance capitale au regard des recherches menées sur les gènes de résistance de la vigne vis-à-vis du mildiou. En effet, le risque de nouvelles introductions ne pouvant être écarté, il conviendra de prendre en compte

l'ensemble de la diversité génétique du mildiou dans les futurs schémas d'amélioration variétale de la vigne.

La gestion des informations relatives aux races à petits effectifs par les systèmes nationaux d'information génétique (SNIG) : contribution et problèmes rencontrés

DOS C.¹, DANCHIN-BURGE C.^{1,2}

¹Institut de l'Élevage, 149 rue de Bercy 75595 PARIS cedex 12

²INRA / AgroParisTech, UMR1236 Génétique et diversité animales, 16 rue Claude Bernard, F-75231 Paris cedex 05, France

Dans l'élevage des ruminants (bovins, ovins, caprins), on appelle système national d'information génétique (SNIG) - comprenant base de données, logiciel de traitement et systèmes d'échanges des données - le système d'information construit pour répondre aux besoins de tous les partenaires de la génétique animale des ruminants, et notamment l'évaluation génétique des reproducteurs. Parallèlement des bases de données ont été créées pour les besoins spécifiques des races à Petits Effectifs (PE). Une présentation des SNIG et des besoins des races en conservation permettront de conclure sur les apports possibles des SNIG et les problèmes rencontrés.

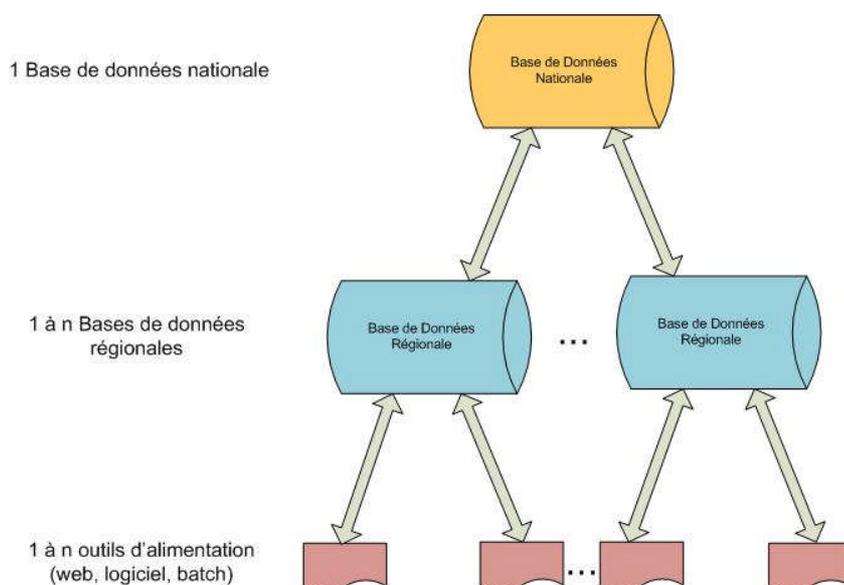
1. PRÉSENTATION DES SNIG

Le développement des contrôles de performance et des programmes collectifs de sélection se sont accompagnés de la création de systèmes d'informations ambitieux permettant de collecter et de valoriser l'ensemble des données, nombreuses et diverses, générées par l'élevage : date de naissance de l'animal, sexe, parentés certifiées ou non certifiées, événements de reproductions (inséminations, montes, mise bas...), performances (lait, viande ...), mouvements, pointages, et, plus récemment, informations moléculaires.

- Un SNIG a été mis en place par filière :
- SIG pour les bovins (lait et allaitant) ;
- OVALL pour les ovins allaitants ;
- SIEOL pour les ovins lait ;
- SIECL pour les caprins (système encore en cours de déploiement).

1.1. Architecture

Dans tous les cas, l'architecture générale est la suivante :



Les organismes de terrain (EDE, Contrôles Laitiers, Contrôles de Performance Viande, Entreprises de Mise en Place de la semence...) transmettent les données recueillies ou calculées à la base de données régionale qui les transmet à son tour à la base de données nationale.

1.2. Données gérées dans les SNIG

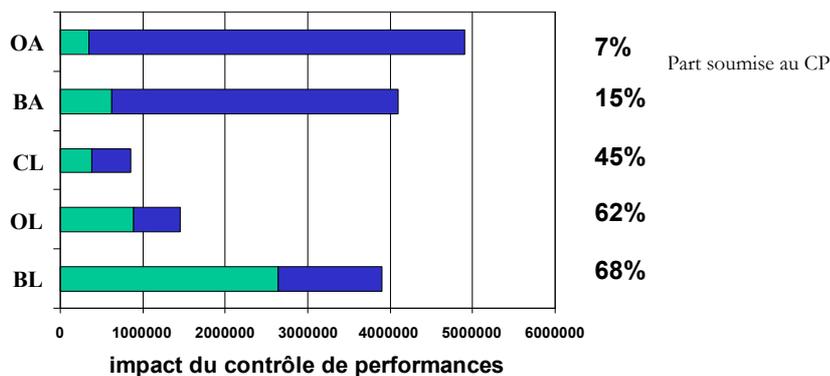
Les données historiques du SNIG sont les suivantes :

- Données d'identification des animaux (numéro national d'identification, sexe, race ...);
- Données de généalogie (filiations);
- Données des contrôles de performances de production (par exemple pour le lait la quantité, ou la composition et en viande la croissance de l'animal, ou son poids à âge type) et données morphologiques (pointages ...);
- Données de reproduction (date de saillie ou d'insémination, de mise bas...);
- Plus récemment, des données de laboratoire concernant des typages à partir des outils de la génétique moléculaire.

C'est à partir de toutes ces données que peuvent être, entre autres, générées les évaluations génétiques ou index, à savoir une estimation – à partir des méthodes de la génétique quantitative – de la valeur génétique d'un animal pour un caractère donné (quantité de lait, fertilité, etc.).

1.3. Périmètre couvert par les SNIG

Le périmètre historique des SNIG sont les animaux en contrôle de performances. La part de la population totale qui est soumise au contrôle de performance est très variable en fonction de la filière considérée comme le montre la figure suivante :



Données : Institut de l'Elevage – JM ASTRUC

OA = ovin allaitant (viande). BA= bovin allaitant. CL= Caprins lait. OL= Ovins lait. BL= Bovins laitiers.

En vert : animaux au contrôle de performance. En bleu : animaux hors contrôle.

Pour les bovins (lait et viande), à la fin 2007, plus de 87% de la population totale (animaux en contrôle de performance + animaux uniquement à l'identification) était gérée dans le SIG. Et ce pourcentage devrait augmenter fortement en 2008, sous l'effet de la nouvelle Loi d'Orientation Agricole, faisant obligation aux EDE d'apporter aux SIG l'exhaustivité des naissances et des actes de reproduction, même effectués sur des animaux n'étant pas au contrôle de performances.

2. GESTION DES INFORMATIONS DES RACES À PETITS EFFECTIFS

L'Institut de L'Elevage effectue un suivi d'une dizaine de races bovines à très petits effectifs (moins de 1000 vaches vivantes) depuis plus de vingt ans pour certaines races et de cinq races locales caprines. Ce travail a permis la

constitution d'un fichier généalogique pour chacune de ces races, grâce au travail décrit ci-après. Pour les ovins, les associations d'éleveurs sont responsables de la constitution et de la gestion des fichiers généalogiques, généralement basés sur le contrôle de performances. L'objectif de ces bases est l'exhaustivité de l'information. Il est indispensable, vu les effectifs, que tous les animaux soient inventoriés, quel que soit le type d'élevage : éleveurs hors contrôle de performance, hors association, éleveurs amateurs, etc.

2.1. Critères d'inventaire des animaux

L'objectif est de constituer des fichiers d'animaux et de généalogies, en général à partir de données "à dire d'éleveurs" car ce sont souvent les seules informations dont on dispose, la très grande majorité voire la totalité des éleveurs de ces races n'adhérant pas à un système de contrôle de performances. La collecte de ces informations ne peut se faire qu'avec un travail de terrain, et c'est là le principal travail de l'intervenant technique. En bovin par exemple, l'actualisation de ces inventaires nécessite de parcourir quelque 25 000 km par an pour rendre visite aux éleveurs.

Les informations collectées *a minima* par animal sont : un identifiant animal, le sexe, la date de naissance (*a minima* une estimation de son année de naissance) et le cheptel d'appartenance. En fonction des races et/ou des élevages, d'autres variables peuvent être collectées, en particulier la généalogie des animaux. Une actualisation des inventaires est faite grâce à un passage annuel dans les élevages, à la même période pour chaque race. Les données sont ainsi comparables d'une année à l'autre.

L'intérêt d'un tel répertoire est de pouvoir suivre l'évolution démographique de la race et, au fur et à mesure de l'accumulation des données généalogiques, d'effectuer un suivi de sa variabilité génétique. Il est en effet fondamental dans ces races de bien connaître les généalogies afin d'éviter les accouplements consanguins directs.

L'Institut de l'Élevage travaille en relation avec l'ensemble des éleveurs concernés, que ce soit des éleveurs isolés, des écomusées ou des fermes pédagogiques. Ce passage permet donc de collecter l'information, mais aussi de connaître les besoins des éleveurs et de les informer.

Cet inventaire représente un soutien psychologique non négligeable pour les éleveurs. Lorsque les programmes ont été lancés, la majorité des éleveurs étaient des personnes âgées et réfractaires aux méthodes "modernes" de l'agriculture. Ils souffraient de la non reconnaissance de leur race, mais d'un autre côté ils n'étaient pas prêts à affronter une intervention lourde de la part des pouvoirs publics. Ce sont souvent des gens jaloux de leur indépendance. L'inventaire a permis de montrer la reconnaissance de leur race : c'est un moyen de montrer aux éleveurs qu'ils sont pris en

considération, et que la conservation ne peut se concevoir que s'ils poursuivent leur travail.

2.2. Système d'information de la Cryobanque nationale

Outre la gestion des animaux vivants « *in situ* », il peut s'avérer nécessaire d'avoir recours, de façon complémentaire, à la conservation par le froid de semences et d'embryons des animaux les plus représentatifs des divers états de nos ressources génétiques patrimoniales. La mise en place d'une la Cryobanque nationale a pour objectif de fédérer les actions en matière de cryoconservation, de rationaliser les collections existantes et surtout de les sécuriser. Trois types de matériel issus de Populations d'Animaux Sélectionnés (PAS) sont stockés dans les collections :

- le matériel de type I, issu de PAS à effectifs réduits, pour lesquelles le risque de disparition est réel,
- le matériel de type II, issu d'animaux exceptionnels ayant certaines aptitudes qui apparaissent en exploitation mais non retenues aujourd'hui par le schéma de sélection qui a d'autres orientations,
- le matériel de type III, constitué d'échantillons représentatifs de l'état génétique d'une PAS soumise à une sélection intense à un instant donné.

Conjointement à la constitution de collections patrimoniales, un des objectifs essentiel de la Cryobanque nationale est de renseigner le matériel biologique cryoconservé et ce de façon la plus fiable. Cette information est nécessaire pour une réutilisation adéquate de ce matériel. Il a donc été défini un système d'information appelé CRYOBASE constitué d'une base de données associée à une application d'acquisition, de traitement et de consultation de ces données sur Internet (www.cryobanque.org).

2.3. Apports et limites du SNIG

2.3.1. Races à petits effectifs

En bovin, la demande de remontée exhaustive des informations de l'Etat Civil Bovin y compris pour les troupeaux hors contrôle de performance permet de faciliter l'inventaire des élevages et des animaux dans un contexte de croissance constante des effectifs. Avec la mise en œuvre de la nouvelle Loi d'Orientation Agricole, la totalité des naissances, que les animaux soient ou non à l'Etat Civil, devront être transmises au SIG. De ce fait, le travail amont d'élaboration d'inventaires d'animaux sera grandement facilité.

Néanmoins le travail de terrain ne peut être abandonné, pour plusieurs raisons : d'une part, parce comme signalé précédemment, le passage dans les élevages est nécessaire pour conseiller les éleveurs, les soutenir et vérifier éventuellement leurs dires – l'informatique ne permet pas de tout contrôler.

D'autre part dans toute base de données les erreurs sont inévitables : or si l'on conçoit que l'impact est minime dans une race comme la Charolaise ou la Prim Holstein où les effectifs se comptent par millions, il est loin d'être anodin dans une race à petits effectifs. Il est donc indispensable que les gestionnaires de races à petits effectifs puissent faire corriger ces erreurs par les organismes responsables de la gestion de la donnée en question, les Etablissements Départementaux d'Elevage pour ce qui concerne le numéro d'identification, le sexe, la race et la date de naissance.

En petits ruminants il n'existe pas d'équivalent à l'Etat Civil Bovin pour l'instant et d'ailleurs, bien que l'identification de tous les individus soit obligatoire, on trouve encore des animaux non ou mal identifiés dans beaucoup de troupeaux, souvent chez les particuliers qui ont peu d'animaux ou au contraire dans les troupeaux en pastoralisme qui se trouvent confrontés à des problèmes récurrents de pertes de boucles. Dans ce cas un identifiant imaginaire est inventé pour prendre en considération l'animal dans les bases de l'Institut de l'Elevage. On peut rester confiant sur la connaissance des animaux « à dire d'éleveurs » dans bien des cas, malgré l'absence d'identification, en raison de la petite taille du troupeau et/ou du phénotype bien particulier de chaque animal (surtout en caprins : la plupart des races locales françaises sont non standardisées sur la couleur).

Le SNIG est mis à contribution dans tous les cas où des animaux de la race locale sont au contrôle de performance, mais le nombre d'animaux concerné est très variable d'une race à l'autre. Par exemple en caprins cela peut aller de 0 (race chèvre des Fossés) à 30 % de la population (chèvre Provençale). On rencontre aussi le problème de la bonne identification du code race des animaux. Son attribution est laissée à disposition du technicien qui enregistre l'animal, et aucune vérification de cohérence n'est faite avec la généalogie de l'animal, ce qui pose deux problèmes :

- certains techniciens ne sont pas au fait des codes races autres que celui des races dominantes, et n'utilisent pas ou pas à bon escient les codes des races locales. Aussi, certaines races suivies par l'Institut de l'Elevage ne sont pas encore reconnues et ne disposent pas encore de code race, elles sont donc qualifiées au mieux comme « autre race », au pire comme « croisée » dans les bases du SNIG ;
- un animal peut être codé d'une race donnée et ses deux parents de races différentes, ce qui pose des problèmes de suivi des généalogies.

Dans ces conditions le rapatriement exhaustif des données de la race ne peut être réalisé qu'à condition qu'elle soit corrélée à une bonne connaissance du terrain, à savoir le repérage en amont des élevages d'une

race donnée au contrôle laitier et la vérification avec l'éleveur de la codification des animaux de la race.

2.3.2. Cryobanque nationale

Les collections de la Cryobanque étant constituées d'un panel d'espèce plus large que celui géré par le SNIG, il n'a pas été possible d'opérer des relations directes entre les bases de la Cryobanque et celles du CTIG : il est donc nécessaire de rapatrier l'information, par extraction, des bases du SNIG vers celle de la Cryobanque, avec les risques d'erreurs que cela comporte.

De plus, si des données sont mises à jour pour un animal au niveau du SNIG il n'existe aucun système d'alerte qui permette de signaler ce changement à l'administrateur de la base cryobanque. D'autre part pour les races à petits effectifs, comme mentionné précédemment, les informations sont en général à rechercher directement auprès des organismes gestionnaires des races, les donneurs étant rarement répertoriés dans le SNIG.

3. CONCLUSION

La mise en place du SNIG par espèce permet d'effectuer un suivi efficace des populations en sélection, dans toutes les espèces de ruminants. Ce système est basé sur la collecte des données du contrôle de performance, dont la représentativité est très variable en fonction de la filière considérée. Dans le cas particulier de l'espèce bovine, les exigences de la nouvelle Loi d'Orientation Agricole permet d'augmenter le périmètre des animaux enregistrés par le SIG pour atteindre une situation qui sera proche de l'exhaustivité.

A l'heure actuelle le SNIG ne prend pas en compte la gestion des répertoires des races locales. C'est un manque qu'il faudra combler à moyen terme, à la fois pour pérenniser, sécuriser et optimiser voire mieux valoriser ces données. Néanmoins l'intégration de ces données ne pourra pas se faire en suivant les règles actuelles des différents SNIG par espèce pour différentes raisons comme l'inexistence d'un identifiant pérenne chez certains animaux.

L'intégration des données des répertoires de l'Institut de l'Elevage impliquera une modification de certaines des données déjà présentes dans les différents SNIG. Au final le travail nécessaire pour sécuriser les bases de données des races locales est relativement lourd au regard de la quantité d'information existante, surtout en comparaison de celle existante pour les autres races : le ministère de l'agriculture a toujours été le financeur principal

de la rénovation des chaînes d'information de la génétique, ce sera à lui d'arbitrer, dans un contexte financier tendu, de la nécessité de mettre en place un tel chantier.

Siregal, un système d'information multi-espèces sur les ressources génétiques végétales françaises

DURAND Sophie

URGI, INRA (sophie.durand@versailles.inra.fr)

Bases des programmes d'amélioration des espèces d'intérêt agronomique et d'études fondamentales, les ressources génétiques patrimoniales et scientifiques sont une des richesses de l'INRA. La valorisation de ces collections de ressources génétiques passe par leur conservation, leur description, la diffusion des matériels et des données associées. Ces étapes doivent être réalisées dans le respect de la traçabilité des informations, d'autant plus que l'INRA souhaite intégrer ces collections dans des Centres de Ressources Biologiques accrédités.

SIReGal (**S**ystème d'**I**nformation coordonné sur les **R**essources **G**énétiques **V**égétales) est un système d'information permettant la gestion des collections et des données associées. Il concerne toutes les espèces végétales conservées par l'INRA, au sein des collections de travail propres à l'institut ou au sein des réseaux de conservation mis en œuvre avec des partenaires publics ou privés.

L'interface conviviale du site web de Siregal s'adresse à plusieurs types d'internautes : le grand public redécouvre la biodiversité au travers des collections de ressources génétiques ; les chercheurs choisissent des variétés en fonction de différents critères et peuvent commander des accessions.

Les principales données stockées sont de deux types : (1) les descripteurs de passeport multicultures concernent toutes les espèces végétales : taxinomie, pays d'origine, statut biologique (sauvage, mutant, hybride, ...), pedigree, ... ; (2) les données spécifiques peuvent être différentes pour chaque groupe d'espèces et décrivent la caractérisation d'un échantillon, en particulier du point de vue morphologique, agronomique ou de la résistance aux maladies.

La version actuelle de Siregal, installée sur le serveur de l'URGI (Unité de Recherche en Génomique-Info INRA), comprend le cœur de l'application : accession, provenance, taxinomie, descripteurs phénotypiques simples. Elle contient pour l'instant les données publiques des collections nationales pour la Vigne (7862 accessions), le Blé (1744 accessions), le Cerisier (178

accessions), le Châtaignier (25 accessions) et le Pois (240 accessions). D'autres espèces végétales d'intérêt agronomique ou scientifique les rejoindront. Certains projets nationaux ou internationaux qui étudient la biodiversité déposent également des données, dont l'accès peut être restreint.

Dans les années à venir, Siregal sera enrichi en renforçant son intégration dans le système d'information global de l'URGI (GnpIS). Les accessions seront ainsi liées (1) à des données phénotypiques et environnementales dans le cadre du projet Ephesis, (2) à des données de géotypage de type polymorphisme moléculaire via GnpSNP, (3) à des données de marqueurs et de cartes physiques et génétiques via GnpMap et enfin (4) aux données d'expressions stockées à l'URGI. Cette forte intégration de données permettra d'apporter une importante valeur ajoutée aux ressources génétiques en les replaçant dans le cadre de leur utilisation scientifique et en facilitant l'accès à l'ensemble des données expérimentales qui les ont caractérisées. Des liens vers des bases de données externes permettront également d'enrichir l'information.

Les données contenues dans GnpIS pourront être exportées pour être analysées dans de nombreuses applications scientifiques. Les données de ressources génétiques de Siregal pourront être utilisées pour l'aide à l'identification d'accessions, l'extraction de core-collections, l'analyse de la diversité, la visualisation de la répartition géographique d'un caractère, le calcul du coefficient de parenté, ... En complétant ces données avec les autres informations présentes à l'URGI, il sera également possible d'étudier les interactions génotype-environnement ou de faire de la génétique des populations par exemple en évaluant le polymorphisme moléculaire au sein d'une collection.

Un module de Siregal sera installé localement pour la gestion quotidienne des Centres de Ressources. Ce module traitera plus particulièrement de la gestion des stocks, des codes-barres, des commandes, etc.

Le développement et la maintenance de Siregal sont réalisés par les bioinformaticiens de l'URGI en collaboration avec les chercheurs et ingénieurs des Centres de Ressources Génétiques de l'INRA, sous le couvert d'un comité de pilotage composé d'experts du domaine. Siregal est un logiciel libre, dont le code source est disponible sur demande.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] La base de données sur les ressources génétiques végétales (Siregal) : <http://urg.versailles.inra.fr/siregal>

- [2] Les collections de ressources génétiques végétales à l'INRA : http://www.inra.fr/les_recherches/ressources_scientifiques/collections/vegetales
- [3] Les ressources génétiques au Département Génétique et Amélioration des Plantes (DGAP) de l'INRA : <http://www.inra.fr/gap/departement/activites/ressources.htm>

Modèles économiques de la valeur d'une ressource génétique

Aziz FADLAOUI

**La connaissance du processus de domestication
nous éclaire t-elle sur les choix de
conservation ?**

Paul GEPTS

Mesure et impact de la stratification génétique intra et inter races chez le chien

Christophe HITTE, Catherine ANDRÉ

Institut de Génétique et Développement, CNRS UMR6061, Université de Rennes1,
2 Av du Pr. Léon Bernard, 35043 Rennes, France

1. INTRODUCTION

Cocker anglais ou berger allemand, pitbull féroce ou yorkshire craintif, Rintintin ou Rantanplan, dans leur infinie variété, tous les chiens (*Canis lupus familiaris*) descendent du loup gris d'Asie orientale (*Canis lupus*) domestiqué par l'Homme il y a environ 15000 ans [1], [2], et se seraient répandus dans toute l'Asie et l'Europe, avant d'accompagner l'Homme dans le nouveau monde.

Le chien appartient à l'ordre des carnivores et à la famille des *Canidae* qui regroupe 34 espèces (renard, chacals, coyote, dingo, loup...) issues d'un ancêtre commun carnivore il y a 10 millions d'années. Les différentes espèces de canidés sont classées en quatre groupes dont le groupe comprenant le chien domestique qui réunit tous les membres du genre *Canis*, ainsi que le dhôle (*Canon alpinus*) et le lycaon (*Lycan pictus*). Ils possèdent tous un caryotype composé de $2n=78$ chromosomes acrocentriques. Bien qu'ils soient de taille, de poids ou possèdent des comportements et des aptitudes très différents, tous ces membres peuvent se reproduire entre eux et sont toujours fertiles avec les autres canidés. Les trois autres groupes ou clades sont les canidés du type Amérique du sud ; le clade type renard et le dernier clade comprenant les autres canidés.

Bien que la divergence chien-loup se soit opérée il y a plus de 150 siècles, le loup, son ancêtre sauvage, est toujours présent permettant des comparaisons de son matériel génétique, de son comportement et de sa physiologie. La domestication a engendré un premier goulet d'étranglement génétique dans l'histoire évolutive du chien avec pour conséquence d'intensifier le processus de dérive génétique à partir d'un pool relativement restreint d'allèles.

Le second goulet d'étranglement génétique intervenu plus récemment dans l'histoire évolutive du chien (XIXe siècle) coïncide avec la sélection par l'homme de plus de 350 races canines modernes. La création des races a pour finalité de disposer de chiens possédant des aptitudes particulières

(chiens de chasse, de garde, de traîneaux), des critères physiques, comportementaux ou esthétiques désirés. Ces pratiques de sélection intensive, souvent effectuées à partir d'un nombre limité de fondateurs et donc d'allèles, ont engendré une diminution de la variabilité génétique et par conséquent, un taux d'homozygotie important au sein de chaque race canine. Cette organisation en races, véritable structuration de l'espèce, se caractérise par une grande homogénéité phénotypique intra-race et une grande diversité inter-race qui reflètent une homogénéité génétique intra-race et une large diversité génétique entre les races [3]. Ainsi chacune des races canines est désormais considérée comme autant d'isolats génétiques à l'instar des populations humaines isolées géographiquement (islandaise par exemple et ilienne en général) ou culturellement (populations Amish ou Ashkénaze).

Le tribut à ce mode de sélection chez le chien est la forte prévalence des maladies génétiques dont la plupart sont homologues des maladies humaines. La présence de maladies génétiques fréquentes et la structuration particulière de sa population font du chien un excellent modèle pour déterminer les relations génotypes/phénotypes [4]. La recherche des causes génétiques des maladies ou des phénotypes d'intérêt très répandus au sein des races canines se réalise principalement par des études de liaisons génétiques sur des familles et par des études d'association de type "cas/contrôles" sur des populations non-apparentées. Dans ces études, une association est déterminée par des analyses statistiques qui prennent en compte notamment les variations entre fréquences alléliques observées et attendues des marqueurs et les déséquilibres de liaisons entre ces marqueurs. Ces tests peuvent être fortement influencés par le phénomène de stratification, c'est-à-dire une hétérogénéité génétique cachée au sein de la population étudiée qui peut donner lieu à la mise en évidence de fausses associations qu'il est difficile de distinguer des véritables associations recherchées.

2. RÉSULTATS

Nous avons caractérisé la stratification intra-race d'origine géographique (Europe *vs* États-Unis) pour quatre races de chien à partir de données de marqueurs SNP couvrant près de la moitié du chromosome 1 canin. Lors de cette étude menée en collaboration avec le groupe du Dr. Elaine Ostrander [3], nous avons modélisé et évalué l'impact de la stratification dans les études d'associations. À partir de plus de 700 marqueurs SNPs génotypés sur 120 chiens appartenant à quatre races (n=30), Bouvier Bernois (Bernese Moutain Dog, BMD), Golden Retriever (GR), Flat Coated Retriever (FCR) et Rottweiller (ROT) échantillonnées pour moitié en Europe et aux États-

Unis, nous avons utilisé plusieurs indicateurs permettant d'évaluer la diversité génétique, telle que la mesure du niveau d'hétérozygotie génétique (F_{st}), les distributions de fréquence alléliques et de diversité haplotypique et des approches de clustering de population. Nous avons déterminé que les races FCR et ROT présentent une homogénéité génétique qui n'est pas affectée par l'origine géographique différente des chiens testés alors que les BMD et de manière encore plus nette les GR présentent une stratification génétique qui corrèle avec l'origine géographique des chiens. La mise en évidence de populations structurées au sein des races canines comme celles observées au sein des BMD et GR reflète l'histoire génétique et sociale propre à chaque race, les patrons d'établissement de standard de races qui peuvent varier entre pays ou continents et les pratiques variables d'élevage et de sélection des races. De plus, l'expansion des races ou à contrario des événements conduisant à une réduction d'effectif (goulet d'étranglement) d'une race, ainsi que des facteurs d'ordre aléatoire au cours du temps (stochastiques) vont aussi contribuer à la stratification génétique d'un race canine. Ainsi, nos résultats montrent que les GR présentent la plus forte diversité génétique attribuable à l'origine continentale des animaux. Cette variabilité est reflétée par le plus fort taux de polymorphisme des SNP, une différenciation génétique significative ($F_{st} = 0,073$) et le plus grand nombre de combinaisons haplotypiques au sein des régions génomiques en déséquilibre de liaison. Les approches de clustering séparent efficacement les deux populations européenne et américaine de GR sur la base de leur distribution allélique. Ces résultats corrélerent avec la popularité de la race GR en Europe et aux États-Unis. Aux États-Unis, les GR ont vu leur popularité augmenter dès les années 1930 et constituent aujourd'hui la 4^e race la plus répandue sur les 155 races enregistrées par l'American Kennel Club (AKC). En France les GR appartiennent à la race qui arrive en seconde position en terme d'effectif dans le classement de 2006. Bien que l'origine ancestrale des GR soit unique, la popularité de la race a établi un large pool d'allèles à la fois en Europe et aux États-Unis combiné à pas ou peu de croisements et donc de flux allélique entre les deux populations. La race BMD bien qu'ayant un faible niveau de SNP polymorphiques présente une différenciation génétique détectable ($F_{st} = 0,05$) entre sa population échantillonnée en Europe et américaine. Ces résultats sont prévisibles par l'histoire de la race créée en Suisse en 1890 dont les chiens ont été exportés aux États-Unis dès 1926 comme l'indiquent les registres de l'AKC. La race a connu une forte popularité dans de nombreux pays dans les années 70 (15^e rang en France en 2006). Alors que les effectifs de cette race sont importants, les chiens originaires de Suisse sont encore régulièrement importés aux États-Unis, expliquant un flux d'allèles toujours présent qui contribue à niveler l'amplitude de la stratification génétique sous-jacente entre les BMD d'Europe et des États-Unis. À l'inverse pas de différenciation

génétique n'a pu être mis en évidence chez les FCR qui présentent peu de polymorphisme allélique et une absence de clusterisation. Ces résultats sont en relation avec les effectifs faibles que connaît cette race des deux côtés de l'Atlantique qui limitent la variabilité génétique et a fortiori une différenciation spécifique des distributions alléliques associées aux continents. Dans le cas des rotweillers, l'histoire de la race présente la particularité d'une réduction drastique des effectifs de la race pendant la seconde guerre mondiale, ces chiens, étant souvent utilisés pour la garde pendant la guerre, ont payé un lourd tribut à ces utilisations à risques. La race a connu par la suite une phase d'expansion et connaît actuellement un nouveau déclin de popularité. Ces goulets d'étranglement semblent être reflétés par la faible diversité génétique mesurable dans nos études. Au delà de la détection de la stratification au sein des races canines, nous avons testé son impact dans les études de cas-contrôles menées sur des cohortes échantillonnées sur des populations structurées. Nos résultats ont montré que des études d'associations réalisées sur une cohorte de GR présentant une stratification génétique peuvent aboutir à la considération de faux-positifs pour les locus candidats associés à un trait d'intérêt.

3. CONCLUSIONS ET PERSPECTIVES

La caractérisation de la diversité génétique au sein des populations de chiens a pour objectif de détecter et corriger la stratification au sein d'une cohorte utilisée pour une étude génétique. Ainsi, dans le cadre du projet Européen LUPA, notre groupe coordonne le programme de recherche de gènes impliqués dans les mélanomes canins. Cette étude porte sur 500 chiens appartenant à deux races principalement (caniche et schnauzer) qui constituent deux cohortes pour des études de type cas/contrôles qui serviront à rechercher les causes génétiques de la prédisposition aux mélanomes. Ces chiens étant recrutés sur plusieurs sites européens et disposant de plusieurs standards de race (couleur et taille différente, caniche nain, moyen, grand, schnauzer moyen et grand), nous étudierons le niveau de stratification de chacune de ces races et leur possibilité d'être différenciées en sous-entités génétiques en raison de leurs origines géographiques différentes, de leur appartenance à des lignées particulières, ou de pratiques ou de standards d'élevage spécifiques.

Une utilisation de la stratification inter races est-elle envisageable dans les études d'association génétique ? Nous élaborons une approche bioinformatique capable de détecter et de mesurer l'amplitude de la variation des distributions alléliques qui existe entre les deux races, puis de la soustraire afin de tester une étude d'association de type cas-race_1 vs contrôles-race_2 par exemple, pour évaluer les variations entre fréquences

alléliques observées et attendues pouvant cibler le ou les locus associés au trait d'intérêt.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T: Genetic evidence for an East Asian origin of domestic dogs. *Science* 2002, 298(5598):1610-1613.
- [2] Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillen S, Vila C: Ancient DNA evidence for Old World origin of New World dogs. *Science* 2002, 298(5598):1613-1616.
- [3] Quignon P, Herbin L, Cadieu E, Kirkness EF, Hedan B, Mosher DS, Galibert F, Andre C, Ostrander EA, Hitte C: Canine Population Structure: Assessment and Impact of Intra-Breed Stratification on SNP-Based Association Studies. *PLoS ONE* 2007, 2(12):e1324.
- [4] Galibert F, Andre C: The dog and its genome. *Med Sci (Paris)* 2006, 22(10):806-808.

La place des projets territoriaux dans les prises de décision sur le devenir des races animales locales

Adeline LAMBERT-DERKIMBA⁽¹⁾, François CASABIANCA⁽¹⁾,
Etienne VERRIER⁽²⁾

¹ INRA SAD, UR 45 Laboratoire de Recherche sur le Développement de l'Élevage,
Quartier Grossetti, 20250 Corte (FRANCE)

² INRA/AgroParisTech, UMR1236 Génétique et Diversité Animale, 16 rue Claude
Bernard, 75231 Paris Cedex 05 (FRANCE)

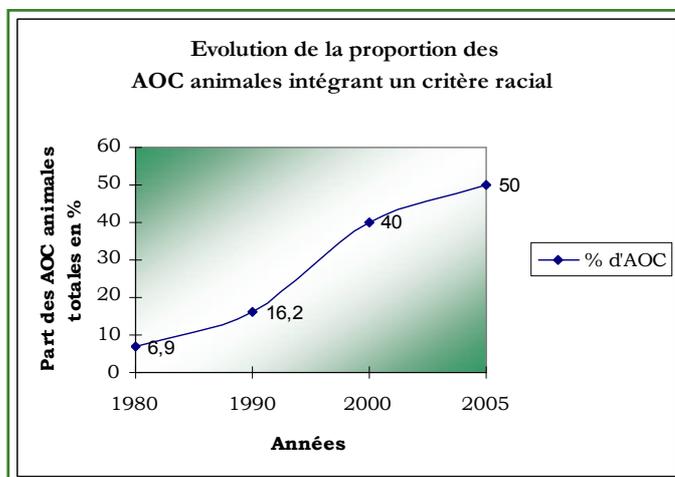
1. INTRODUCTION

Notre travail s'intéresse aux races animales locales, c'est-à-dire aux races dont l'élevage est restreint à un territoire délimité, et à la manière dont certains projets territoriaux ont des répercussions sur le développement et l'orientation de ces races dès lors qu'ils les mobilisent. Parmi ces projets, nous avons choisi d'analyser avec précision la place des produits AOC, liés aux territoires par les ressources qu'ils mobilisent (les races, les systèmes d'élevage, les savoir-faire locaux), dans les processus de décision sur le devenir des races locales concernées. Pour cela nous avons analysé deux questions majeures : (1) quel impact l'association entre race(s) et AOC peut avoir sur les prises de décisions concernant les ressources génétiques d'un point de vue quantitatif (développement et diffusion de la race) et qualitatif (objectifs de sélection, structure génétique des troupeaux), et (2) quelles configurations rendent possible la prise en compte effective des règles de production sous AOC dans la définition des objectifs de sélection ? L'analyse a porté sur les cas des Alpes du Nord (fromages AOC Abondance, Beaufort, Reblochon et Tome des Bauges / races bovines Abondance, Tarentaise et Montbéliarde, cette dernière n'étant pas locale), et de la Corse (projet d'AOC charcuterie de Corse / race porcine Nustrale).

2. DYNAMIQUE D'ÉVOLUTION DU COUPLAGE AOC/RACES LOCALES

On note une franche évolution de la part des AOC comprenant un critère racial à partir des années 90. Cela correspond aux volontés conjointes de l'INAO, via son règlement de 1992 visant à assurer une meilleure

légitimité territoriale, et des groupes professionnels, qui veulent protéger les ressources locales et renforcer les règles de production. Cette évolution est particulièrement forte pour les produits fromagers dans les montagnes de l'Est de la France (Alpes du nord et Jura) et dans le sud (Pyrénées, Corse). Les produits carnés, quant à eux, inscrivent quasi systématiquement les races dans leur cahier des charges (Maine Ajou, Taureau de Camargue, volaille de Bresse).



En conséquence, le système d'acteurs autour de la gestion des races locales évolue. On voit en effet apparaître de nouveaux acteurs regroupés en syndicats AOC qui sont partie prenante du devenir des races locales (Roncin, 2000). Ces syndicats, orientés par les injonctions des comités de l'INAO et des commissions d'enquêtes, tentent de transmettre leur vision du projet racial. Les associations de gestion des races qui, jusqu'à présent, avaient surtout à faire à des utilisateurs individuels de ces ressources génétiques, doivent donc apprendre à composer avec ces nouveaux utilisateurs collectifs (Audiot *et al.*, 2004).

3. PRISE EN COMPTE DES CONDITIONS DE PRODUCTION DES PRODUITS SOUS AOC

3.1. Développement et diffusion des races

Le couplage AOC / race locale permet, dans les cas étudiés, de stimuler le développement et la diffusion des races (Verrier et Bouffartigue, 1993). L'obligation raciale induit en effet une nécessité pour les exploitations de se fournir en animaux de race pure, ce qui peut augmenter fortement la

demande en reproducteurs. On observe alors une mobilisation territoriale plus ou moins forte favorisant la structuration des exploitations de producteurs de reproducteurs, et la mise en réseaux des éleveurs pour la valorisation des jeunes de race pure. Cette mobilisation des instances territoriales (interprofessions, organismes de développement local) s'explique en grande partie par la nécessité, pour les AOC, de trouver une disponibilité en animaux suffisante pour satisfaire les filières de production sous AOC. Le développement de la race devient ainsi indispensable à la réussite de la filière AOC.

Cette configuration permet dans le cas de la race porcine Nustrale, en Corse, de dynamiser fortement le marché des reproducteurs de race pure. Des financements pour l'équipement des exploitations (parc saillie, cabanes à mise bas) sont débloqués par l'ODARC (Office de Développement Agricole et Rural de Corse) sur les lignes « produits de qualité ». Cela permet d'augmenter le nombre d'exploitations susceptibles de produire des reproducteurs, pour une filière de production qui, jusqu'à aujourd'hui, n'était que très peu structurée (à l'échelle individuelle et collective).

Dans le cas des Alpes du Nord, un groupe de travail pour le développement de la race Abondance est monté afin de répondre à la demande du syndicat d'augmenter la proportion d'animaux de race Abondance dans l'aire de production de l'AOC Abondance. Ce groupe, financé par l'interprofession, entamera une sensibilisation des producteurs à la valorisation des jeunes animaux de race pure et à leur diffusion dans des réseaux gérés par l'UPRA.

3.2. Orientation génétique des races locales

La question de l'orientation génétique des races a été analysée sous l'angle de l'impact des règles de production, tout particulièrement de la limite de productivité laitière (AOC Beaufort et Tome des Bauges) sur les objectifs de sélection et sur la structure génétique des troupeaux. Ces analyses ont concerné les cas de couplage dans les Alpes du Nord.

3.2.1. Impact sur les objectifs de sélection

Nous avons considéré l'index synthétique (multi caractères) de sélection, dit ISU, comme indicateur des orientations génétiques des races bovines. Dans le cas de la race Tarentaise, on remarque dans l'ISU une pondération de l'index synthétique laitier (INEL) de 0,45, alors que toutes les autres races laitières lui affectent un poids de 0,5. Si cette différence n'a que peu d'impact sur les progrès génétiques attendus, cette décision traduit la prise en compte de la limite de productivité de l'AOC Beaufort, à laquelle la race est fortement liée. L'UPRA, en cohérence avec le syndicat AOC, promeut ici un type de système de production qui ne pousse pas à l'intensification laitière.

Dans le cas de la race Montbéliarde, on remarque qu'il n'y a pas de prise en compte des AOC des Alpes du Nord dans l'ISU, et encore moins de la limite de productivité imposée par certaines d'entre elles. La race Montbéliarde est fortement associée aux AOC de sa région d'origine (la Franche Comté) et a une extension nationale, voire internationale. Dans ce contexte, les projets territoriaux des Alpes du Nord ne peuvent avoir qu'un impact marginal sur le développement et l'orientation de la race.

Dans le cas de la race Abondance, on observe une difficulté des gestionnaires raciaux à construire un ISU (jusqu'à juin 2008). En effet, la race est présente sur un territoire hétérogène (piémont/montagne), avec des projets territoriaux différents (conditions de production diverses) et ses éleveurs sont tiraillés entre divers débouchés (AOC avec ou sans limite de productivité, non AOC). Cela explique la difficulté pour ses responsables à construire un objectif de sélection clair et accepté par tous (Lambert-Derkimba *et al.*, 2006). On peut souligner ici un paradoxe : le développement numérique de la race (race bovine laitière locale au cheptel le plus important en France) conduit à son hétérogénéité de localisation et donc aux difficultés présentées ci-dessus.

3.2.2. Impact sur la structure génétique des troupeaux

Dans les Alpes du Nord, l'analyse de la variabilité des index de valeur génétique des vaches selon la localisation des troupeaux fait ressortir des différences entre zones (résultats en cours de publication). Notamment, pour les races Abondance et Montbéliarde, la moyenne des index de valeur génétique pour la quantité de lait produite par vache et par lactation est significativement plus faible dans les zones AOC que dans les zones non AOC et, en race Abondance, cette moyenne est significativement plus faible dans les zones AOC avec limite de productivité que dans les zones AOC sans limite.

4. CONFIGURATIONS RENDANT POSSIBLE L'EXISTENCE DE CES PRISES EN COMPTE

4.1. Critères explicatifs

L'analyse de la répartition des effectifs de race représente un facteur explicatif majeur de la prise en compte des conditions de production du Beaufort pour la gestion de la race Tarentaise ($\approx 50\%$ des effectifs de la race), de la non prise en compte des AOC des Alpes du Nord pour la race Montbéliarde (moins de 2%), et de la prise en compte des AOC Reblochon et Beaufort pour la race Abondance (respectivement $\approx 26\%$ et $\approx 10\%$ des effectifs de la race).

L'analyse des interactions entre les deux collectifs de gestion, collectif « race » et collectif « AOC », met en exergue deux éléments clé de la compréhension de ces prises en compte. Tout d'abord, les organismes de sélection (organismes raciaux, au sens de la LOA de 2006), peuvent être financés en partie par les syndicats AOC. Dans ce cas-là, le syndicat est en droit d'attendre une prise en compte de sa vision du projet racial. En absence de financement, les arguments des syndicats AOC peuvent recevoir une écoute plus « distraite ». Le deuxième élément est l'existence de représentations croisées entre les deux collectifs. On entend par là le fait qu'une même personne soit membre du conseil d'administration ou du bureau, à la fois du syndicat AOC et de l'organisme de sélection. Cette représentation croisée peut alors être considérée comme étant formelle (mandaté par le syndicat pour le représenter au sein de l'association) ou informelle (non mandaté mais à même de faire passer des idées).

Comme montré précédemment, la marge de manœuvre dont disposent les éleveurs est primordiale (gestion de la structure génétique des troupeaux). Mais cette marge de manœuvre dépend de la gamme de reproducteurs proposée aux éleveurs, qui va varier selon les races. De plus, les éleveurs ne réalisent pas systématiquement le choix des animaux qu'ils intègrent à leur troupeau. En effet, les agents inséminateurs sont aussi considérés comme des conseillers pour la planification des accouplements. Les éleveurs ne choisiront alors pas eux-mêmes leurs reproducteurs (ils ont cependant généralement confiance dans les choix de l'agent inséminateur qui connaît leur exploitation depuis des années).

4.2. Lecture des cas analysés à travers les critères d'interprétation

Ces différents critères nous permettent de lire les situations étudiées et de les comprendre, comme ils peuvent représenter des grilles de lecture pour de nouvelles situations à explorer.

Synthèse des critères sur les cas des Alpes du Nord et de la Corse

Race	AOC	Financement spécifique	Doubles casquettes	Représentativité effectifs	Gammes des reproducteurs	Choix effectif	Divers spécifiques
Abondance	Abondance	Non	Informelles	2.1%	Orientations diversifiées	Varié : technicien ou éleveur	Même nom
	Beaufort	Non	Non	9.7%			Poids politique
	Reblochon	Non	Informelles	26%			
	Tome des Bauges	Non	Informelles	0.6%			
Tarentaise	Abondance	Non	Non	0.3%	Orientation claire	Varié : technicien ou éleveur	
	Beaufort	Oui	Formelles et informelles	48%			Projet territorial dans la durée
	Reblochon	Non	Non	3.3%			
	Tome des Bauges	Oui	Non	2.5%			
Montbéliarde	Abondance	Non	Non	0.2%	Orientations diversifiées	Varié : technicien ou éleveur	
	Reblochon	Non	Non	1.6%			
	Tome des Bauges	Non	Non	<0.1%			
Corse	Charcuterie	Oui	Informelles	?	Faible	Préconisation technicien	Synchronicité

5. CONCLUSION

L'existence d'une prise en compte des règles de production des AOC dans les processus de décision sur le devenir des races locales n'est pas systématique (Lambert-Derkimba, 2007). Cette prise en compte est, dans nos cas, particulièrement présente lors d'un couplage des AOC avec les races originaires du territoire concerné (Abondance, Tarentaise et Nustrale). Mais les projets AOC ne sont qu'un type de projet territorial parmi d'autres. Dans ces jeux de construction des décisions sur le devenir des races locales, qu'en est-il des autres projets territoriaux (PNR, autres produits...)? Comment ces différents projets interagissent – ils entre eux dans la prise de décision sur le devenir des ressources génétiques ?

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] AUDIOT, A., BOUCHE, R. B., H., CASABIANCA, F., GAILLARD, C., ROCHE, B., TRIFT, N., et STEYAERT, P. (2004). Populations animales locales et produits de qualité. Comment valoriser transformer la ressource génétique. *In* Les Actes du BRG n°5 "Un dialogue pour la diversité génétique", Lyon. 577-592
- [2] LAMBERT-DERKIMBA, A. (2007). L'inscription des races locales dans les conditions de production des produits animaux sous AOC: enjeux et con-

séquences pour la gestion collective des races mobilisées, AgroParisTech, Paris, 284 p.

- [3] LAMBERT-DERKIMBA, A., CASABIANCA, F., et VERRIER, E. (2006). La race bovine Abondance à la croisée des chemins. *In* "6ème colloque du BRG "Des ressources partagées"", La Rochelle.
- [4] RONCIN, F. (2000). Les enjeux de la conservation des races locales, position du problème pour les signes de qualité: cas des appellations d'origine. *In* "Races locales et biodiversité animale", Rambouillet.
- [5] VERRIER, E., et BOUFFARTIGUE, B. (1993). Les AOC, élément du maintien des races régionales? L'exemple des Alpes du Nord et réflexion sur d'autres situations. *In* "Races domestiques en péril: 4ème Journée d'Etudes de la Société d'Ethnozootechnie ", Vol. 52, Paris.

Caractérisation et utilisation de la diversité génétique de la vigne

L. Le CUNFF, M. DI VECCHI STARAZ, R. BACILIERI, S. NICOLAS,
V. LAUCOU, T. LACOMBE, A. DOLIGEZ, A-F ADAM-BLONDON,
J-M BOURSQUOT, P. THIS

Une séquence haute qualité du génome de la vigne est maintenant disponible. C'est une étape indispensable pour accéder à l'ensemble des gènes de la vigne. Elle va aussi permettre une meilleure exploitation des ressources génétiques naturelles de cette espèce notamment au travers de la génétique d'association pour la dissection de caractères complexes d'intérêt viticole. Cette exploitation nécessite cependant une bonne connaissance de la diversité génétique de l'espèce mais également de sa structuration.

La diversité génétique de la vigne a été étudiée à l'aide de 20 marqueurs microsatellites bien répartis sur le génome. Du matériel de l'espèce *Vitis vinifera* et de son compartiment sauvage *Vitis sylvestris* a été analysé ainsi que des accessions de différentes espèces de *Vitis* dans le cadre d'un projet Tripartite "Coregrapegen" en vue de définir des core-collections représentant la diversité de chaque compartiment.

L'exposé présentera le travail réalisé pour *Vitis vinifera* à partir de l'analyse de la collection de vigne de l'INRA (2486 variétés). Une core-collection emboîtée de 92 accessions a ainsi été définie pour représenter l'ensemble de la diversité SSR de la collection. Le noyau de base de cette core-collection (24 variétés) est actuellement utilisé pour développer une carte de diversité SNP de la vigne. De même la connaissance de la structuration de la diversité de l'espèce a permis la définition d'échantillons pour des études de génétique d'association pertinents.

Métagénomique et diversité génétique et métabolique

Denis LE PASLIER

CEA / Genoscope

La métagénomique ou génomique environnementale est une nouvelle approche permettant d'analyser l'ensemble des génomes de tous les microorganismes d'un environnement ou d'une niche écologique. Cependant, vu la très grande taille des génomes eucaryotes, cette approche est généralement réduite à la population microbienne. Pour ce faire, la population microbienne est isolée, son ADN extrait et séquencé. L'analyse bioinformatique des lectures permet d'identifier les gènes, peut permettre de reconstituer de grands morceaux de génomes, voire des génomes complets. Une des grandes difficultés des approches métagénomiques réside dans le fait que l'immense majorité des microorganismes n'est pas cultivable. Cette situation résulte en partie du fait que de nombreuses divisions de l'arbre phylétique bactérien restent inexplorées, notamment en raison de notre incapacité actuelle à cultiver des représentants de ces divisions. Il en résulte que de nombreuses informations de séquence ne peuvent être reliées facilement à l'organisme dont ils dérivent.

Le but du projet "Cloaca maxima" est de contribuer massivement à l'inventaire des gènes des procaryotes des usines de traitement des eaux usées domestiques. Le processus de traitement des eaux usées est une succession de plusieurs écosystèmes interconnectés dans lesquels les processus biochimiques qui se produisent sont encore mal compris. Ces processus dépendent des activités métaboliques de communautés microbiennes incomplètement inventoriées. Les objectifs principaux du projet sont : 1) d'établir un inventaire des micro-organismes présent aux étapes principales des usines de traitement des eaux ; 2) de commencer un inventaire des gènes de ces micro-organismes à partir du métagénome ; 3) d'établir un inventaire des enzymes à partir de l'analyse métagénomique ; 4) de caractériser de grands segments de génome des micro-organismes présents, avec un intérêt particulier pour les plus représentés et pour les microorganismes non encore cultivés ; 5) d'utiliser cette ressource métagénomique pour caractériser de nouvelles enzymes et voies métaboliques.

Tout comme pour de nombreux projets de génomique, le développement des technologies de séquençage à haut débit est en train de révolutionner les approches métagénomiques.

Whole genome sequencing of the fungal plant pathogens *Botrytis cinerea* and *Sclerotinia sclerotiorum*

Marc-Henri LEBRUN

Botrytis cinerea and *Sclerotinia sclerotiorum* are destructive fungal pathogens of many economically important crops provoking grey and white mould respectively. The Broad Institute has released the assembly of a 4-5 x genomic sequence from *B. cinerea* strain B05-10 (TMRI/Syngenta) as well as the 7-8 x genomic sequence of *S. sclerotiorum* (Broad Institute). In parallel, the French national sequencing center (Genoscope) has sequenced the genomes of grapevine and of two of its pathogens: Stolbur phytoplas and *B. cinerea* strain T4. A large collection of ESTs (> 80 000) have been obtained for both *B. cinerea* and *S. sclerotiorum*. Automatic gene prediction with *ab initio* and similarity softwares have been applied to the three genomes revealing 14.000 to 16.000 potential gene calls for both species. Comparative analyses of both *B. cinerea* strains and of *B. cinerea* vs. *S. sclerotiorum* has revealed a common set of 8400 gene calls, 1200 of which are absent from other fungal genomes. In addition both species present a high degree of sequence similarity (80 % average protein sequence identity among orthologues), and of synteny at the contig level. These two species significantly differ in their repetitive DNA content, since *S. sclerotiorum* has more repetitive elements (7.7 %) than *B. cinerea* (3 - 4 %). This difference is likely related to the expansion of some families of transposable elements (TEs) in *S. sclerotiorum*. TEs seem to have been subject to RIP in both species although at higher levels in *B. cinerea*. The comparison between *B. cinerea* and *S. sclerotiorum* offers an interesting opportunity to compare the genomes of two closely related necrotrophic plant pathogens helping to understand evolutionary trends that have shaped their genomes. This comparison has highlighted differences between *B. cinerea* and *S. sclerotiorum* in secondary metabolism gene clusters, transcription factors and mating type loci, while sets of secreted proteins and metabolic enzymes are almost similar. These findings suggest that both species have evolved with different strategies to behave as non specialized necrotrophic plant pathogens.

Diversité moléculaire et structure du génome chez les levures *S. cerevisiae* œnologiques

JL LEGRAS, B. BLONDIN*, S. CASAREGOLA, S. DEQUIN*

*UMR1083 Sciences pour l'Oenologie, INRA, 2 place Viala,
34060 Montpellier, France
dequin@supagro.inra.fr

Alors que la séquence du génome de la levure *S. cerevisiae* S288C a été achevée en 1996, le lien entre diversité génétique et origine des levures n'a été montré que tout récemment. Les souches œnologiques présentent une relativement faible diversité liée à une origine unique, située vraisemblablement en Mésopotamie et la structuration des populations semblent indiquer une migration simultanée avec la vigne. De plus, certains groupes de levure ont été visiblement sélectionnés par l'homme pour leur propriétés technologiques particulières (levures de voile, levures de Champagne, hybrides cryophiles...). Aujourd'hui le génome de plusieurs souches de levure vient d'être séquencé, dont celui de la souche EC1118 (diploïde, hétérozygote) isolée en Champagne et commercialisée dans le monde entier, sélectionnée pour ses propriétés technologiques. Le génome de cette souche diffère de celui de la souche de laboratoire de référence de 0,5%, et présente un taux d'hétérozygotie interne de 0,08%. Ce génome présente des particularités remarquables, en particulier la présence de nouvelles régions, non présentes dans l'espèce *S. cerevisiae*, qui portent des gènes principalement impliqués dans des fonctions métaboliques et de transport. L'étude de la distribution, au sein de l'espèce *S. cerevisiae*, de ces fragments potentiellement impliqués dans l'adaptation des souches, est en cours.

Une information hors norme : les savoirs traditionnels

Anne LUXEREAU

Génomique et évolution des Actinobactéries

NORMAND P⁽¹⁾, ALLOISIO N⁽¹⁾, PUJIC P⁽¹⁾, VALLENET D⁽²⁾,
MEDIGUE C⁽²⁾, BERRY A⁽³⁾, SANTOS C⁽⁴⁾, TAVARES F⁽⁴⁾, DAUBIN V⁽⁵⁾

⁽¹⁾UMR 5557 Ecologie Microbienne, Université Lyon1, 43 bd du 11 novembre 1918,
Bât 741, 69622 Villeurbanne cedex, France

⁽²⁾Genoscope/UMR-CNRS 8030, Atelier de Génomique Comparative, 2 rue Gaston
Crémieux, F-91006 Evry, France

⁽³⁾Department of Plant Sciences, Mailstop 1, PES 1210, University of California,
Davis, CA 95616. USA

⁽⁴⁾Institute for Molecular and Cellular Biology, Cellular and Applied Microbiology,
University of Porto, Rua Campo Alegre, 823, 4150-180 Porto, Portugal

⁽⁵⁾UMR5558, Laboratoire de Biométrie et Biologie Évolutive, Université Lyon I,
CNRS, F-69622 Villeurbanne, France
Philippe.Normand@univ-lyon1.fr

Les interactions plantes-microorganismes conditionnent la croissance des plantes et la productivité végétale, leur étude constitue donc une possibilité d'améliorer la qualité et la quantité de nourriture disponible ainsi que le couvert végétal pour prévenir l'érosion, produire du bois de chauffage ou à d'autres fins biotechnologiques. La capacité des plantes de distinguer entre microorganismes symbiotiques et pathogènes, la capacité des microorganismes d'induire des programmes de développement spécifiques et des cascades de signalisation dans la plante, le contrôle de la physiologie de fonctions spécifiques sont autant d'aspects qui peuvent nous permettre ces améliorations. La génétique classique a permis des avancées considérables dans la compréhension de ces interactions mais a montré ses limites s'agissant de deux des microorganismes symbiotiques majeurs, les champignons endomycorhiziens impossibles à cultiver et l'actinobactérie fixatrice d'azote *Frankia* impossible à transformer génétiquement.

Cette bactérie appartient au sous-ordre des Frankinae aux caractéristiques physiologiques contrastées, avec ainsi les *Acidothermus* (acidothermophile des sources hydrothermales), les *Geodermatophilaceae* (radiorésistantes colonisant les pierres), les *Sporichthyaceae* (morphologie unique et complexe), les *Kineosporiaceae* (radiorésistantes retrouvées dans les ateliers de traitement des combustibles nucléaires). C'est pourquoi plusieurs programmes de détermination de la séquence génomique ont été entrepris par le Génoscope et par le DoE. Ces programmes sont terminés pour certaines souches de *Frankia*, *Acidothermus* et de *Kineococcus* et en cours pour d'autres souches de *Frankia* et pour les *Geodermatophilaceae*.

Dans le cas des *Frankia*, la principale conclusion qui émerge de ces travaux (Normand *et al.*, 2007a) de génomique sont que la taille des génomes varie fortement entre souches très proches avec des génomes de 5,3 et 9,1 Mb pour des souches n'ayant que 2% de distance dans le 16S. Une deuxième conclusion est que les duplications, gains et pertes de gènes sont importants et localisés principalement au terminus de réplication. Enfin une troisième conclusion est l'absence des gènes *nod* canoniques (Normand *et al.*, 2007b) et donc que la signalisation chimique ne se fait pas de façon exactement semblable que chez les *Rhizobium*. Cette découverte contraste avec la publication récente de la découverte chez la plante symbiotique *Casuarina* de la kinase SymRK (Gherbi *et al.*, 2008) qui permet la nodulation par *Frankia*, kinase impliquée dans la cascade de signalisation symbiotique chez *Rhizobium*. Une première analyse transcriptomique permet de voir de nombreux gènes surexprimés dans la fixation d'azote *in vitro* et aussi en symbiose. Un programme en cours au Génoscope de séquençage des EST d'*Alnus* et de *Casuarina* devrait permettre de savoir si les autres éléments de la cascade de signalisation symbiotique des légumineuses sont présents.

Chez *Acidothermus cellulolyticus*, le plus proche voisin phylogénétique de *Frankia*, le génome non encore publié est encore plus petit, à 2,44 Mb, avec de nombreux gènes liés à la cellulolyse. Ce génome d'un thermophile montre un début d'adaptation thermique par surreprésentation de certains acides aminés, ce qui est surprenant chez un microorganisme n'ayant divergé de la lignée des *Frankia* que depuis peu de temps puisque la distance dans le 16S n'est que de 5% (Barabote *et al.* In preparation).

L'obtention de nombreux génomes dans ce groupe contrasté permettra d'identifier plus finement les transferts et duplications de gènes qui se sont produits à taux élevé. Ces nombreux génomes disponibles d'organismes aux caractéristiques physiologiques contrastées permettront aussi de corrélérer le nombre de gènes appartenant à différentes catégories avec le biotope choisi.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Gherbi H, *et al.* (2008) SymRK defines a common genetic basis for plant root endosymbioses with AM fungi, rhizobia and *Frankia* bacteria. *Proc Natl Acad Sci USA* 105:4928–4932.
- [2] Normand P, *et al.* (2007a) Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Research* 17: 7-15.
- [3] Normand P, *et al.* (2007b) Exploring the genomes of *Frankia* sp. *Physiologia Plantarum* 130: 331-346.

De la collecte de données à l'information pour les décideurs : la production d'indicateurs de biodiversité

Dominique RICHARD⁽¹⁾ et Eléonore CHARVOLIN⁽²⁾

⁽¹⁾Centre Thématique Européen sur la Diversité Biologique/
Agence Européenne de l'Environnement

⁽²⁾Fondation Française pour la Recherche sur la Biodiversité

La diversité génétique est l'un des trois composants de la diversité biologique, telle que reconnue par la Convention sur la Diversité Biologique (Rio, 1992). En s'engageant à la mise en œuvre de cette convention, les Parties Contractantes, dont la France et l'Union Européenne, se sont fixées des objectifs de conservation se traduisant respectivement par une stratégie nationale pour la biodiversité et une stratégie européenne, accompagnées de plans d'action spécifiques. Ces démarches s'inscrivent elles-mêmes dans l'objectif général d'«arrêter la perte de biodiversité d'ici 2010». Afin de suivre les progrès dans la réalisation de cet objectif général, les pays signataires de la CDB se sont entendus sur la définition et le développement d'un jeu d'indicateurs, très agrégés, destinés à renseigner les décideurs sur l'état et les tendances des composants de la biodiversité (habitats/écosystèmes, espèces, ressources génétiques), l'intégrité des écosystèmes et les services écosystémiques, les menaces à la biodiversité et l'utilisation durable de cette biodiversité. Au niveau européen, un programme intitulé «Streamlining European Biodiversity Indicators 2010» (SEBI 2010), coordonné par l'Agence Européenne de l'Environnement, a été lancé en 2005, visant à assurer la production de 26 indicateurs de biodiversité, sur la base des données disponibles auprès des pays. Ces indicateurs devront être mis à jour régulièrement jusqu'en 2010 et au-delà. Concernant la diversité génétique, un indicateur relatif aux ressources génétiques animales a émergé des propositions d'un groupe d'experts nommés par quelques pays européens actifs au sein du RFP (Réseau Européen sur la gestion des ressources génétiques animales – ERFP). Cet indicateur sera, au même titre que les 25 autres indicateurs de biodiversité, publié en 2009 dans un rapport de synthèse SEBI 2010 commandité par la Commission Européenne, et en cours de préparation par l'Agence Européenne de l'Environnement. Un indicateur relatif aux ressources génétiques végétales devrait par la suite être développé.

Lors de l'élaboration des différentes propositions d'indicateurs relatifs aux ressources génétiques animales une question majeure s'est posée au groupe d'experts : à partir de quelles données possibles travailler ? En effet, plusieurs bases de données sont accessibles et ont chacune des avantages et des inconvénients. Les bases de données mondiales sont facilement accessibles mais aucune donnée « évolutive », nécessaire à l'élaboration d'un indicateur pertinent, n'y est répertoriée. D'autres, régionales, sont en construction, et plusieurs des champs qui seraient utiles ne sont pas encore ou sont insuffisamment renseignés. Enfin les bases de données nationales, très riches, alimentées directement à partir de données de terrain, grâce aux différents acteurs de la gestion des ressources génétiques, sont difficilement comparables pour une analyse au niveau européen. Les systèmes sociaux, économiques et réglementaires sont différents d'un pays à l'autre et entraînent une implémentation et utilisation variable des différents champs. L'un des enjeux est de faire les ponts entre les acteurs de terrain détenant les données, les acteurs de la réglementation, les compétences méthodologiques existantes pour les indicateurs et les utilisateurs de ces indicateurs, qu'ils soient nationaux et/ou européens. Le même langage est nécessaire pour pouvoir produire des indicateurs européens qui reflètent les tendances sur l'ensemble des acteurs. Il en découlera, si ces conditions sont remplies, un panorama reflétant la réalité des tendances et des évolutions. Il permettra aux décideurs d'anticiper les besoins des acteurs de la conservation des ressources génétiques au moyen de mesures réglementaires, de mieux adapter les programmes de recherche, d'élaborer des politiques publiques etc. Les enjeux autour des indicateurs, et l'expérience acquise mettent clairement en évidence que la coordination, du local au régional, puis du national à l'europpéen, et à travers les 3 niveaux de la biodiversité doit être renforcée.

La mesure de la diversité génétique : quels outils, quelles méthodes, quel futur ?

Joëlle RONFORT

Nouveaux outils moléculaires pour décrire la biodiversité

Pierre TABERLET

Les développements technologiques dans l'analyse des génomes

Patrick WINCKER

Les analyses à l'échelle génomique ont connu une accélération remarquable au cours des dernières années, en particulier grâce à l'apparition de nouvelles technologies de séquençage de l'ADN. Nous présenterons les évolutions en cours de ces technologies, ainsi que plusieurs de leurs applications majeures. Le séquençage de génomes (*de novo* et re-séquençage comparatif) sera abordé, ainsi que les nouvelles possibilités d'étude du transcriptome et des interactions ADN-protéines. Nous montrerons dans quels cas ces méthodes peuvent se substituer économiquement à la méthode de séquençage jusqu'alors utilisée, et dans quels cas elles ne le permettent pas encore. Nous montrerons quelques exemples de développements informatiques nécessaires pour tirer pleinement partie de la quantité très importante de données générées par ces nouvelles méthodes. Enfin, nous décrirons les méthodes actuellement en phase de développement pré-commercial dont l'apparition attendue permettra le séquençage à très bas coût.