Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Eva M. Wolf
born in: Karlsruhe, Germany
Oral examination: 30.08.2017

The evolutionary history of *Cochlearia* L.

Cytogenetics, phylogenomics and metabolomics of a cold relic in a warming world

Referees: Prof. Dr. Marcus A. Koch
apl. Prof. Dr. Claudia Erbar

Table of Contents

Summary

The cold-adapted genus *Cochlearia* L. (scurvy grass), a young polyploid complex within the Brassicaceae family (Cruciferae), displays a range of highly interesting cytogenetic and ecotypic characteristics and might serve as an excellent model system to study general evolutionary mechanisms such as polyploidization, hybridization, or cold and edaphic adaptation. The presented study provides, for the first time, comprehensive cytogenetic and highly-resolving phylogenomic analyses, and first metabolomic insights into the *Cochlearia* cold response. Thus, the findings presented herein might constitute a good starting point for further in-depth analyses of said evolutionary aspects e.g. based on population-genomic datasets.

In **chapter 1**, the cytogenetic evolution within the genus *Cochlearia* is analyzed via both conventional chromosome counts and flow cytometry measurements. Based on a comprehensive literature review on published chromosome counts, the geographical distribution of cytogenetic variability is described, suggesting an early evolutionary separation of the two diploid karyotypes (2n=12 and 2n=14). The high frequency of aberrant chromosome numbers in polyploid taxa is interpreted as a result of frequent interploidal hybridization, given the near absence of interspecific fertility barriers, and thus reflecting the dynamics of polyploid evolution within the genus *Cochlearia*. Moreover, a correlation between genome size and chromosome number, as well as genome downsizing in polyploid taxa are revealed.

Chapter 2 provides comprehensive phylogenomic analyses based on Illumina high-throughput sequencing data. Chloroplast and mitochondrial phylogenies are largely in congruence and indicate a glacial survival of the whole genus in arctic refuge areas as well as repeated adaptation to alpine habitats in Central Europe. Divergence time estimates, based on complete chloroplast genomes, imply a diversification of the whole genus over the course of several Pleistocene glaciations within the last ~700,000 years. Results from nuclear data analyses support both the clear evolutionary separation of the two diploid karyotypes as described in chapter 1, as well as the basal phylogenetic position of arctic taxa as shown in organellar phylogenies, and they reveal new insights into the evolutionary origins of the different polyploid taxa.

In **chapter 3**, the metabolomic analysis of the *Cochlearia* cold response via metabolite profiling using gas chromatography-mass spectrometry (GC-MS) is described. Based on temperature-related bioclimatic variables (WorldClim), four bioclimatic ecotypes are defined and utilized as group priors for statistical analyses. All studied plants/ecotypes show strong metabolomic adjustments in reaction to 20 days of cold treatment under 5°C, especially with regard to increased levels of soluble carbohydrates and amino acids. Statistical analyses do not show a strong discrimination of the four bioclimatic ecotypes based on the analyzed metabolites, yet slight intrageneric variation among the bioclimatic clusters is described, implying similarities in the cold response between arctic and alpine taxa.

Zusammenfassung

Die kalt-adaptierte Gattung *Cochlearia* L. (Löffelkräuter) aus der Familie der Brassicaceae (Kreuzblütler) bietet als evolutionäres Modellsystem exzellente Möglichkeiten zur vertiefenden Erforschung verschiedener evolutionärer Vorgänge, wie Polyploidisierung, Hybridisierung oder der Anpassung an arktische und alpine Kaltstandorte. Ziel der vorliegenden Arbeit war es, mittels umfassender zytogenetischer und phylogenomischer Analysen eine hochauflösende Darstellung der Evolution dieser Gattung in Zeit und Raum zu erstellen, sowie anhand metabolomischer Untersuchungen erste Einblicke in die Physiologie verschiedener Ökotypen im Hinblick auf deren Metabolismus unter Temperaturstress zu gewinnen. Die hierbei gewonnenen Erkenntnisse stellen unter anderem eine ideale Ausgangslage für weiterführende populations-genomische Studien bezüglich der vermutlich parallel entstandenen Adaptation an extreme Kaltstandorte dar.

In **Kapitel 1** wird die zytogenetische Evolution innerhalb der Gattung *Cochlearia* anhand konventioneller Chromosomenzählungen sowie durch Genomgrößenbestimmung mittels Durchflusszytometrie analysiert. Eine umfassende Literaturrecherche bezüglich publizierter Chromosomenzahlen ermöglicht die detaillierte Beschreibung der geographischen Verbreitungsmuster zytogenetischer Variabilität innerhalb der Gattung und legt eine frühe evolutionäre Trennung der beiden diploiden Karyotypen (2n = 12 und 2n = 14) nahe. Das gehäufte Auftreten numerischer Chromosomenaberrationen in polyploiden Arten wird als Folgeerscheinung der durch das Fehlen ausgeprägter Fertilitätsbarrieren häufigen Hybridisierungsereignisse zwischen Taxa unterschiedlicher Ploidiestufen interpretiert und reflektiert somit die hochdynamische Evolution polyploider Taxa innerhalb der Gattung *Cochlearia*. Ferner werden eine Korrelation zwischen Genomgröße und Chromosomenzahl, sowie *Genome Downsizing* in Polyploiden beschrieben.

Kapitel 2 liefert umfassende phylogenomische Analysen basierend auf Illumina-Hochdurchsatz-Sequenzdaten. Plastidäre und mitochondriale Phylogenien stimmen in weiten Teilen überein und legen unter anderem die Existenz arktischer Refugialareale an der Basis der Gattung, sowie parallel entstandene Adaptation an hochalpine Standorte in Zentraleuropa nahe. Divergenzzeiten innerhalb der Cochlearieae, ermittelt auf der Basis kompletter Plastidengenome, belegen die Diversifikation der Gattung *Cochlearia* innerhalb der letzten 700.000 Jahre, im Verlauf mehrerer Eiszeiten des Pleistozäns. Kerngenomische Daten stützen sowohl die in Kapitel 1 beschriebene deutliche evolutionäre Trennung der diploiden Karyotypen, als auch die basale phylogenetische Position arktischer Arten, und liefern zudem neue Einblicke in die Evolution der polyploiden Arten.

Kapitel 3 fasst die metabolomische Untersuchung der Gattung mittels Gaschromatographie mit Massenspektrometrie-Kopplung (GC-MS) zusammen. Basierend auf temperaturbezogenen bioklimatischen Variablen (WorldClim) werden vier klimatische Ökotypen definiert, die als Gruppen-Prioren für die Metabolitenanalyse fungieren. Auf eine 20-tägige Kältebehandlung bei 5°C zeigen alle untersuchten Pflanzen/Ökotypen eine deutliche Reaktion, was sich erwartungsgemäß besonders in einer starken Akkumulation von Kohlenhydraten und Aminosäuren äußert. Die physiologischen Phänotypen der vier bioklimatischen Gruppen lassen sich anhand der erhobenen metabolomischen Daten statistisch nicht deutlich separieren, dennoch werden leichte Unterschiede in der Reaktion auf Kalt- und Kontrollbedingungen (20°C) ersichtlich, die Ähnlichkeiten in der Kälteadaptation zwischen arktischen und alpinen Arten nahelegen.

General Introduction

Evolution in the context of a changing environment

Contemporary research in evolutionary biology is facing plenty of challenges and opportunities given the unprecedented wealth of data produced by the latest technologies in different fields (reviewed by e.g. Gomez-Cabrero et al., 2014; Li and Chen, 2014). The combination of different approaches and high-throughput, or 'omics', studies promises new insights and discoveries in topics that urgently need to be advanced such as the impacts of recent climate and environmental change on biodiversity and plant evolution (reviewed by e.g. Nadeau and Jiggins, 2010; Franks and Hoffmann, 2012). Here, case studies resolving the evolutionary history of ecologically interesting taxa can provide valuable information in order to get the big picture one day.

Present-day climate warming and environmental changes are affecting life on earth in various ways (e.g. physiological effects, changes in the distribution range; for review see Parmesan, 2006) but the genetic consequences of these abiotic fluctuations are complex and not yet fully understood (Alsos et al., 2012; Franks and Hoffmann, 2012). Cold terrestrial habitats, like the circum-arctic regions will be among the first to be strongly influenced by global warming and the resulting changes in abiotic conditions, such as earlier snowmelt and prolonged growing-seasons (Parmesan, 2006; Alsos et al., 2012; Barrett et al., 2015). Besides other factors like nutrient supply, the temperatures of air and soil have already been shown to be associated with different plant traits in arctic regions (e.g. Chapin et al., 1995; Arft et al., 1999; Hollister et al., 2005; Hudson et al., 2011). Earlier snowmelt followed by a return to freezing temperatures can severely damage the arctic vegetation without the insulating snow cover (e.g. Bokhorst et al., 2009), thus against expectation, cold tolerance will be even more important for plants of these cold habitats under global warming (Armstrong et al., 2015).

Three very different strategies allow for the survival of an organism facing environmental change: 1) to deal with the new abiotic stresses by phenotypic plasticity, 2) to move into more suitable habitats by migration or 3) to adapt to the new conditions by local adaptation (Aitken et al., 2008; Williams et al., 2008). Especially for immobile plants, option number 2 might not be fast enough in many cases, given the speed of the current and future environmental change (Davis and Shaw, 2001; Jump and Penuelas, 2005). Likewise, phenotypic plasticity might often be limited and not sufficient for drastic ecological changes. Since adaptation is one of the most important evolutionary responses to new environmental conditions and yet many details about the genetic prerequisites for local adaptation are still in the dark (Savolainen et al., 2013), there is a pressing need to improve our knowledge in this area of research. Besides field studies and experimental approaches measuring the impact of today's global warming, another possibility is to study the genetic footprints in taxa that once underwent adaptation to new ecological conditions arising during past climatic events like the recurrent cycles of glaciation and deglaciation in the Pleistocene.

In order to fully understand the mechanisms behind these adaptations, it is necessary to study different levels and aspects of evolution in the study system. Evolutionary changes affecting the genome as a whole, e.g. via duplication of whole genomes or single chromosomes, likely have an impact on the diversification of the study system (Soltis and Soltis, 1999; Otto and Whitton, 2000; Adams and Wendel, 2005; Soltis et al., 2009; Soltis et al., 2014; Tank et al., 2015). These events may act on different taxonomic levels, within families or genera and even within a single species. Nevertheless, in most cases it won't be sufficient to analyze genome scale changes in order to reconstruct the evolutionary history of these taxa. Thus, fine-scale phylogenetic studies are needed in order to fully resolve these past processes in space and time. Here, present-day high-throughput sequencing technologies, the so-called next generation sequencing (NGS), provide the best opportunities for data generation and analysis enabling the generation of highly resolved phylogenetic insights at different taxonomic levels (e.g. Whittall et al., 2010; Mandel et al., 2014; Edger et al., 2015). Finally, in parallel with today's ever-growing genomic resources there is a growing need to link the new genomic information with ecological data of the respective phenotypes (Faria et al., 2014). Aside from classical approaches such as QTL mapping or functional tests, metabolomic analyses offer new possibilities for a better knowledge of a taxon's physiological phenotype under a given or changing environment with its specific biotic and abiotic stresses (for reviews see e.g. Sardans et al., 2011; Brunetti et al., 2013) and can be utilized as a first step connecting phenotype and genotype (Fiehn, 2002).

The study system: *Cochlearia* and tribe Cochlearieae

Model organisms like *Arabidopsis thaliana* (L.) Heynh. are providing a wealth of scientific information and evolutionary insights based hereon (for review see Hedges, 2002). Together with their often short life cycles, this facilitates experimental setup as well as data interpretation, which is why they usually are the first choice to study pending evolutionary questions. Yet, at the same time, the fit of a model species for many studies might be limited (Bolker, 2012). *A. thaliana*, for example, is not a perfect model in comparative studies on genome evolution within the Brassicaceae family, because of the drastic processes of genome and chromosome downsizing that have acted on its genome (Schranz et al., 2006). Likewise, *Arabidopsis*, being an annual species, can't reveal every aspect of the evolutionary responses in perennial plants and since its natural distribution does not cover extreme arctic/alpine habitats, it should not be the only candidate to study the evolution of climate adaptation to extreme cold.

However, the famous model species *A. thaliana* - and the vast store of knowledge that has been accumulated about it - is of course one of the reasons why research interest in the Brassicaceae family is and has been very high for several decades (see e.g. Franzke et al., 2009). Other interesting model systems (e.g. *Arabis*, *Cardamine*, *Draba*) as well as ornamentals (e.g. *Aubrieta*, *Matthiola*) and economically important plants like the different *Brassica* species (e.g. canola, cabbage) further strengthen the scientific relevance of the family (Hohmann et al., 2015). Besides this, the general diversity of the family with regard to species richness, genetic diversity, ecological

and physiology diversity is strikingly high and the list of well-studied non-model Brassicaceae genera remarkably long, making this family are very good playground for comparative studies on polyploidy and other evolutionary aspects (Marhold and Lihová, 2006; Kagale et al., 2014).

The taxonomical system of the Brassicaceae family is complex and has been revised several times (Warwick et al., 2006; Warwick and Al-Shehbaz, 2006; Al-Shehbaz, 2012; Al-Shehbaz et al., 2014; German et al., 2015). In 2006, Beilstein et al. proposed a division of core Brassicaceae into three major lineages based on chloroplast sequence data (ndhF; Beilstein et al., 2006), which were supported by several studies hereafter (Bailey et al., 2006; Koch et al., 2007; Franzke et al., 2009; Lysak et al., 2009). More recently, a family-wide transcriptome-based multi-locus study by Huang et al. (2015) identified six major clades, thereof five (clades A to E) within the so called core Brassicaceae, that is all extant lineages excluding Aethionemeae (e.g. Price et al., 1994; Galloway et al., 1998). According to the online-accessible *BrassiBase* knowledge database (Kiefer et al., 2013), the Brassicaceae family currently comprises 3990 species that have been ascribed into 52 tribes and 338 genera.

Together with its sister-genus *Ionopsidium* Rchb., the genus *Cochlearia* L. (engl. scurvy grasses) forms the tribe Cochlearieae, a rather small tribe, comprising only about 30 taxa depending on taxonomical treatment (24 accepted species and 5 subspecies according to *BrassiBase* (Kiefer et al., 2013); see Table 1), that forms an isolated evolutionary lineage within the Brassicaceae family. As revealed by the study of Huang et al. (2015), the Cochlearieae are one of four tribes showing conflicting tree topologies and thus having an uncertain phylogenetic position within the Brassicaceae. This was interpreted as a result of putative ancient allopolyploidization events near the origin of the respective tribes. Owing to several of its enthralling characteristics, research interest in the tribe Cochlearieae has been growing over the last decades in fields as diverse as cytogenetics/cytogenomics (e.g. Ceccarelli and Cionini, 1993; Mandakova et al., 2017), comparative developmental genetics (Bowman, 2006), studies of salt and heavy-metal tolerance (Reeves, 1988; Nawaz et al., 2017) or secondary metabolite analyses (Brock et al., 2006; Dauvergne et al., 2006) among others (see Koch, 2012).

The genus *Cochlearia* as a study system profits from the comparatively close relationship with *A. thaliana* and other Brassicaceae model species. Furthermore, like many Brassicaceae species it has a rather small genome size, facilitating data generation and analysis and it shows a wealth of interesting and diverse traits, which seem to have evolved within a very short evolutionary timespan (e.g. Koch et al., 1996; Koch et al., 1998; Koch et al., 1999; Koch, 2012). The genus contains mainly short-lived perennials but also annuals and facultative annuals. Morphological differences between the different taxa are often insufficient for a clear species determination (Gill et al., 1978; Koch et al., 1996) and ideally, a combination of morphology, ecotype, location, and, if possible, chromosome number should be considered for species identification. Due to the lack of unambiguous morphological characters, taxonomy within the tribe Cochlearieae has often been confusing and under debate for decades (e.g. Saunte, 1955; Gill, 1973; Gill et al., 1978). Therefore, out of the previously six genera that once belonged to the tribe (*Bivonaea* DC., *Cochlearia* L., *Ionopsidium* Rchb., *Minaea* Lojac., *Pastorea* Tod. ex Bertol. and *Thlaspi* L.), only two were retained after the last major revision (Koch, 2012), based on the internal transcribed spacer regions ITS1 and ITS2, the nuclear chalcone synthase (CHS) gene and the chloroplast intergenic region

trnL-trnF. *Minaea* and *Pastorea* had already been considered synonyms of *Ionopsidium* before, *Thlaspi* had been shown to be unrelated with the tribe Cochlearieae (Mummenhoff and Koch, 1994; Mummenhoff et al., 1997b, a; Koch and Mummenhoff, 2001; Koch and Al-Shehbaz, 2004) and several species that were formerly placed into *Bivonaea*, *Cochlearia*, *Minaea*, *Pastorea* or *Thlaspi* were now finally assigned to *Ionopsidium* (now containing six species) based on present-day methods and information. A new tribe Bivonaeeae was established, and all Mediterranean taxa within Cochlearieae are now members of the genus *Ionopsidium*, forming a western Mediterranean clade (Koch, 2012), adapted to at least seasonally warm and dry habitats, with scattered distribution areas in Spain, Portugal, Italy and some regions in North Africa. The distribution of the genus *Cochlearia* ranges from the coasts of northern Portugal up to circum-arctic regions of Canada and Alaska with Central Europe as a center of distribution and species diversity. Many species are endemics or show highly scattered occurrences and most of the taxa are (highly) endangered. Although the taxonomic picture became much clearer over the last decades, many open questions remained, especially regarding the taxonomic confusion along the coasts, in Great Britain and in arctic regions.

Several previous studies have revealed first insights into the complex and putatively reticulate phylogenetic evolution of the study group (e.g. Koch et al., 1996; Koch et al., 1998; Koch et al., 1999; Koch, 2002; Koch et al., 2003; Kochjarová et al., 2006; Cieślak et al., 2007) including divergence times for the main lineages within the tribe Cochlearieae (Koch, 2012). According to these, the two genera diverged about 13.8 million years ago (mya) in mid-Miocene. Split times of the main groups within *Ionopsidium* could be dated to mid to late Miocene, either before or in concert with the Messinian salinity crisis, that lasted from about 5.96 to 5.33 mya (Krijgsman et al., 1999) and had a dramatic impact on the Mediterranean ecosystems (Krijgsman, 2002; Rodríguez-Sánchez et al., 2008; Garcia-Castellanos et al., 2009). Further divergence within the genus *Cochlearia* did not start before mid or late Pleistocene (Koch, 2012) and split times of this presumably rapid radiation could not be fully resolved based on the genetic markers used so far. Yet several other evolutionary aspects of this species complex have been revealed. It has been shown to have evolved two base chromosome numbers, namely n=6 and n=7 (Saunte, 1955; Löve and Löve, 1961; Gill, 1971a), several major phylogenetic lineages and a range of ploidy levels (from diploid to octoploid/dodecaploid) putatively over several cycles of glaciation and deglaciation during the Pleistocene (e.g. Koch et al., 1996; Koch et al., 1998; Koch et al., 1999). Likewise, the genus developed a wide ecological amplitude, with taxa occurring in as different habitat types as coastal regions (especially along the Atlantic coastlines), lowland/mountainous habitats associated with calcareous cold springs, high alpine areas on siliceous bedrocks, arctic habitats or even heavy-metal contaminated soils. A connection between the different cytotypes and ecotypes is the cold character of the described habitats – this way contrasting with the sister clade *Ionopsidium* which is purely found in arid habitat types (Koch, 2012). Figure 1 summarizes previous phylogenetic findings and hypotheses with the respective combination of ecotype and substrate specificity. This evolutionary and ecological setup of the study system nicely illustrates the excellent possibilities to investigate different evolutionary aspects such as the (parallel?) evolution of cold adaptation, the evolution of substrate specificity, the evolutionary impact of polyploidization, hybridization and introgression. A thorough knowledge of the species complex -

the different diploid and polyploid taxa, their ecology and evolutionary background - is of utmost importance for any of these investigations.

The different taxa within the genus *Cochlearia* (16 species and 4 subspecies according to BrassiBase) can be best categorized according to their ploidy level and ecotype. Central European diploid taxa include a single coastal species, namely *C. aestuaria* (Lloyd) Heywood (2n=2x=12), restricted to the coast of northern Spain. This and a basal phylogenetic position in previous studies let to the assumption of the involvement of *C. aestuaria* progenitors in the formation of the different polyploid coastal taxa (Koch et al., 1996). The diploid inland species *C. pyrenaica* DC. (2n=2x=12) shows a wide but scattered distribution in mountainous regions from the French Pyrenees to the Ukraine with few northern occurrences in Great Britain. It is mainly growing on base-rich substrate, in or near calcareous cold-springs but is also found on heavy-metal sites in Belgium and Great Britain. Based on its present-day distribution combined with information gained from previously applied molecular markers (e.g. Koch et al., 1996; Koch et al., 1998; Koch 2002) a role in the formation of different inland polyploid taxa is likely (see below). *C. macrorrhiza* (Schur) Pobed. (2n=2x=12) was a diploid species endemic to Austria, formerly found on lower elevations in a restricted area of lowland fens near Moosbrunn (Vienna basin, Lower Austria), and despite strong efforts for a reestablishment, it is meanwhile extinct in the wild (Koch and Bernhardt, 2004; Mrkvicka et al., 2015). It was shown to constitute a separate lineage, originating from other diploid inland species (Koch et al., 2003). Finally, *C. excelsa* Zahlbr. Ex Fritsch (2x=2n=12) is a rare high alpine species, growing on siliceous bedrocks in the eastern Austrian Alps and as for *C. macrorrhiza*, close relationships with other inland diploids, namely *C. pyrenaica*, have been determined (Koch et al., 1996; Koch et al., 2003). In general, for the recent diploid inland taxa in Central Europe a rather young age has been hypothesized based on both distribution patterns and close phylogenetic relationships (Koch et al., 1996). Besides these Central European diploids, that all share a base chromosome number of n=6, there is a circum-arctic group of diploid species for which both in-depth evolutionary histories and, based on these, a critical taxonomical re-evaluation have been missing so far (Koch, 2012). This arctic group comprises exclusively diploids with a base chromosome number of n=7 and a single case of n=6 in southern Iceland. *C. groenlandica* L. (2n=2x=14) has a wide, circum-arctic distribution (including Iceland) and shows considerable variability in size and morphology depending on habitat type, e.g. tundra, rocky beaches and slopes, tidal flats, mud flat bird nesting sites (Al-Shehbaz and Koch, 2010; Zmudczyńska-Skarbek et al., 2013). Several taxa have meanwhile been reduced to synonyms of *C. groenlandica*, such as *C. arctica* Schltdl. ex DC., *C. fenestrata* R. Br., *C. oblongifolia* DC. (Warwick et al., 2006; for a full list of synonyms refer to the BrassiBase (Kiefer et al., 2013)). Nordal and Laane (1990) suggested to place all arctic diploids into *C. groenlandica* L. including the 2n=12 diploids from southern Iceland but despite the current status of a synonym, a separate taxonomic status of the latter as *C. islandica* Pobed. (2n=2x=12) is preferred herein. *C. sessilifolia* Rollins (2n=2x=14) is a presumably annual Alaskan diploid found on gravel bar, at lagoon-type estuaries where it can be underwater at high tide, or along seashores (Al-Shehbaz and Koch, 2010). Due to its distinct morphological characters, it is treated as a separate arctic taxon, although detailed phylogenetic studies have been missing so far (Al-Shehbaz and Koch, 2010). *C. tridactylites* Banks ex DC. (2n=2x=14) is a biennial/perennial species distributed along western Canadian coastal areas

where it is growing e.g. on calcareous sandstone or limestone gravel substrate, on cliffs and shores (Al-Shehbaz and Koch, 2010).

The polyploid section of the genus *Cochlearia* is divided into coastal and inland polyploids. The coastal environments inhabit a series of taxa with different ploidy levels, distributed along the coasts of continental Europe and Great Britain, with partly overlapping distribution ranges. Ecological preferences between coastal polyploid taxa are slightly different, especially with regard to the degree of salt tolerance, but still contact zones can arise occasionally and since barriers to gene flow are only poorly established throughout the genus, hybridization even across ploidy levels has been described frequently in the wild (e.g. Saunte, 1955; Fearn, 1977; Pegtel, 1999) and has been furthermore demonstrated in various hybridization experiments, also including inland and arctic taxa (Crane and Gairdner, 1923; Gill, 1971a, b, 1973, 1976; Gill et al., 1978). Therefore, the coastal species complex is taxonomically very complicated also due to a high plasticity in response to the habitat type (Elkington, 1984; Nordal, 1988). According to BrassiBase, the coastal polyploid taxa comprise *C. officinalis* L. (2n=4x=24), *C. scotica* Druce (2n=4x=24), *C. danica* L. (2n=6x=42) and *C. anglica* L. (2n=8x=48 or 2n=10x=60) as well as three subspecies of *C. officinalis*, namely *C. officinalis* subsp. *integrifolia* (Hartm.) Nordal & Stabbetorp (2n=4x=24), *C. officinalis* subsp. *norvegica* Nordal & Stabbetorp (2n=4x=24) and *C. officinalis* subsp. *officinalis* L. (2n=4x=24). The latter have been described in Northern Scandinavia, representing three different ecotypic occurrences on or near a) beaches and bird cliffs (*C. officinalis* subsp. *officinalis*), b) estuarine habitats (*C. officinalis* subsp. *norvegica*), c) cold springs (*C. officinalis* subsp. *integrifolia*) (Nordal et al., 1986; Nordal and Stabbetorp, 1990). *C. officinalis* s. str. is a tetraploid biennial to perennial species that is widely distributed along the northern coasts of continental Europe and Great Britain especially on gravelly beaches, drier areas of salt-marshes and on (nesting) sea cliffs (Gill et al., 1978; Nordal and Stabbetorp, 1990). It was suggested to be of autopolyploid origin from a diploid taxon by several authors, most likely related to present-day coastal *C. aestuaria* (Gill et al., 1978; Elkington, 1984; Koch et al., 1996; Koch et al., 1998). The tetraploid *C. scotica* was described to have a more compact phenotype with shorter inflorescences compared to the larger *C. officinalis* s. str., but due to weak morphological distinctiveness it was also considered to constitute a morphological variant of *C. officinalis* (e.g. Gill, 1971a, 1973). The annual hexaploid *C. danica* is a pioneer species in coastal sand dune systems experiencing seasonally high temperatures. Besides these characteristics that are exceptional within the genus, it is also the only coastal species in Central Europe with a base chromosome number of n=7 and a specific chromosomal constitution with clear size differences between the three chromosome sets (Crane and Gairdner, 1923; Saunte, 1955; Gill, 1976). It is common along the coasts from Portugal to Great Britain and southern Scandinavia and furthermore spreading inland along highways. Despite the shared base chromosome number of n=7 (Gill, 1976), *C. groenlandica* was ruled out as progenitor species of *C. danica* based on cpDNA analyses (Koch et al., 1996) and a relation to *C. tatrae*, another hexaploid species with 2n=42 (see below), was also negated based on its distinct chromosomal constitution (Gill, 1976). A putative allopolyploid origin from tetraploid *C. officinalis* relatives and *C. aestuaria*-like (*C. pyrenaica* s.l.) diploids was hypothesized (Koch et al., 1996; Koch et al., 1998) but the evolutionary origin of this cryptic species has not been finally resolved yet. Lastly, *C. anglica* is an octoploid species found in estuarine habitats or along

tidal rivers. It shows some distinct morphological characters especially regarding the shape of leaves which are ovate to oblong. According to previous studies, it likely originated from *C. officinalis* via (auto-)polyploidization (Koch et al., 1996; Koch et al., 1998).

Another group of polyploid taxa is found at inland stations of continental Europe, most of them highly endemic and endangered. *C. bavarica* Vogt (2n=6x=36) is endemic to two mountainous areas in southern Germany where it is found near carbonate-rich cold water springs. An allopolyploid origin predating human influence (inter- or postglacial) from *C. pyrenaica* and *C. officinalis*-related parental taxa was suggested for this hexaploid species based on RAPD and isozyme data, and postglacial fragmentation of a formerly wider distribution area was hypothesized (Koch et al., 1996; Koch et al., 1998; Koch, 2002). *C. polonica* Fröhl. (2n=2x=36) is a highly endangered endemic found along river banks in a small area in southern Poland and its persistence in the natural localities required extensive conservation efforts (Cieślak et al., 2011). As for *C. bavarica*, allopolyploidy from *C. pyrenaica* and *C. officinalis* was suggested for *C. polonica* based on detected isozyme variation (Koch et al., 1998). *C. tatrae* Borbás (2n=6x=42) is a high-alpine taxon from the High Tatra mountains in Slovakia and Poland. In contrast to *C. bavarica* and *C. polonica* it has a base chromosome number of n=7, yet as already suggested by Bajer (1951), allozyme data by Koch et al. (1998) likewise indicated an allopolyploid origin from *C. pyrenaica* and *C. officinalis*. Moreover, affinities between *C. groenlandica* and *C. tatrae* have been detected based on cpDNA data (Koch et al., 1996) and a close relationship with East Carpathian hexaploid *C. borzaeana* has been revealed from AFLP data (Koch et al., 2003) but no final conclusions on the formation of its distinct karyotype and its phylogenetic origin have been made. Finally, *C. borzaeana* (Com. et Nyár.) Pobed. (2n=8x=42/48) is an octoploid inland species endemic to East Carpathian mountain ranges in Romania. Is has been formerly analyzed cytogenetically and described as *C. pyrenaica* (with chromosome counts of 2n=42; Stefureac and Lungeanu, 1976) and *C. pyrenaica* var. *borzaeana* (2n=48; Stefureac and Panzaru, 1978), but was finally considered a separate taxon (Pobedimova, 1970; Kochjarová et al., 2006). The evolutionary origin of this octoploid species has not been fully revealed yet, but similar processes and putative parental taxa as described for other inland polyploids have been suggested (Koch et al., 2003; Kochjarová et al., 2006).

In Great Britain, the taxonomic situation is more complex. Besides the above mentioned populations of diploid *C. pyrenaica*, tetraploid inland populations have been described that are facing somewhat colder, more severe conditions and show a more alpine phenotype. There exists some taxonomic confusion as these plants have been referred to as either inland populations of *C. officinalis* (Clapham et al., 1990; Koch et al., 1998) or as a separate taxon *C. pyrenaica* ssp. *alpina* (Bab.) Dalby (e.g. Gill, 2008) with closer affinities to the diploid *C. pyrenaica* complex. The latter taxonomic treatment is preferred in this study, hereafter abbreviated as *C. alpina*. Diploid inland populations have also been distinguished as *C. alpina* (Bab.) H.C. Watson but it has been suggested to include these in *C. pyrenaica* (Gill, 1971b). *C. micacea* E.S. Marshall, with a karyotype of 2n=4x=26, represents another tetraploid inland species endemic to high altitude locations in Scotland. It was originally described to be a separate taxon based on morphological characteristics and later defined to be cytogenetically distinct from the other upland populations by its karyotype (Gill et al., 1978; e.g. Clapham et al., 1990). It was also suggested to be synonymous

with *C. officinalis* subsp. *integrifolia* (Hartm.) Nordal & Stabbetorp (2n=4x=24) found on inland stations in Norway (Nordal and Stabbetorp, 1990) but its status as a separate taxon was recently questioned by (Gill, 2008), based on morphological and genetic AFLP data that did not show a clear separation from other upland taxa in northern Great Britain.

Detecting the basic mechanisms that triggered the rapid diversification and adaptation of the genus *Cochlearia* to the different habitat types might bring forward valuable information on general aspects of plant evolution.

Table 1. List of accepted *Cochlearia* taxa according to *BrassiBase* (Kiefer et al., 2013) in alphabetical order together with information on cytotype, ploidy level, ecology and distribution area.

Taxon name	Cytotype	Ecology	Distribution
Cochlearia aestuaria (Lloyd) Heywood	2n=2x=12	coastal	Northern Spain
Cochlearia anglica L.	2n=8x=48/60	coastal	Coasts of Central Europe (incl. UK)
Cochlearia bavarica Vogt	2n=6x=36	mountainous	Southern Germany
Cochlearia borzaeana (Coman & Nyaudy) Pobed.	2n=8x=48	mountainous	Romania
Cochlearia danica L.	2n=6x=42	coastal	Coasts of Central Europe
Cochlearia excelsa Zahlbr. ex Fritsch	2n=2x=12	high alpine	Eastern Austrian Alps
Cochlearia groenlandica L. (incl. *Cochlearia islandica* Pobed. with 2n=2x=12)	2n=2x=14	arctic	Circum-arctic regions and Iceland
Cochlearia macrorrhiza (Schur) Pobed.	2n=2x=12	lowland	Eastern Austria (extinct in the wild)
Cochlearia micacea E.S. Marshall	2n=4x=26	mountainous	Uplands of northern UK
Cochlearia officinalis L.	2n=4x=24	coastal	Coasts of Central Europe
Cochlearia officinalis subsp. *integrifolia* (Hartm.) Nordal & Stabbetorp	2n=4x=24	coastal	Northern Scandinavia
Cochlearia officinalis subsp. *norvegica* Nordal & Stabbetorp	2n=4x=24	coastal	Norway
Cochlearia polonica A. Fröhl.	2n=6x=36	calcareous springs	Poland
Cochlearia pyrenaica DC.	2n=2x=12	calcareous springs	Central Europe (incl. UK)
Cochlearia pyrenaica subsp. *alpina* (Bab.) Dalby	2n=2x=12	calcareous springs	UK, Scotland
Cochlearia scotica Druce	2n=4x=24	coastal	Northern UK, Ireland
Cochlearia sessilifolia Rollins	2n=2x=14	arctic	Alaska
Cochlearia tatrae Borbás	2n=6x=42	high alpine	High Tatra Mountains; Czech. Rep., Poland
Cochlearia tridactylites Banks ex DC.	2n=2x=14	arctic	Eastern Canada

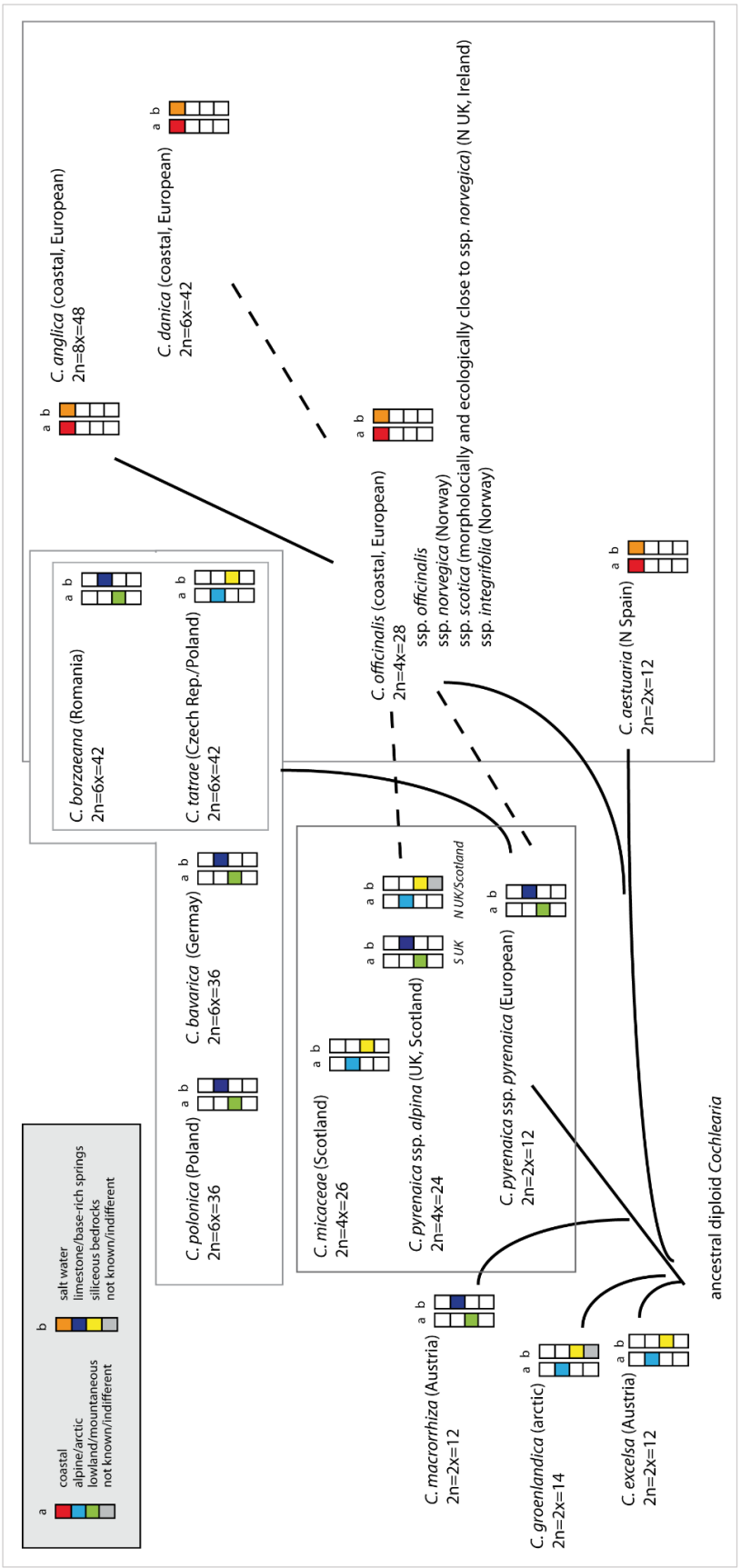

Figure 1. Figure made by M.A. Koch (unpublished), summarizing previous evolutionary hypotheses regarding the genus *Cochlearia*, combined with information on ploidy levels and ecology. The latter is given as a combination of a) species distribution, and b) substrate specificity.

References

Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. Current opinion in plant biology **8,** 135-141.

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T., and Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications **1,** 95-111.

Al-Shehbaz, I.A. (2012). A generic and tribal synopsis of the Brassicaceae (Cruciferae). Taxon **61,** 931-954.

Al-Shehbaz, I.A., and Koch, M.A. (2010). *Cochlearia*. Flora of North America Editorial Committee , eds. 1993. +Flora of North America North of Mexico. 20+ vols. New York and Oxford. **7,** 514-516.

Al-Shehbaz, I.A., German, D.A., Mummenhoff, K., and Moazzeni, H. (2014). Systematics, tribal placements, and synopses of the *Malcolmia* s.l. segregates (Brassicaceae). Harvard Papers in Botany **19,** 53-71.

Alsos, I.G., Ehrich, D., Thuiller, W., Eidesen, P.B., Tribsch, A., Schönswetter, P., Lagaye, C., Taberlet, P., and Brochmann, C. (2012). Genetic consequences of climate change for northern plants. Proceedings of the Royal Society B: Biological Sciences.

Arft, A., Walker, M., Gurevitch, J.et a., Alatalo, J., Bret-Harte, M., Dale, M., Diemer, M., Gugerli, F., Henry, G., and Jones, M. (1999). Responses of tundra plants to experimental warming: Meta-analysis of the international tundra experiment. Ecological monographs **69,** 491-511.

Armstrong, J.J., Takebayashi, N., Sformo, T., and Wolf, D.E. (2015). Cold tolerance in *Arabidopsis kamchatica*. American journal of botany **102,** 439-448.

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane, S.L., Warwick, S.I., Windham, M.D., and Al-Shehbaz, I.A. (2006). Toward a global phylogeny of the Brassicaceae. Molecular Biology and Evolution **23,** 2142-2160.

Bajer, A. (1951). Cytological studies on *Cochlearia tatrae* Borb. Bulletin International de l'Academie Polonaise des Sciences et des Lettres Série B **3,** 89-118.

Barrett, R.T., Hollister, R.D., Oberbauer, S.F., and Tweedie, C.E. (2015). Arctic plant responses to changing abiotic factors in northern Alaska. American journal of botany **102,** 2020-2031.

Beilstein, M.A., Al-Shehbaz, I.A., and Kellogg, E.A. (2006). Brassicaceae phylogeny and trichome evolution. American journal of botany **93,** 607-619.

Bokhorst, S.F., Bjerke, J.W., Tømmervik, H., Callaghan, T.V., and Phoenix, G.K. (2009). Winter warming events damage sub-Arctic vegetation: consistent evidence from an experimental manipulation and a natural event. Journal of Ecology **97,** 1408-1415.

Bolker, J. (2012). Model organisms: There's more to life than rats and flies. Nature **491,** 31-33.

Bowman, J. (2006). Molecules and morphology: comparative developmental genetics of the Brassicaceae. Plant Systematics and Evolution **259,** 199-215.

Brock, A., Herzfeld, T., Paschke, R., Koch, M., and Dräger, B. (2006). Brassicaceae contain nortropane alkaloids. Phytochemistry **67,** 2050-2057.

Brunetti, C., George, R.M., Tattini, M., Field, K., and Davey, M.P. (2013). Metabolomics in plant environmental physiology. Journal of experimental botany**,** ert244.

Ceccarelli, M., and Cionini, P. (1993). Tissue-specific nuclear repatterning in plant cells. Genome **36,** 1092-1098.

Chapin, F.S., Shaver, G.R., Giblin, A.E., Nadelhoffer, K.J., and Laundre, J.A. (1995). Responses of arctic tundra to experimental and observed changes in climate. Ecology **76,** 694-711.

Cieślak, E., Ronikier, M., and Koch, M.A. (2007). Western Ukrainian *Cochlearia* (Brassicaceae)—the identity of an isolated edge population. Taxon **56,** 112-118.

Cieślak, E., Kaźmierczakowa, R., and Ronikier, M. (2011). *Cochlearia polonica* Fröhl. (Brassicaceae), a narrow endemic species of southern Poland: history of conservation efforts, overview of current population resources and genetic structure of populations. Acta Societatis Botanicorum Poloniae **79,** 255-261.

Clapham, A.R., Tutin, T.G., and Moore, D.M. (1990). Flora of the British isles. (CUP Archive).

Crane, M.B., and Gairdner, A. (1923). Species-crosses in *Cochlearia*, with a preliminary account of their cytology. Journal of Genetics **13,** 187-200.

Dauvergne, X., Cérantola, S., Salaün, S., Magné, C., Kervarec, N., Bessières, M.-A., and Deslandes, E. (2006). General occurrence of the glucosinolate glucocochlearin within the *Cochlearia* genus. Carbohydrate research **341,** 2166-2169.

Davis, M.B., and Shaw, R.G. (2001). Range shifts and adaptive responses to Quaternary climate change. Science **292,** 673-679.

Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glöckner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., and Tang, M. (2015). The butterfly plant arms-race escalated by gene and genome duplications. Proceedings of the National Academy of Sciences **112,** 8362-8366.

Elkington, T. (1984). Cytogenetic variation in the British flora: Origins and significance. New Phytologist **98,** 101-118.

Faria, R., Renaut, S., Galindo, J., Pinho, C., Melo-Ferreira, J., Melo, M., Jones, F., Salzburger, W., Schluter, D., and Butlin, R. (2014). Advances in ecological speciation: an integrative approach. Molecular ecology **23,** 513-521.

Fearn, G. (1977). A morphological and cytological investigation of *Cochlearia* populations on the Gower peninsula, Glamorgan. New Phytologist **79,** 455-458.

Fiehn, O. (2002). Metabolomics–the link between genotypes and phenotypes. Plant molecular biology **48,** 155-171.

Franks, S.J., and Hoffmann, A.A. (2012). Genetics of climate change adaptation. Annu Rev Genet **46,** 185-208.

Franzke, A., German, D., Al-Shehbaz, I.A., and Mummenhoff, K. (2009). *Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae. Taxon **58,** 425-437.

Galloway, G.L., Malmberg, R.L., and Price, R.A. (1998). Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. Molecular Biology and Evolution **15,** 1312-1320.

Garcia-Castellanos, D., Estrada, F., Jiménez-Munt, I., Gorini, C., Fernández, M., Vergés, J., and De Vicente, R. (2009). Catastrophic flood of the Mediterranean after the Messinian salinity crisis. Nature **462,** 778-781.

German, D., Garden, S.-S.B., and Friesen, N. (2015). *Shehbazia* (Shehbazieae, Cruciferae), a new monotypic genus and tribe of hybrid origin from Tibet. Turczaninowia.

Gill, B., McAllister, H., and Fearn, G. (1978). Cytotaxonomic studies on the *Cochlearia officinalis* L. group from inland stations in Britain. Watsonia **8,** 395-396.

Gill, E. (2008). Conservation genetics of the species complex *Cochlearia officinalis* L. s.l. in Britain. PhD thesis. University of Edinburgh.

Gill, J. (1971a). Cytogenetic studies in *Cochlearia* L. The chromosomal homogeneity within both the 2n = 12 diploids and the 2n = 14 diploids and the cytogenetic relationship between the two chromosome levels. Annals of Botany **35,** 947-956.

Gill, J. (1971b). The cytology and transmission of accessory chromosomes in *Cochlearia pyrenaica* DC. (Cruciferae). Caryologia **24,** 173-181.

Gill, J. (1973). Cytogenetic studies in *Cochlearia* L. (Cruciferae). The origins of *C. officinalis* L. and *C. micacea* Marshall. Genetica **44,** 217-234.

Gill, J. (1976). Cytogenetic studies in *Cochlearia* L. (Cruciferae). The chromosomal constitution of *C. danica* L. Genetica **46,** 115-127.

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A., and Tegnér, J. (2014). Data integration in the era of omics: current and future challenges. BMC systems biology **8,** I1.

Hedges, S.B. (2002). The origin and evolution of model organisms. Nature Reviews Genetics **3,** 838-849.

Hohmann, N., Wolf, E.M., Lysak, M.A., and Koch, M.A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. The Plant Cell **27,** 2770-2784.

Hollister, R.D., Webber, P.J., and Tweedie, C.E. (2005). The response of Alaskan arctic tundra to experimental warming: Differences between short-and long-term responses. Global Change Biology **11,** 525-536.

Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., and Edger, P.P. (2015). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Molecular biology and evolution **33,** 394-412.

Hudson, J., Henry, G., and Cornwell, W. (2011). Taller and larger: shifts in Arctic tundra leaf traits after 16 years of experimental warming. Global Change Biology **17,** 1013-1021.

Jump, A.S., and Penuelas, J. (2005). Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters **8,** 1010-1020.

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., and Links, M.G. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. The Plant Cell **26,** 2777-2791.

Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., and Koch, M.A. (2013). BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Plant and Cell Physiology**,** pct158.

Koch, M. (2002). Genetic differentiation and speciation in prealpine *Cochlearia*: Allohexaploid *Cochlearia bavarica* Vogt (Brassicaceae) compared to its diploid ancestor *Cochlearia pyrenaica* DC. in Germany and Austria. Plant Systematics and Evolution **232,** 35-49.

Koch, M., and Mummenhoff, K. (2001). *Thlaspi* s. str. (Brassicaceae) versus *Thlaspi* s.l.: morphological and anatomical characters in the light of ITS nrDNA sequence data. Plant Systematics and Evolution **227,** 209-225.

Koch, M., and Al-Shehbaz, I.A. (2004). Taxonomic and phylogenetic evaluation of the American "*Thlaspi*" species: identity and relationship to the Eurasian genus *Noccaea* (Brassicaceae). Systematic Botany **29,** 375-384.

Koch, M., and Bernhardt, K. (2004). *Cochlearia macrorrhiza*, a highly endangered lowland species from Eastern Austria. Conservation genetics, ex situ and in situ conservation efforts. Scripta Bot Belg **29,** 157-164.

Koch, M., Hurka, H., and Mummenhoff, K. (1996). Chloroplast DNA restriction site variation and RAPD-analyses in *Cochlearia* (Brassicaceae): Biosystematics and speciation. Nordic Journal of Botany **16,** 585-603.

Koch, M., Huthmann, M., and Hurka, H. (1998). Isozymes, speciation and evolution in the polyploid complex *Cochlearia* L. (Brassicaceae). Botanica acta **111,** 411-425.

Koch, M., Mummenhoff, K., and Hurka, H. (1999). Molecular phylogenetics of *Cochlearia* (Brassicaceae) and allied genera based on nuclear ribosomal ITS DNA sequence analysis contradict traditional concepts of their evolutionary relationship. Plant systematics and evolution **216,** 207-230.

Koch, M., Dobeš, C., Bernhardt, K., and Kochjarová, J. (2003). *Cochlearia macrorrhiza* (Brassicaceae): A bridging species between Cochlearia taxa from the Eastern Alps and the Carpathians? Plant Systematics and Evolution **242,** 137-147.

Koch, M.A. (2012). Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). Taxon**,** 76-92.

Koch, M.A., Dobeš, C., Kiefer, C., Schmickl, R., Klimeš, L., and Lysak, M.A. (2007). Supernetwork identifies multiple events of plastid trnF (GAA) pseudogene evolution in the Brassicaceae. Molecular Biology and Evolution **24,** 63-73.

Kochjarová, J., Valachovič, M., Bureš, P., and MrÁZ, P. (2006). The genus *Cochlearia* L. (Brassicaceae) in the Eastern Carpathians and adjacent area. Botanical Journal of the Linnean Society **151,** 355-364.

Krijgsman, W. (2002). The Mediterranean: mare nostrum of earth sciences. Earth and Planetary Science Letters **205,** 1-12.

Krijgsman, W., Hilgen, F., Raffi, I., Sierro, F., and Wilson, D. (1999). Chronology, causes and progression of the Messinian salinity crisis. Nature **400,** 652-655.

Li, Y., and Chen, L. (2014). Big biological data: challenges and opportunities. Genomics, proteomics & bioinformatics **12,** 187-189.

Löve, Á., and Löve, D. (1961). Chromosome numbers of central and northwest European plant species. Opera Botanica (Lund) **5,** 1-581.

Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A., and Leitch, I.J. (2009). The dynamic ups and downs of genome size evolution in Brassicaceae. Molecular Biology and Evolution **26,** 85-98.

Mandakova, T., Li, Z., Barker, M.S., and Lysak, M.A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. The Plant Journal.

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H., and Burke, J.M. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. Applications in Plant Sciences **2,** 1300085.

Marhold, K., and Lihová, J. (2006). Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. Plant Systematics and Evolution **259,** 143-174.

Mrkvicka, A.C., Pfundner, G., Pfundner, P., and Sauberer, N. (2015). Zweimal ausgestorben – Die gescheiterte Wiederansiedlung des Dickwurzel-Löffelkrauts (*Cochlearia macrorrhiza*) im Naturdenkmal Brunnlust (Moosbrunn, Niederösterreich). BCBEA **1/2,** 252–261.

Mummenhoff, K., and Koch, M. (1994). Chloroplast DNA restriction site variation and phylogenetic relationships in the genus *Thlaspi* sensu lato (Brassicaceae). Systematic Botany**,** 73-88.

Mummenhoff, K., Franzke, A., and Koch, M. (1997a). Molecular phylogenetics of *Thlaspi* s.l. (Brassicaceae) based on chloroplast DNA restriction site variation and sequences of the internal transcribed spacers of nuclear ribosomal DNA. Canadian Journal of Botany **75,** 469-482.

Mummenhoff, K., Franzke, A., and Koch, M. (1997b). Molecular data reveal convergence in fruit characters used in the classification ofThlaspi sl (Brassicaceae). Botanical Journal of the Linnean Society **125,** 183-199.

Nadeau, N.J., and Jiggins, C.D. (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends in Genetics **26,** 484-492.

Nawaz, I., Iqbal, M., Bliek, M., and Schat, H. (2017). Salt and heavy metal tolerance and expression levels of candidate tolerance genes among four extremophile *Cochlearia* species with contrasting habitat preferences. Science of The Total Environment **584,** 731-741.

Nordal, I. (1988). *Cochlearia pyrenaica* DC., a species new to Scotland. Watsonia **17,** 49-52.

Nordal, I., and Stabbetorp, O. (1990). Morphology and taxonomy of the genus *Cochlearia* (Brassicaceae) in Northern Scandinavia. Nordic journal of botany **10,** 249-263.

Nordal, I., and Laane, M. (1990). Cytology and Reproduction in arctic *Cochlearia*. Sommerfeltia **11,** 147-158.

Nordal, I., Eriksen, A., Laane, M., and Solberg, Y. (1986). Biogeographic and biosystematic studies in the genus *Cochlearia* in Northern Scandinavia. Symbolae Botanicae Upsalienses **27,** 83-93.

Otto, S.P., and Whitton, J. (2000). Polyploid incidence and evolution. Annual review of genetics **34,** 401-437.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. Annual Review of Ecology, Evolution, and Systematics**,** 637-669.

Pegtel, D.M. (1999). Effect of ploidy level on fruit morphology, seed germination and juvenile growth in scurvy grass (*Cochlearia officinalis* L. s.l., Brassicaceae). Plant Species Biology **14,** 201-215.

Pobedimova, E. (1970). Revisio generis *Cochlearia* L., 2 Novosti sistematiky Vysshikh Rastenii. Leningrad **7,** 167-195.

Price, R.A., Palmer, J.D., and Al-Shehbaz, I.A. (1994). Systematic Relationships of *Arabidopsis*: A Molecular and Morphological Perspective. Cold Spring Harbor Monograph Archive **27,** 7-19.

Reeves, R.D. (1988). Nickel and zinc accumulation by species of *Thlaspi* L., *Cochlearia* L., and other genera of the Brassicaceae. Taxon**,** 309-318.

Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P., and Arroyo, J. (2008). The Strait of Gibraltar as a melting pot for plant biodiversity. Quaternary Science Reviews **27,** 2100-2117.

Sardans, J., Penuelas, J., and Rivas-Ubach, A. (2011). Ecological metabolomics: overview of current developments and future challenges. Chemoecology **21,** 191-225.

Saunte, L.H. (1955). Cytogenetical studies in the *Cochlearia officinalis* complex. Hereditas **41,** 499-515.

Savolainen, O., Lascoux, M., and Merila, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics **14,** 807-820.

Schranz, M.E., Lysak, M.A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends in Plant Science **11,** 535-542.

Soltis, D.E., and Soltis, P.S. (1999). Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution **14,** 348-352.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. American journal of botany **96,** 336-348.

Soltis, P.S., Liu, X., Marchant, D.B., Visger, C.J., and Soltis, D.E. (2014). Polyploidy and novelty: Gottlieb's legacy. Philosophical Transactions of the Royal Society B **369,** 20130351.

Stefureac, T., and Lungeanu, I. (1976). Considérations caryologiques sur l'éspèce relicte *Cochlearia pyrenaica* DC. de Roumanie. Rev. Roum. Biol., Biol. Veg **21,** 117-119.

Stefureac, T., and Panzaru, G. (1978). Statiunea cu *Cochlearia pyrenaica* DC de la Silhoi (Maramures) si ocrotirea sa (La station avec *Cochlearia pyrenaica* DC de Silhoi (Maramures) et sa sauvegarde.). Ocrot. Nat **22,** 39-42.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., and Harmon, L.J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytologist **207,** 454-467.

Warwick, S., and Al-Shehbaz, I. (2006). Brassicaceae: chromosome number index and database on CD-Rom. Plant Systematics and Evolution **259,** 237-248.

Warwick, S., Francis, A., and Al-Shehbaz, I. (2006). Brassicaceae: species checklist and database on CD-Rom. Plant Systematics and Evolution **259,** 249-258.

Whittall, J.B., Syring, J., Parks, M., Buenrostro, J., Dick, C., Liston, A., and Cronn, R. (2010). Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. Molecular Ecology **19,** 100-114.

Williams, S.E., Shoo, L.P., Isaac, J.L., Hoffmann, A.A., and Langham, G. (2008). Towards an integrated framework for assessing the vulnerability of species to climate change. PLoS Biology **6,** e325.

Zmudczyńska-Skarbek, K., Barcikowski, M., Zwolicki, A., Iliszko, L., and Stempniewicz, L. (2013). Variability of polar scurvygrass *Cochlearia groenlandica* individual traits along a seabird influenced gradient across Spitsbergen tundra. Polar Biology **36,** 1659-1669.

Chapter 1 – *Cochlearia* Cytogenetics

This chapter summarizes and extends cytogenetic studies presented by Rainer Schulz in his master thesis "Evolutionary insights from cytogenetic analyses: the case of *Cochlearia* L." (2015).

1.1 Abstract

In this chapter, the cytogenetic variation and evolution within the genus *Cochlearia* is analyzed via both conventional chromosome counting and flow cytometry. A correlation between genome size and ploidy level/chromosome number is revealed and genome downsizing in polyploid taxa is described. Moreover, in combination with own chromosome counts, a comprehensive literature review on *Cochlearia* chromosome numbers enables the analysis of genus-wide distributional patterns of cytotypic evolution with a special emphasis on the presence of aneuploidies which are rarely found in diploid taxa but frequently in polyploids. These chromosomal aberrations are interpreted as a common result of frequent interploidal hybridization reflecting the highly dynamic evolution of polyploid taxa within the genus. The clear geographical separation of the two diploid karyotypes (2n=12 and 2n=14) suggests an early evolutionary separation and renews the question of the ancestral base chromosome number.

For a more detailed reconstruction of the evolutionary history of the genus *Cochlearia*, especially regarding the cytogenetic evolution within in the genus, chromosome numbers and genome size measurements alone will not be sufficient, since those might originate via different or repeated processes (e.g. allo- and autopolyploidization, diploidization). In order to fully understand these complex past processes, a combination of cytogenetic analyses, in-depth phylogenetic studies (ideally based on whole-genome assemblies) and comparative cytogenetic strategies are needed. The phylogenetic aspect will be the subject of chapter 2.

1.2 Introduction

1.2.1 Genome evolution in flowering plants

Genome sizes of angiosperms (flowering plants) vary dramatically both between and within families, genera or even species (Bennett and Leitch, 2005), with the smallest and the largest known C-values (the amount of DNA in picograms (pg) in an unreplicated, haploid nucleus) showing a difference of about 2400-fold (Bennett and Leitch, 2005; Doležel et al., 2007a; Dodsworth et al., 2015). These magnitudes of variation have been studied intensively over the last decades both in large-scale studies, e.g. analyzing C-values throughout the land plants (Embryophyta, Leitch et al., 2005), or on smaller scales, for example within a family (e.g. the Brassicaceae family; Lysak et al., 2009), or a single genus/species (e.g. Hawkins et al., 2006; Zuccolo et al., 2007; Costich et al., 2010). Although many details on the regulation of genome size are still in the dark, several drivers of genome size changes have been identified, with the two most important ones for genome size increase being the amplification of repetitive genome fractions, the transposable elements (Hawkins et al., 2006) and the multiplication of chromosomes or whole genomes (polyploidy) leading to a stepwise gain in genome size (Hu et al., 2011; Dodsworth et al., 2015). Genome size increase via polyploidization is often followed by rapid reorganization and loss of redundant duplicated genomic blocks, ergo genome downsizing and diploidization (e.g. Wolfe, 2001; Mandáková et al., 2010; Liu et al., 2014; Dodsworth et al., 2016; Mandáková et al., 2016; Mandakova et al., 2017) as recently reviewed for the Brassicaceae family by Hohmann et al. (2015).

Polyploidization may either be the result of a genome duplication/triplication within a species (autopolyploidy) or arise via the hybridization of different taxa resulting in allopolyploidy. Due to the resulting difficulties in meiotic pairing of the different chromosome sets in a diploid hybrid, which may be avoided in polyploid genomes, hybridization frequently leads to the establishment of polyploids (Otto, 2007). Whilst moderately spotted in animals (e.g. platyhelminths, insects, fishes, amphibians), polyploidy frequently occurs in plants, especially in angiosperms (Leitch and Leitch, 2008), where a ploidy increase was estimated to be involved in about 15% of all speciation events (Wood et al., 2009). Over the course of the last years and especially with more and more fully sequenced plant genomes available, a wealth of new insights have been gained, showing for example that probably all flowering plants have undergone at least one polyploidization event (Soltis et al., 2004; Albert et al., 2013; Mayfield-Jones et al., 2013). The effects of these whole genome duplications (WGDs) are manifold (reviewed by Otto, 2007) and they concern all aspects of an organism's essential biological processes (e.g. Costich et al., 2010). With regard to the ecological implications of polyploidy, Thompson et al. (2004) described its effect on the evolutionary ecology of the interactions between plants and herbivores/pollinators. Also, as recently proposed by Guignard et al. (2016), polyploidy and variation in genome size might play a critical role in building up biomass under differences in the availability of the nutrients phosphorus and nitrogen, central components of nucleic acids. Thus, variation at the level of chromosome numbers and genome size seems to be of ecological relevance as well. However, the effects of

WGDs depend on the time since the occurrence of the duplication. Whereas newly obtained polyploidy (neopolyploidy) strongly affects general processes like the cell cycle and gene regulation, mesopolyploid genomes (at intermediate timescales after the duplication event) are often characterized by trends of fractionation and diploidization, sometimes leaving only slight marks, such as an over-retention of dosage-sensitive genes (reviewed e.g. by Edger and Pires, 2009), in the paleopolyploid genome with more time passing (Mayfield-Jones et al., 2013). Thus, the effects of the ancient duplication may still be present in gene copy number, gene expression and the downstream metabolomic responses, finally resulting in the organism's phenotype (Mayfield-Jones et al., 2013; Costich et al., 2010; Liu et al., 2014). Therefore, it is commonly accepted meanwhile that these evolutionary changes on the genome level played an important role in the evolution of plants and polyploidization has been suggested as one of the major drivers of diversification in plants (Soltis and Soltis, 1999; Otto and Whitton, 2000; Adams and Wendel, 2005; Soltis et al., 2009; Soltis et al., 2014; Tank et al., 2015). Therefore, interest in genome evolution and polyploidization has been and still is very high (e.g. Wendel, 2000; Soltis and Soltis, 1999; Leitch and Leitch, 2008) and by now, many insights into the general patterns of plant genome organization and evolution have been gained, triggered by new methods that arose for example with the availability of fully sequenced genomes and comparative genomic studies based hereon, and/or karyotype analyses based on comparative chromosome painting (e.g. Lysak et al., 2006; Mandáková et al., 2010; Liu et al., 2014; Lysak et al., 2016; Mandáková et al., 2016; reviewed e.g. by Schranz et al., 2006). However, the dynamics of these processes are not yet fully understood and further studies, especially in mesopolyploid genomes, might shed new light on this topic (Lysak et al., 2016).

Since major WGD events in different angiosperm lineages are chronologically linked to the mass Cretaceous–Tertiary extinction event ~65 my ago, it was hypothesized that polyploidy might have been beneficial for the adaptation to a changing environment (Fawcett et al., 2009; Vanneste et al., 2014; Hohmann et al., 2015; reviewed by Franzke et al., 2011). Therefore, deeper insights into genome evolution e.g. in the context of changing climatic conditions might be of special interest with regard to the ongoing global warming.

1.2.2 Genome evolution in the Brassicaceae family

Research on genome evolution within the Brassicaceae family has been shown to be highly interesting and enlightening for several reasons. As reviewed recently by Hohmann et al. (2015) the family shows the highest speciation rates reported for any plant group so far and diversification seems to have taken place at increased rates within lineages and tribes during the last 30 million years (Couvreur et al., 2010; Karl and Koch, 2013). As revealed from studies on the *Arabidopsis thaliana* genome, the evolutionary history of the Brassicaceae family is characterized by several rounds of WGDs and number and timing of these events have been studied intensively over the course of the last decades (e.g. Blanc et al., 2003; Bowers et al., 2003; Barker et al., 2009; Soltis et al., 2009; Schranz et al., 2012; Kagale et al., 2014; Magallón et al., 2015; Edger et al., 2015). Three major WGDs (known as At-γ, At-β, At-α) have been detected (Bowers et al., 2003; Barker et al., 2009; reviewed by Franzke et al., 2011). Whereas the oldest of these WGDs, the At-γ

duplication, most probably occurred after the monocot-eudicot divergence and likely in context with the diversification of the eudicots, the two younger events putatively occurred within Brassicales (De Bodt et al., 2005; Soltis et al., 2009). The At-β duplication was proposed to have taken place after the Caricaceae-Brassicaceae split (Ming et al., 2008) and the youngest of these major WGDs, the At-α duplication, likely occurred within core Brassicaceae (Vision et al., 2000; Bowers et al., 2003; Fawcett et al., 2009; Schranz et al., 2012). According to recent analyses based on Brassicales-wide transcriptome data, the At-β duplication could have occurred ~88 mya and the At-α duplication ~40 mya (Edger et al., 2015). However, divergence time estimates for the Brassicaceae family are problematic, mainly due to the poor fossil record, and thus have been discussed intensively since they strongly depend on the different approaches that were applied (see e.g. Kagale et al., 2014; Magallón et al., 2015; Hohmann et al., 2015; reviewed by Franzke et al., 2016).

Several tribe-specific post-At-α mesopolyploidization events have been described (e.g. Mandáková et al., 2010; Mandáková et al., 2012; Cheng et al., 2013; Geiser et al., 2015; Mandakova et al., 2017; reviewed by Hohmann et al., 2015) and within recent Brassicaceae, about 37% of species have been detected to be of a putative (neo- or meso-)polyploid origin under a polyploid definition of $n \geq 14$ (Appel and Al-Shehbaz, 2002; Warwick et al., 2006; Schranz and Mitchell-Olds, 2006). Thus, besides the three major paleopolyploidy events, polyploidization is still acting as an important factor in Brassicaceae evolution and there is large interest in additional cytogenetic insights. Based on extensive cytogenetic and comparative genomic studies within Brassicaceae, an Ancestral Crucifer Karyotype (ACK) has been proposed by Schranz et al. (2006) and recently updated to consist of 22 genomic blocks organized in 8 chromosomes (Lysak et al., 2016). Besides this ACK, the so called Proto-Calepineae Karyotype (PCK), another highly abundant ancestral Brassicaceae genome, made up of seven chromosomes, has been reconstructed (Mandáková and Lysak, 2008). The reconstruction of these ancestral karyotypes provided valuable insights into general patterns of Brassicaceae genome organization and evolution. According to Lysak et al. (2016), genome evolution within Brassicaceae has been described to be either a) very stable over long evolutionary timescales, with no changes in the ancestral karyotypes, b) stable with regard to chromosome numbers but showing strong reshuffling within, or c) instable in chromosome numbers and showing a decrease from the ACK/PCK (e.g. from n=8 to n=5 in *A. thaliana*). Aside from these new cytogenetic insights, the description of the ACK further facilitated the analyses of Brassicaceae whole genome sequencing data thus promoting the field of phylogenomics for this family (Lysak et al., 2016).

As for the whole Angiosperms, variation in genome size is comparatively high within the Brassicaceae family, yet in general, Brassicaceae genomes are comparatively small (Lysak et al., 2009). As shown recently by Hohmann et al. (2015), there is a general trend of genome size reduction towards or even below an estimated ancestral genome size of ~0.5 pg (Lysak et al., 2009). Thus (neo-)polyploidization followed by the stabilizing effects of genome downsizing and diploidization were suggested to be of major relevance for the family-wide diversification processes.

1.2.3 Genome evolution in the genus *Cochlearia*

With regard to the immense effects of genome-scale changes on an organism's basic biological functions as well as on the putative evolutionary consequences, it is an important prerequisite for studies on phylogenomics (chapter 2) or metabolomics (chapter 3) to gain thorough insights into the genome evolution of the study system (Costich et al., 2010). Besides interesting aspects focusing on general mechanisms/patterns of genome evolution, cytogenetic analyses also fulfil very practical purposes with regard to correct identification and distinction of taxa (Elkington, 1984). As stated above, this can be especially useful in a genus like *Cochlearia*, where morphological differences between taxa are poorly developed. Since the genomic constitutions of the different *Cochlearia* species seem to be the result of a very dynamic, yet quite recent past, involving several (possibly recurring) events of hybridization and polyploidization (Koch et al., 1998; Koch et al., 1999; Koch, 2012) it is of large interest to know the karyotypes of the analyzed populations regarding taxonomy, phylogeny, genome evolution and ecological implications. Owing to the highly complex cytotaxonomic situation within *Cochlearia*, a range of studies has been carried out early on, in order to unravel cytogenetic characteristics and relationships within the genus (e.g. Crane and Gairdner, 1923; Saunte, 1955; Gill, 1965, 1971a, 1973, 1976; Gill et al., 1978). Till now, many insights on the various karyotypes in different species have been gained, yet a genus-wide overview of the measured chromosome numbers throughout the distribution area has been missing so far.

From a cytogenetic point of view, (ancient) polyploidization and hybridization are clearly the most important processes in the complex evolution of the genus. Using both comparative chromosome painting and/or whole-transcriptome data, Mandakova et al. (2017) recently investigated the origins of multiple base chromosome numbers in different Brassicaceae genera and tribes including *Cochlearia*. Their results indicate two rounds of whole genome duplication for the tribe Cochlearieae, represented by *C. pyrenaica* and *C. officinalis*, which is therefore regarded as a mesohexaploid tribe. This mesohexaploidy had not been revealed before in a study of Kagale et al. (2014) based on *K*s analysis (Blanc and Wolfe, 2004) and likewise including *C. officinalis*. Today we see a large range of naturally of occurring ploidy levels (diploids up to octoploids or even dodecaploids) within the genus *Cochlearia*. Some of them are considered autopolyploids; most of them might actually be of hybrid origin (allopolyploids) as a result of very low interspecific barriers to gene flow (e.g. Koch et al., 1998). Hybrids between different *Cochlearia* taxa have been described early on both in natural habitats as well as in artificial crossings (e.g. Crane and Gairdner, 1923; Gill, 1975; Fearn, 1977). Among the known naturally occurring hybrids are several coastal hybrid cytotypes within the *C. officinalis* s.l. complex such as *C. danica* x *C. officinalis* (2n=33), *C.* x *hollandica* (2n=36; likely *C. officinalis* x *C. anglica*) with different cytotypes of parental taxa and hybrids showing differences in their ecological preferences - especially regarding salt tolerance (Fearn, 1977; Pegtel, 1999). Regarding this highly dynamic cytogenetic evolution, the genus *Cochlearia* could well serve as a model system for ancient, recent and recurrent formation of both auto- and allopolyploidy.

Finally, besides the above mentioned insights into general trends of genome organization and evolution, the flow-cytometric measurement of genome sizes within the organisms of interest

enables a rough estimation of the expected sequencing depth in next generation sequencing approaches, this way facilitating planning and preparation of NGS projects as well as downstream data analyses.

1.3 Material and methods

Most of the practical and analytical work in this chapter was done by Rainer Schulz within the scope of his master thesis "Evolutionary insights from cytogenetic analyses: the case of *Cochlearia* L." (2015). Flow-cytometric measurements were continued after this work and updated in here together with respective statistical analyses.

1.3.1 Cytogenetic methods

In order to complement the list of previously published chromosome counts and genome sizes for the genus *Cochlearia*, and also to obtain the cytogenetic information on (most of) the populations that were chosen for phylogenetic analyses (see chapter 2), microscopical chromosome countings from root tips and flower buds as well as flow cytometry measurements of genome sizes from fresh leaf material were performed.

Plant material

The plants used in this study where either collected in the wild or grown from seed material under greenhouse conditions in the Botanical Garden Heidelberg (HEID accession numbers are given in Table 1-3). Both wild collections and grown plants were cultivated in the greenhouse in a substrate consisting of seedling potting soil, quartz sand, and either composted earth or Ökohum's peat based substrate (Ökohum GmbH, Herbertingen, Germany). In order to avoid high temperatures during the summer period, parts of the collection were temporarily moved to a plant room with controlled conditions of 20°C and a 16/8h day/night light cycle.

Flow cytometry

Flow cytometry measurements within the genus *Cochlearia* turned out to be very challenging, especially for the diploid taxa with comparatively small genome sizes, but also for other ploidy levels and due to high levels of endopolyploidy and high concentrations of certain cytosolic compounds, putatively tannins, that have been shown to cause large amounts of debris in other plant taxa before (e.g. Loureiro et al., 2006; Doležel et al., 2007b). Therefore, several steps of method improvement have been worked out on the basis of Doležel et al. (2007a; for a detailed best practice report refer to Schulz (2015)), with the most important aspects being the following: 1) optimization of cytometer instrument performance (periodical maintenance is recommended) 2) selection of high quality tissue material (free of parasites (e.g. fungi) and other contaminants; young tissues decreasing the amount of endopolyploid nuclei) 3) determination of the best performing nuclear isolation buffer and staining protocol 4) proper adjustment of the amount of sample tissue (strongly depending on species and age of tissue material) 5) optimal balancing of the amount of sample vs. standard material.

For *Cochlearia*, the following sample preparation steps gave the best results. Very young, high quality leaf material was chosen for flow cytometric analyses. On the one hand, this improved the

chances of cutting leaves before they got infected by fungi or other contaminating factors leading to additional peaks or larger amounts of debris background. On the other hand, it decreased the amount of cells showing endopolyploidy. In Brassicaceae, and especially for species with small genome sizes, endopolyploidy is a common phenomenon (e.g. Barow and Meister, 2003; Barow, 2006) which has also been described for *Cochlearia* before (Kochjarová et al., 2006; Brandrud, 2014). Therefore, flow cytometry measurements of older *Cochlearia* leaves were often hindered by very strong 2C (but also 4C and 8C) peaks. Using very young leaves reduced this effect to a certain degree. Sample preparation steps were performed on ice. Preferably, Partec's two-step kit CyStain PI Absolute P (Partec GmbH, Münster, Germany) was used for nuclei extraction and staining, with the slight modifications of adding 15 mM β-mercaptoethanol and 1% w/v polyvinylpyrrolidone 25 (PVP) to the staining buffer. Leaf material of the sample and the chosen internal standard was chopped using a fresh razor blade in a petri dish containing 500 µl of the extraction buffer. A list of standard plants used in here, together with their respective 2C-values is given in Table I-1. Due to some slightly deviating 2C-values of standard sizes, all used standard plants were finally readjusted according to *Solanum* which was considered as a reference value. The chopped suspension was then filtered through a 50 µm CellTrics filter (Sysmex Partec GmbH, Görlitz, Germany) and incubated with 2000 µl of the staining buffer on ice for at least 30 to 60 minutes. Alternatively, lysis buffer LB01 (15 mM Tris, 2 mM Na2EDTA, 0.5 mM spermine tetrahydrochloride, 80 mM KCl, 20 mM NaCl, 0.1% v/v Triton X-100; 15 mM β-mercaptoethanol and 0.1 mg/ml RNAse A added right before preparation) as specified by (Doležel et al., 1989) was used instead of the Partec kit, providing similar peak qualities. Samples were analyzed in a Partec CyFlow Space flow cytometer (Sysmex Partec GmbH) using a 30 mW green solid state laser (λ=532 nm). The Partec FloMax software version 2.4 was used for gating and peak analysis.
Flow cytometry measurements were carried out by R. Schulz and E. Wolf (as documented in Supplemental Data Set I-1, sheet 1 (column A) and sheet 2 (column C)).

Table I-1 Standard plants used for flow cytrometric analyses with respective 2C-values as given in Doležel et al. (2007a). Genome sizes for the unknown cultivar of *Glycine max* and *Atropa belladonna* were measured secondarily by R. Schulz (as documented in Supplemental Data Set I-1, sheet 3 'standard plants').

Standard plant	**2C-value**
Raphanus sativus L. cv. "Saxa" (radish)	1.11 pg
Solanum lycopersicum L. cv. "Stupické polní rané" (tomato)	1.96 pg
Glycine max (L.) MERR. cv. "Polanka" (soybean)	2.50 pg
Glycine max (L.) MERR. (cultivar unknown)	2.46 pg
Atropa belladonna L. (deadly nightshade, Botanical Garden Heidelberg)	3.95 pg

Chromosome counting

Chromosome counts were generated on root-tip cells for the majority of analyzed accessions. Therefore, actively growing root-tips were collected in the early morning, pre-treated for at least 2 hours in 8-hydroxyquinoline at 3 °C and hereafter fixed in Carnoy's solution II (75% ethanol, 25% acetic acid). The fixative was changed for 1-2 times after about 2 h and finally kept at 3 °C overnight.

After washing the root tips 3 times for 5 minutes in distilled water and hereafter in citrate buffer (10 mM: 0.294% trisodium citrate dihydrate, 0.21% citric acid monohydrate; stored at 3 °C), they were incubated for at least 2 h at 37 °C in citrate buffer combined with an enzyme mix (1% cellulase, 1% pectolyase, 1% cyclohelicase; together with citrate buffer stored at -20 °C) for mazeration of the root tissue. The tissue mazeration was completed in a drop of 42% acetic acid on a clean microscopic glass slide, placed on a heating plate of 42 °C. The preparations were then first surrounded by Carnoy's solution II for 10 seconds, flooded with the same mixture and finally air-dried on the heating plate. Vector's VECTASHIELD® HardSetTM Mounting Medium with DAPI was used for mounting and staining of the slides. These were stored in the dark under refrigeration until further analysis.
Chromosome counts were examined using a Zeiss Axioskop (Zeiss Filter Set 43: BP550/25 (HE), Beam Splitter FT 570 (HE), BP 605/70 (HE)) at a 1000 x magnification. Pictures of the analyzed cells were taken using the Axiovision Rel. 4.8.2.0 software and either an AxioCam MRc5 (color) or an AxioCam HRm3 (monochrome). Further processing steps for quality improvement of (most of) the generated pictures involved 1) a blue channel extraction for the colored pictures using the software IrfanView v.4.38 2) a greyscale value readjustment in GIMP v.2.8.14 (GNU General Public License) and 3) a greyscale inversion.
Besides the chromosome counts gained from root tip material, for several accessions (Table I-3 and Supplemental Data Set I-2, sheet 4) chromosome counts and analyses were made on flower buds. Therefore, whole inflorescences or single flower buds from plants cultivated under greenhouse conditions were harvested right before the opening of the buds and fixed overnight in 6:3:1 ethanol : chloroform : acetic acid. Bud samples were then stored in 70% ethanol at −20°C until use. Further sample preparation (as described e.g. in Lysak and Mandáková, 2013) analysis was performed by Dr. Martin Lysak and Dr. Terezie Mandáková at the Central European Institute of Technology (CEITEC, Brno).

Statistical methods

The statistical analyses were carried out using R version 3.3.1 (R Core Team, 2014). Simple linear regression analyses as well as rank correlation tests were performed based on measured C-values and chromosome numbers (input data given with Supplemental Data Set I-1, sheet 4 'R input') in order to determine 1) the relationship between 2C values and chromosome numbers as well as 2) the relationship between DNA content per chromosome and the respective chromosome numbers in different ploidy levels. In both cases, chromosome numbers determined for other individuals of the same population were assigned to measured genome size values if no counts for the same individual were generated, that way accepting the possibility of errors, due to intrapopulational karyotype variability. Respective scatter plots for linear regressions were produced via the plot() function from the R 'graphics' package and the lm() function as embedded in the R 'stats' package was used to fit a linear model. Summaries of the fitted models, including significance tests, were generated and evaluated with the summary() function, and the respective quantile-quantile plots and other diagnostic plots for the linear regression models were inspected using the plot () function in order to assure the fit of a linear regression for the given dataset. Furthermore, both Spearman rank correlation and Kendall rank correlation tests were performed for the full dataset via the R

cor.test() function. Since normality cannot be assumed for this dataset, the chosen correlations were favored over Pearson linear correlations.

1.3.2 Geographical distribution of chromosome numbers/genome sizes

In addition to own measurements, published literature from the last 100 years has been reviewed for information on measured chromosome numbers and/or genome sizes in *Cochlearia* (see Schulz, 2015). A list of the literature included in the survey is shown in Table I-2.

Table I-2 Publications considered for literature review.

Author(s)	Year	Included therein
Craine, Gairdner	1923	
Saunte	1955	Matsuura and Sutô (1935), Böcher (1938), Maude (1939), Flovik (1940), Löve and Löve (1948), Bajer (1950), Böcher and Larsen (1950), Holmen (1952)
Gill	1965	
Gill	1971a 1971b	Zhukova (1965), Mosquin and Hayley (1966), Hedberg (1966), Zhukova (1967)
Gill	1973	
Gill	1976	
Gill et al.	1978	
Fearn	1977	
Fernández-Casas	1975	
Elkington	1984	
Vogt	1985	
Nordal et al.	1987	
Vogt	1987	Ludwig (1960), Kakes (1973)
Nordal and Lane	1990	
Heubl	1996	
Koch et al.	1996	
Nordal and Lane	1996	Lövkvist (1963a), Lövkvist (1963b)
Krisai and Greilhuber	1997	
Koch et al.	1998	
Pegtel	1999	
Peer et al.	2003	
Kochjarová et al.	2006	
Lysák et al.	2009	
BrassiBase (Kiefer et al.)	2014	

Information of own results and the literature survey was merged in a comprehensive database by R. Schulz, given with Supplemental Data Set I-3. If coordinates of the respective sampling points were given in the literature, they were transferred to the data table as well. Otherwise, coordinates were extracted as precisely as possible given the authors' information on the respective population localities. Data points with too little or too imprecise sampling point descriptions (e.g. "Arctic Russia") were finally excluded from the georeferenced dataset.

In total, more than 600 database entries were collected, treating every population given by one author as an individual data point. Further information on the database entries with respective coordinates (original or approximated), chromosome counts, the number of accessory chromosomes (if present), genome size measures, et cetera can be found in Supplemental Data Set I-3.

In order to visualize the geographical chromosome number distribution within the genus *Cochlearia*, maps were created by R. Schulz based on the information given in the database (Supplemental Data Set I-3, sheet 2 'ArcGIS') and using the software ESRI® ArcMapTM 10.

1.4 Results

1.4.1 Chromosome counting and flow cytrometric analysis

In total, chromosome counts (generated from root tips or flower buds) and/or genome size measurements of 133 individuals out of 60 populations/accessions were finally considered to be of sufficient quality for further analyses (see Table I-3; more detailed information on the produced results are given with Supplemental Data Sets I-1 and I-2).

Chromosome counting

Chromosome numbers were determined for 71 individuals out of 54 accessions (Table I-3 and Supplemental Data Set I-2) with the smallest measured chromosome number being 12 (diploid samples with 2n=12) and the largest one being 50 in octoploid *C. borzaeana* (regular species-specific karyotype 2n=48). If several individuals per accession were analyzed, chromosome numbers were often consistent but there are several cases where detected aneuploidies led to a slight variation within accessions (population numbers 144, 153, 158, 204, 207). None of these cases was found in diploid accessions but this might also be due to the generally small number of individuals analyzed per accession. Detected chromosome numbers were taken into account for taxon identification in several cases, especially for some indistinct coastal populations (see Table I-3). As mentioned earlier, this information is often needed because of the weak morphological differentiation within the genus. For example, none of the sampled British populations that were assumed to be *C. anglica* based on morphological (especially leaf) characters, had the species-specific karyotype of 2n=48. Instead, chromosome numbers ranging from 2n=24 to 2n=39 were detected (see Table I-3). On the contrary, some putative populations of *C. officinalis* had chromosome counts of 2n=36 and 2n=39 instead of 2n=24, which is indicative of a hybrid origin as well.

Table I-3 Results of chromosome counting and/or flow cytometry measurements, with information on analyzed populations/individuals. Asterisks (*) indicate progeny from open pollination. Table generated by R. Schulz taken from Schulz (2015), edited and complemented with additional data points. Chromosome counts marked with 'FB' were made on flower buds, generated by Dr. Martin Lysak and Dr. Terezie Mandáková at the Central European Institute of Technology (CEITEC, Brno).

presumed species	Nat.	Provenance	Pop.	Ind.	2n	2C
C. alpina (2n = 4x = 24)	UK	England, Somerset, Cheddar Gorge near Cheddar +51.283333, -2.75 (138 m)	010	HEID112859*	-	1.72 (1)
				HEID112863*	27	1.60 (2)
				HEID112864*	-	1.71 (1)
				HEID112986*	-	1.75 (2)
		Scotland,Teesdale +54.737165, -2.351277 (621 m) → *C. pyrenaica*	148	HEID113039	-	0.83 (2)
				HEID113042	-	0.86 (1)
				Calp_1320	12	-
		Scotland, Cairngorm Mountains; Corrie an't snechda +57.106775, -3.659078 (1013-1134 m)	153	HEID113036	29	-
				HEID112910	29	-
				Calp_0828	29-30	-
				Calp_1377	-	1.70 (1)
C. pyrenaica (2n = 2x = 12) or	UK	England, Ribblesdale (Ingleborough) +54.17, -2.34 [extr.]	200	HEID112876	12	0.75 (1)
				HEID112872	-	0.82 (1)
				HEID112874	-	0.83 (1)

C. alpina (2n = 4x = 24)		→ *C. pyrenaica*		HEID112877	-	0.81 (2)
				HEID112878	-	0.83 (3)
				HEID112880	-	0.80 (2)
		Scotland, Hartfell Rig near Moffat +55.408057, -3.401385 (714 m) [extr.]	203	HEID112850	-	1.69 (1)
				HEID112987	24	1.52 (3)
				HEID112849	-	1.53 (1)
		→ *C. alpina*		HEID112410	24	-
				Calp_1352	-	1.53 (1)
C. anglica (2n = 6x = 48)	UK	Scotland, north of Creetown +54.893354, -4.384783	206	HEID112727	-	2.20 (3)
				HEID112416	36 (FB)	-
		→ *C.* x *hollandica*?				
		Scotland, Sandyhills, south of Dumfries +54.877616, -3.729505 (6 m)	207	HEID112826	39	2.24 (3)
				HEID112828	39	-
				HEID112831	39 (38)	-
		→ *C.* x *hollandica*				
		Scotland, Cromarthy Firth, Ross. +57.616667, -4.383333	266	HEID112977	24	-
				HEID112978	-	1.56 (1)
		→ *C. officinalis*				
C. bavarica (2n = 6x = 36)	DE	Bayern, Glonn +47.991867, 11.85681	193	HEID112979	38	-
		Mariengrotte near Unteregg (Katzbrui) +47.971317, +10.454743 (683 m)	267	HEID112734	36	-
				HEID113056	-	2.28 (1)
				HEID113058	-	2.26 (1)
C. borzeana (2n = 8x = 48)	RO	Eastern Carpathians, Valea Salhoi (Distr. Maramures) +47.65, +24.983 [extr.]	139	HEID112995	-	2.61 (3)
				HEID112998	50	-
		Muntii (Distr. Maramures) +47.9, +24.5 [extr.]	140	HEID106723-5	50	3.09 (3)
		Valea Salhoi, Distr. Maramures +47.65, +24.983 [extr.]	270	HEID109736-1	-	2.72 (3)
C. danica (2n = 6x = 42)	FR	Barbâtre en Noirmoutier (Dép. Vendée) +46,910833, -2,163756	178	Cdan_1314	42	1.60 (1)
				Cdan_1315	-	1.59 (1)
				Cdan_1427	42	-
	ES	Carnota (Galicia), Stol Sea Farm in Lira +42.798818, -9.150195	281	HEID113102	41?	1.50 (2)
				HEID113107	-	1.48 (2)
				HEID113108	42	1.51 (2)
			296	HEID113110	42	1.50 (2)
				HEID113111	42	1.49 (2)
			297	HEID113112	42	1.50 (3)
				HEID113115	42	1.50 (2)
C. excelsa (2n = 2x = 12)	AT	Styria, Seckau Tauern, Mt. Seckauer Zinken +47.34006, +14.73758 (2299 m)	274	HEID125644	-	0.87 (2)
				HEID125645	-	0.86 (2)
		Styria, Gurktal Alps, Mt. Eisenhut, Dieslingsee +46.95134, +13.93578 (2019 m)	273	HEID125646		0.78 (2)
				HEID125647		0.77 (2)
C. islandica (2n = 2x = 12)	IS	Stokkseyri (South Coast) +63.816667, -21.033333 (0 m)	210	HEID125633	-	0.65 (2)
				HEID125638	-	0.65 (2)
				HEID125643	12 (FB)	0.69 (1)
				HEID113122	-	0.77 (3)
				HEID113126	-	0.77 (2)
				HEID115632	12	-
C. macrorrhiza (2n = 2x = 12)	AT	Niederösterreich, Moosbrunn +48.0 -16.43333 (ca. 200 m)	021	HEID109697-2	12	-
				HEID106724-2	-	0.74 (2)
				HEID109699-1	-	0.75 (2)
C. micacea (2n = 4x = 26)	UK	Scotland, Beinn an Dothaidh +56.533072, -4.724218 (743 m)	162	HEID112649	24	-
				HEID112832	24	-
				HEID112892	-	1.55 (1)
				HEID112894	-	1.54 (2)
				HEID112895	25?	1.55 (1)
		Scotland, Meall nan Tarmachan, (Ben Lawers region) +56.53788, -4.291608 (590 m)	163	HEID112725	-	1.52 (3)
				HEID112901	27	1.72 (2)
				HEID112903	-	1.69 (1)

				HEID112904	-	1.73 (2)
				HEID112905	27	-
C. officinalis (2n = 4x = 24)	NO	Storfjord, Skibotn (Trs) +69.366667, +20.216667 (0,5 m)	169	HEID112964	≥25	-
				HEID112965	-	1.55 (3)
	UK	England, North Somerset, Isle Brean Down west of Weston-Super-Mare +51.3281173, -3.0344457 [extr.]	144	HEID112835	36/37	-
				HEID112833	36	-
		→ *C.* x *hollandica*?				
		Scotland, northern Littleferry +56.186968, -2.812142 (5 m)	150	HEID112846	-	1.61 (3)
				HEID112847	25	-
		Scotland, Anstruther +56.24119, -2.651395 (1 m)	151	HEID112817	-	1.50 (4)
				HEID112820	25	-
				HEID112824	-	1.48 (3)
		Scotland, Dunure (coast south of Ayr) +55.40545, -4.761227 (4 m)	204	HEID112413	26 (FB)	-
				HEID112772	-	1.45 (4)
				HEID112773	24	-
		Scotland, Girvan (coast south of Ayr) +55.22599, -4.861633	205	HEID112788	39	2.29 (1)
		→ *C.* x *hollandica*				
		England, Northumberland, Holy Island +55.67876, -1.87443 (1.5 m)	292	HEID006209	-	1.61 (2)
		Scotland, east coast, E Lothian, Dunbar +56.00443, -2.51911 (2 m)	293	HEID006210	-	1.60 (2)
C. pyrenaica (2n = 2x = 12)	AT	Niederösterreich; Retzbach valley near Türnitz (Mostviertel) +47.889667, +15.467667 (525 m)	024	HEID112736	15	-
		Niederösterreich; Molterboden (Mostviertel) between Annaberg and Ulreichsberg +47.855, +15.387333 (894 m)	026	HEID113013	12	-
		Steiermark; Dobrein (Bez. Mürzzuschlag) +47.671983, +15.4762 (819 m)	028	HEID113032	12	-
		Steiermark; Siebenquellen (Bez. Mürzzuschlag) +47.683333, +15.558333 (775 m)	029	HEID113024	12	-
		Niederösterreich; Lehen near Lunz (Bez. Scheibbs); +47.836167, +15.017 (581 m)	032	HEID112750	12	-
		Niederösterreich, between Annaberg und Ulreichsberg (Mostviertel) +47.8505, +15.387167 (923 m)	268	HEID112740	12	-
	DE	Baden-Württemberg, near Dörzbach (Hohenlohekreis) +49.370556, +9.715278 (250 m)	040	HEID106720	12	-
		Baden-Württemberg, Mauchenmühle near Mühlhausen / Eberhardzell +47.958889, +9.827778 (583 m)	213	HEID112704	12	-
		Baden-Württemberg, Ochsenhausen +48,05083, +9,9235 (610 m)	214	HEID112694	12	-
				HEID112698	12	-
	ES	Asturias, between Villar de Vildas and La Pornacal +43.071833, -6.333333 (1000 m)	212	HEID112715	12	-
				HEID113047	-	0.74 (1)
				HEID113053	-	0.70 (1)
	FR	Auvergne (Dép. Cantal), 10 km east of Salers +45.127717, +26.11533 (1526 m)	001	HEID112955	-	0.78 (1)
				HEID112956	-	0.78 (1)
				HEID112957	-	0.80 (1)
				HEID112958	12	-
				HEID112959	-	0.80 (2)
				HEID112962	-	0.81 (2)
	SK	Velka Fatra, 1,5km southwest of Rosenberg/Jazierce +49.017167, +19.275 (589 m)	033	HEID113006	12	-
	UK	England, Malham +54.1, -2.175 [extr.]	199	HEID112868	13	0.81 (1)
				HEID112870	-	0.87 (1)

		England, Nenthead near Alston +54.785846, -2.341452 (439-465 m)	201	HEID112983	14	-
		Scotland, Yorkshire Dales, Greenhow, ~ 12km east of Grassington +54.073593, -1.818443 (348-380 m)	208	HEID112785	15	-
C. scotica (2n = 4x = 24)	UK	Scotland, Culbin Sands +57.59621, -3.83835 (1 m)	155	HEID112663	25	-
		Scotland, Burghead +57.704068, -3.495818 (5 m)	156	HEID112803	28	-
		Scotland, eastern Tairn, Portmahomack; +57.704068, -3.495818 (1 m)	157	HEID112945	32	1.81 (1)
				HEID112951	-	1.66 (2)
		Scotland, Brora +58.010304, -3.847068 (2 m)	158	HEID112931	25	-
				HEID112932	24	1.46 (1)
				HEID112934	-	1.45 (3)
		Scotland, Highland, Sutherland: Achmelvich +58.167536, -5.310338	159	HEID112769	-	1.45 (1)
				HEID112654	28 (FB)	-
		Scotland, Ullapool +57.901323, -5.168667 (1 m)	160	HEID112928	28	1.57 (3)
				Cscot_1332	-	1.45 (1)
		Scotland, Ballachulish +56.687867, -5.184737 (2 m)	161	HEID112937	-	1.48 (3)
				HEID112940	24	-
				HEID112656	24	-
C. sessilifolia (2n = 2x = 14)	Ala	Kodiak Quad.: Kodiak Island, Anton Larson Bay +57.845740, -152.627175 [extr.]	221	Cses_1294	-	0.76 (3)
		Kodiak Quad.: Kodiak Island, Buskin River mouth and lagoon +57.754783, -152.482748 [extr.]	223b	Cses_1295	-	0.75 (3)
				Cses_b	14 (FB)	-
		Kodiak Island, Buskin River mouth and lagoon +57.75529, -152.48523	222c	Cses_1807		0.72 (1)
				Cses_1808		0.72 (1)
C. spec (2n = ?)	UK	Scotland, harbour of Anstruther +56.2212, -2.6985 [extr.] → *C. danica*	152	HEID112809	-	1.48 (2)
				HEID112813	42	-
C. tatrae (2n = 6x = 42)	PL	Western Carpathians, Tatra Mts., Mieguszowiecka Przelecz pod Chlopkiem +49.18361, +20.065 [extr.]	240	HEID112722	38	-
	SL	Western Carpathians, Tatra Mts., Velicka Dolina +49,15, 20,15 [extr.]	145	HEID112992	41	2,10 (3)
		Western Carpathians, Tatra Mts., Velicka Dolina +49,15, 20,15	167	HEID108776	45	2,42 (3)

Geographical distribution of ploidy levels

The analysis of the geographical distribution of ploidy levels in *Cochlearia*, based on published and own cytogenetic analyses (more than 600 georeferenced data points as given with Supplemental Data Set I-3; Figure I-1), revealed some marked trends. First, there is a clear separation of arctic vs. Central European karyotypes in the diploid taxa – 2n=14 vs. 2n=12, or base chromosome numbers of n=7 vs. n=6 respectively (Figure I-2). Gairdner described a count of 2n=14 for the United Kingdom (in Maude, 1939), which would likely represent the southernmost documented occurrence of *C. groenlandica*, but this count could not be confirmed hereafter and was doubted by other authors (Elkington, 1984). Thus, the only contact zone is found in Iceland where the arctic karyotype (2n=14) is found in the north and the European one (2n=12) in the south.

Interestingly, when analyzing chromosome pictures, the southern Icelandic 2n=12 karyotype was initially suspected to carry two B-chromosomes, but based on more precise chromosome pictures

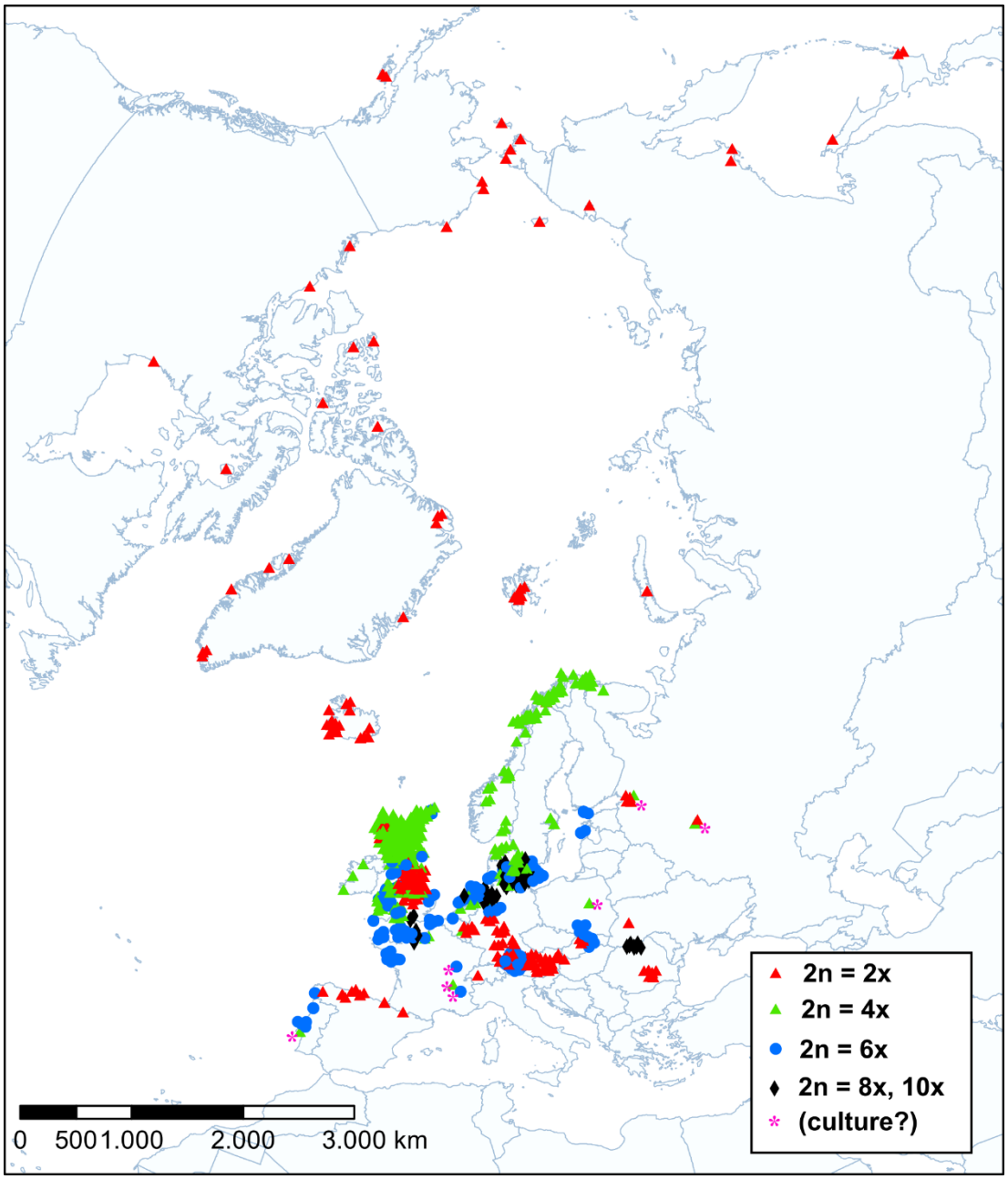

Figure I-1. Distribution of ploidy levels throughout the distribution range of *Cochlearia* according to own and published chromosome counts (Supplemental Data Set I-3). Aberrant chromosome numbers were grouped within the next lower ploidy level with few exceptions (if chromosome numbers were only slightly smaller than the expected number). Data points from the same or neighboring locations are shown slightly dispersed for a better visualization. Asterisks mark samples presumably originating from cultivation. Figure made by R. Schulz (taken from Schulz, 2015).

generated in collaboration with Dr. Martin A. Lysak and Dr. Terezie Mandáková (CEITEC, Brno) from flowering buds (harvested and fixed by E. Wolf from samples cultivated in Heidelberg), these were shown to be rather fragile heterochromatic knobs and/or nucleolar organizing regions (NORs) (see Figures I-3 and I-7). During sample preparation, these sites might likely break, leading to

artificially higher chromosome numbers. This could also be one of the explanations for the frequently described occurrences of B-chromosomes in the genus (see e.g. Saunte, 1955; Gill, 1971b). In own chromosome counts, B-chromosomes could rarely be identified based on smaller size compared to the remaining A-chromosomes (e.g. population numbers 24, 201; see Schulz (2015) for chromosome pictures). In most cases, no clear size differences between the chromosomes could be detected and therefore odd chromosome numbers were interpreted as aneuploidies rather than B-chromosomes (chromosome numbers in Table I-3 are given as total numbers).

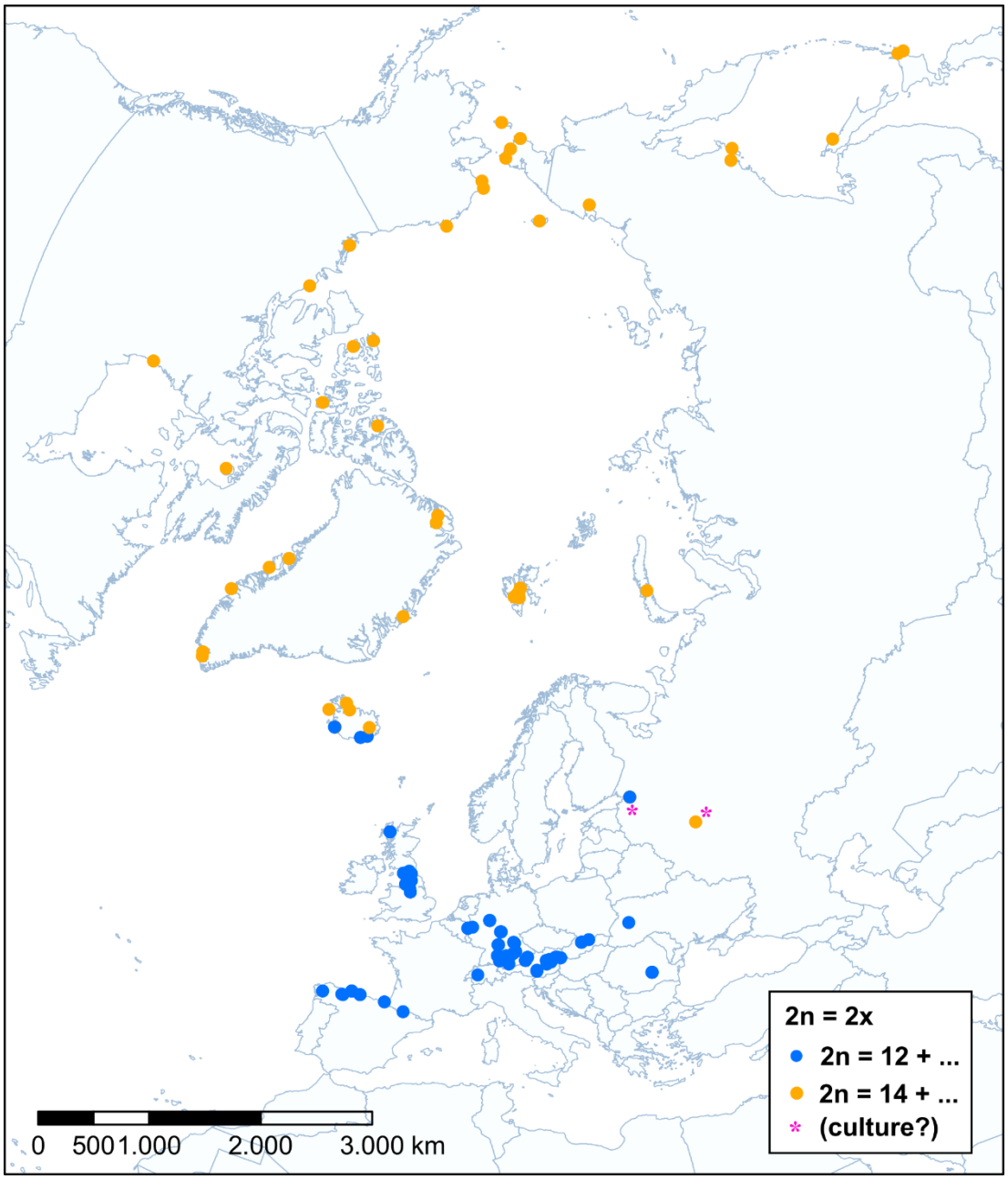

Figure I-2. Distribution of published and own chromosome counts for diploid *Cochlearia* samples (Supplemental Data Set I-3) showing a clear separation between two different base chromosome numbers of n = 6 (European distribution) and n = 7 (arctic distribution). Asterisks mark samples presumably originating from cultivation. Figure made by R. Schulz (taken from Schulz, 2015).

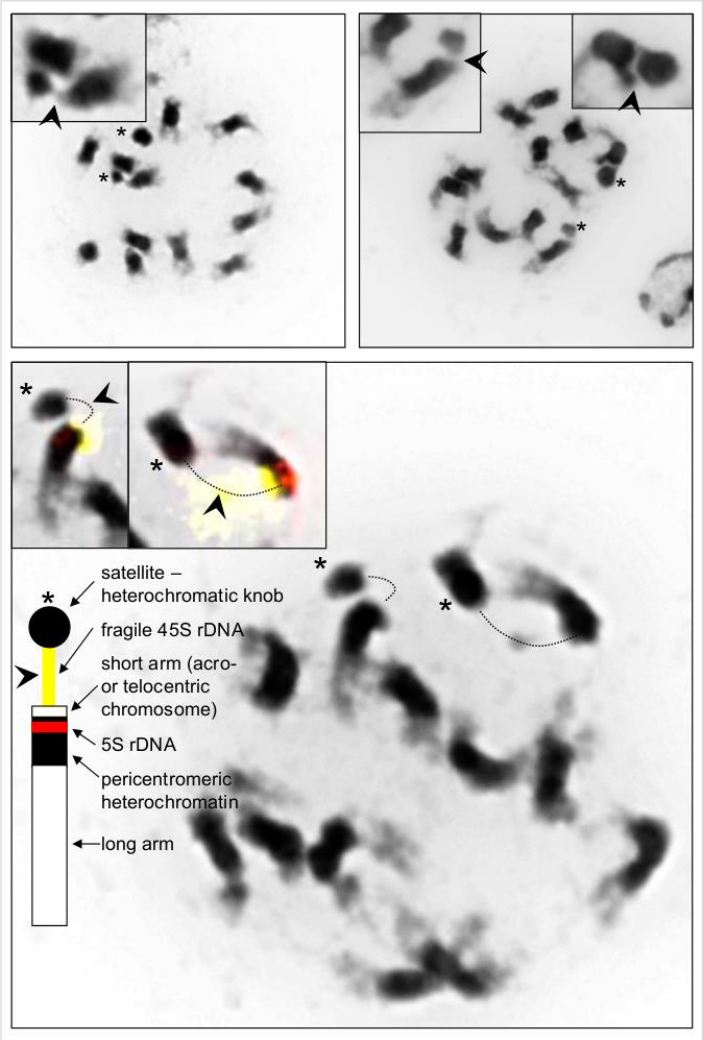

Figure I-3. Chromosome pictures of Icelandic *C. islandica* (2n = 12) generated by Dr. Martin Lysak and Dr. Terezie Mandáková (CEITEC, Brno) using flower buds. Asterisks indicate the presence of satellite heterochromatic knobs.

Another interesting observation concerned the occurrence of these aneuploidies (Figures I-4 and I-5). They were very rarely spotted in diploid taxa with only two counts of 2n=15 instead of 2n=12 in an Austrian (Pop. 24, HEID112736) and a Scottish population (Pop. 208, HEID112785) of *C. pyrenaica* and a single case of 2n=13 in an English population (Pop. 199, HEID112868). This picture was supported by the literature survey where few aberrations are documented for diploid taxa and especially regarding the arctic regions with only four described cases of aneuploidy. In contrast to this, aberrant chromosome numbers were frequently found in higher ploidy levels and especially along the coasts. Particularly Great Britain, the northern coast of Norway and the coasts between Kattegat and the Baltic Sea turned out to be hotspots of aneuploidies throughout all ploidy levels present. As illustrated in Figure I-6, only cases of hyper-aneuploidy but no observations of hypo-aneuploidy were found in diploid and tetraploid individuals whereas counts on the hexaploid level revealed both additional and missing chromosomes relating to their regular species-specific karyotypes. The only two octoploid samples analyzed in here, namely two individuals of *C. borzaeana*, had aneuploid chromosome numbers with two additional chromosomes each (Pop. 139, HEID112998 and pop. 140, HEID106723-5).

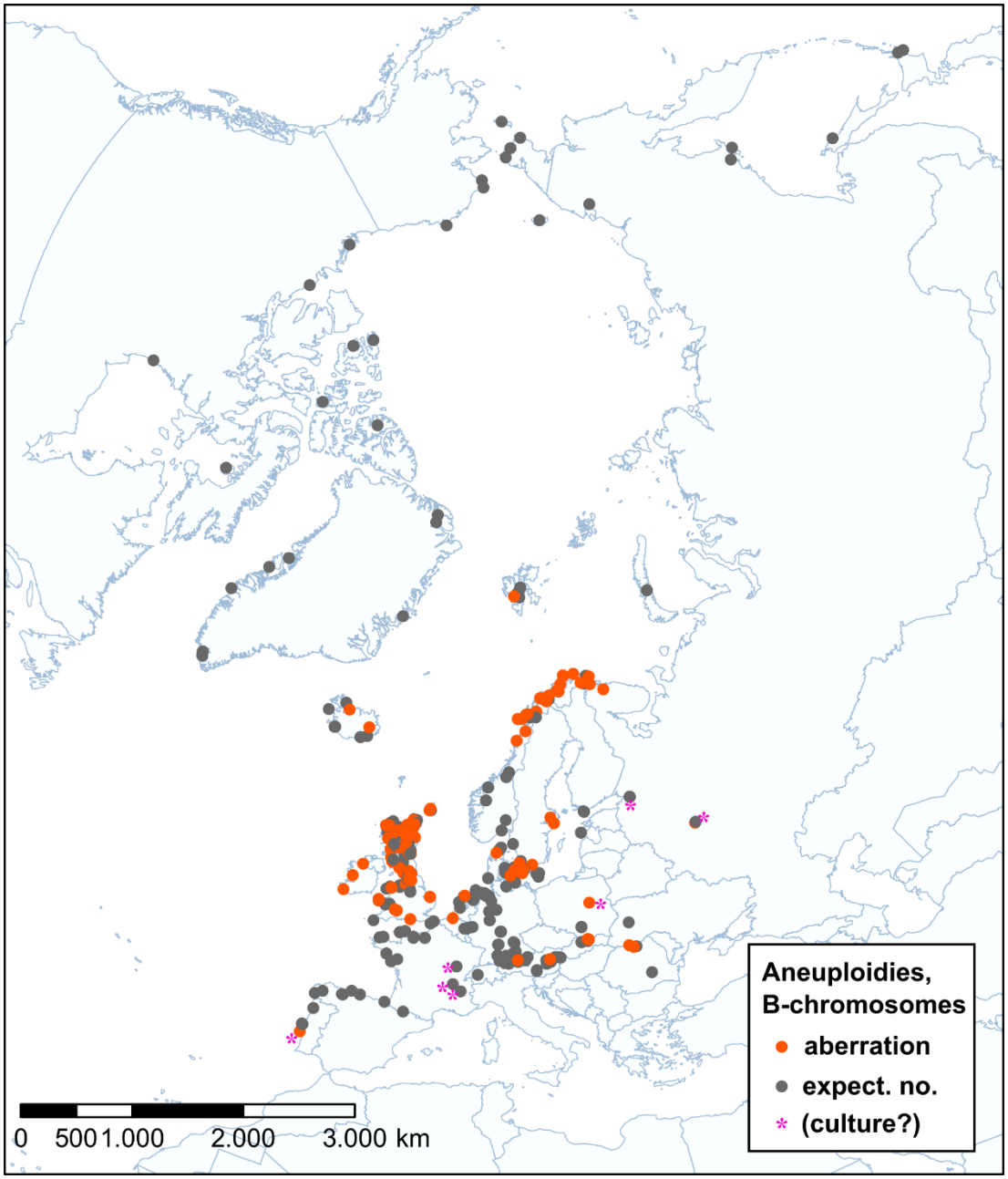

Figure I-4. Distribution of documented aneuploidies/B-chromosomes (literature and own measurements; Supplemental Data Set I-3) in samples/populations of *Cochlearia*, with expected chromosome numbers being 2n = 12/14/24/26/36/42/48. Besides documented B-chromosomes and other additional or lacking chromosomes, the hybrids *C. x hollandica* (2n=36) and *C. officinalis x danica* (2n=33) were treated as aberrations as well. Asterisks mark samples presumably originating from cultivation. Figure made by R. Schulz (taken from Schulz, 2015).

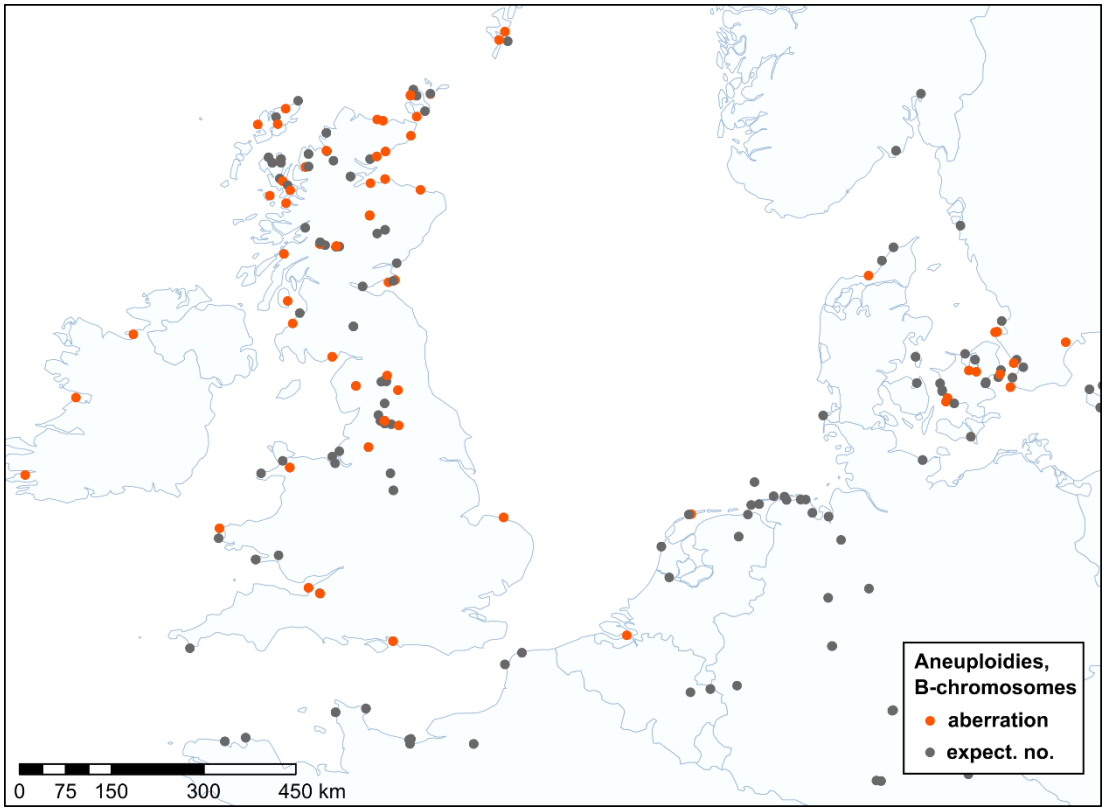

Figure I-5. Distribution of documented aneuploidies/B-chromosomes (literature and own measurements; Supplemental Data Set I-3) with a focus on Great Britain and adjacent coastal regions of continental Europe. Karyotypes of 2n = 12/14/24/26/36/42/48 are regarded as expected numbers whereas documented B-chromosomes and other additional or lacking chromosomes as well as the hybrids *C. x hollandica* (2n=36) and *C. officinalis x danica* (2n=33) are treated as aberrations. Figure made by R. Schulz, taken from Schulz (2015).

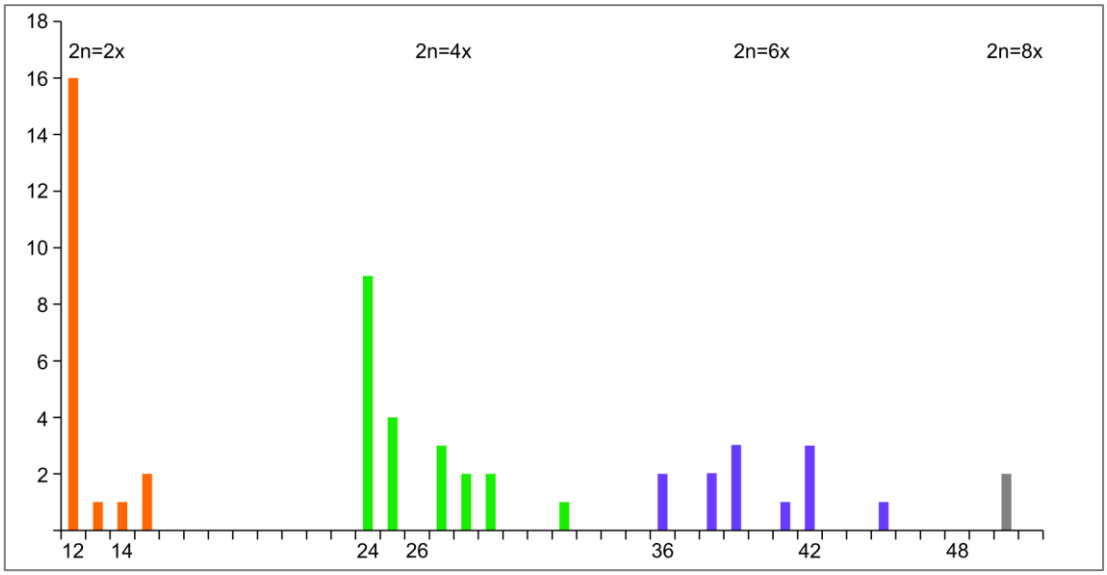

Figure I-6. Distribution of *Cochlearia* chromosome numbers in chromosome counts by R. Schulz and P. Sack (Supplemental Data Set I-2) with "regular" karyotypes given on the x-axis. Figure made by R. Schulz redrawn from Schulz (2015).

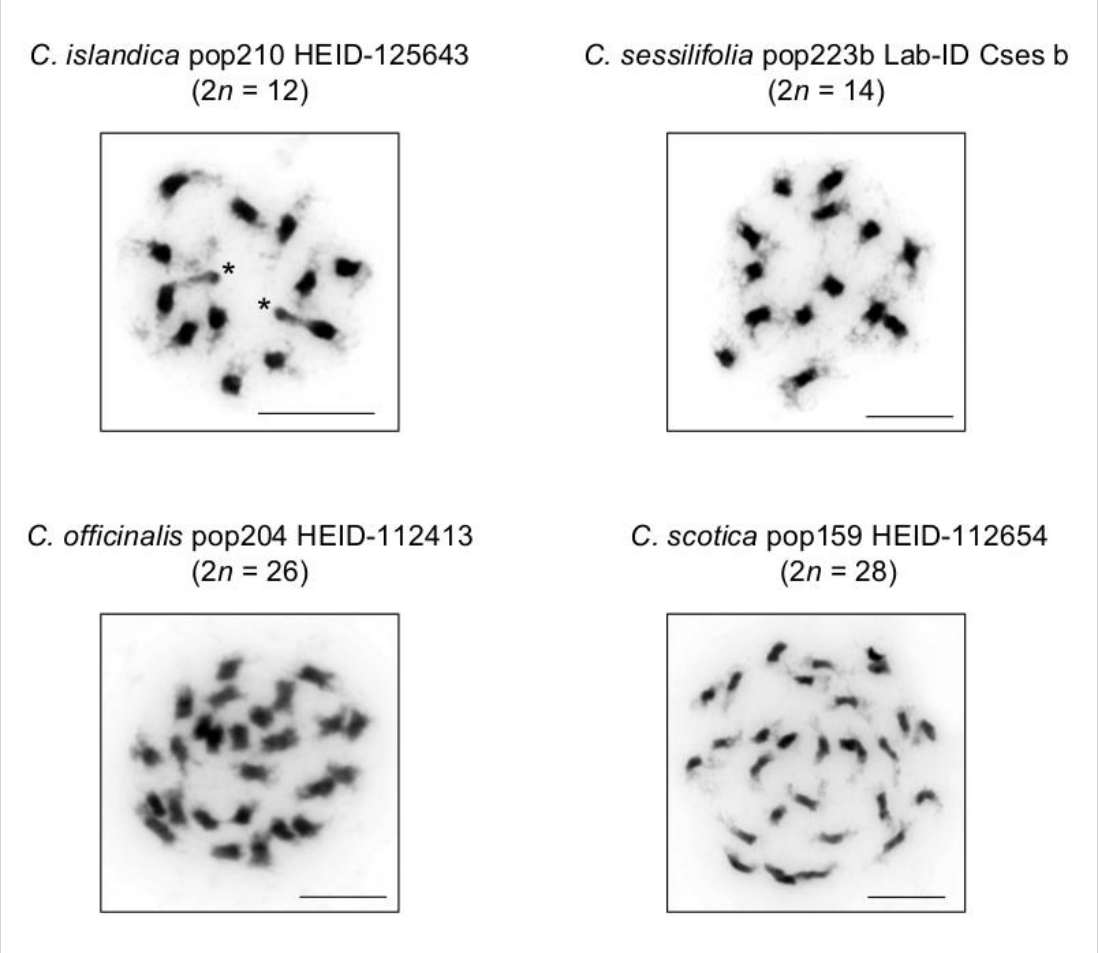

Figure I-7. Chromosome pictures produced by Dr. Martin Lysak and Dr. Terezie Mandáková (CEITEC, Brno) using flower buds. Asterisks indicate the presence of satellite heterochromatic knobs. Chromosome numbers in pop. 204 and pop. 159 are deviating from species-specific karyotypes.

1.4.2 Flow cytometry

Genome size measurements were generated for a total of 84 individuals out of 43 populations (Table I-3 and Supplemental Data Set I-1). Individual 2C-values ranged from 0.65 pg to 3.09 pg with the smallest measured 2C-value found in Icelandic *C. islandica* (2n=2x=12) and the largest one in Romanian *C. borzaeana* (2n=8x=50, HEID106723-5).

If more than one sample per accession was analyzed, detected genome sizes within several accessions turned out to vary to a certain degree (see Table I-3, e.g. populations 10, 157, 160, 163, 203, 210) while other populations had very consistent values. The largest difference within a single population was found in population no. 163 with 0.21 pg difference between the smallest and the largest measured value. While technical issues cannot be completely ruled out (see Schulz, 2015), it still seems apparent that genome size variation is a common phenomenon within populations and especially if aberrant chromosome numbers have been detected as well. A quite strong variation was observed between the genome size measurements of different *C. islandica* samples, which

might be a technical issue concerning the flow cytometer (before and after maintenance) or due to aneuploidies. However, the latter have not been documented for this population so far and since they are generally rare in diploid taxa, here, additional measurements should be performed in order to either confirm natural variation within the population or to reveal the true genome size. In contrast to this, genome sizes within the Central European mainland 2n=12 diploid samples were found to be comparatively stable also compared to higher ploidy levels where a higher variation was observed (see Figures I-8 and I-9). Interestingly, the two populations of high-alpine *C. excelsa* turned out to differ in their genome size. Whereas the average genome size measured for the population from the Gurktal Alps was ~0.78 pg, the population from the Seckau Tauern had an average genome size of ~0.86 pg which was the largest value measured for any of the 2n=12 populations.

Relationship between chromosome number and genome size

As revealed from a simple linear regression analysis including all measured *Cochlearia* samples (Supplemental Data Set I-1, sheet 4 'R input': *2C*, *ChrAll*) genome sizes within *Cochlearia* significantly increase with ploidy level (R^2 value: 0.74; Figure I-8 and Supplemental Data Set I-4). However, according to the R-squared value (the proportion of total variation explained by the model), the accuracy of the correlation was much higher after excluding arctic diploids and samples of *C. danica* as putative outliers (R^2 value: 0.97; Figure I-9). Also, as shown in Supplemental Data Set I-4 the quantile-quantile (QQ) plot for the second regression was reasonably consistent with the normality of the data.

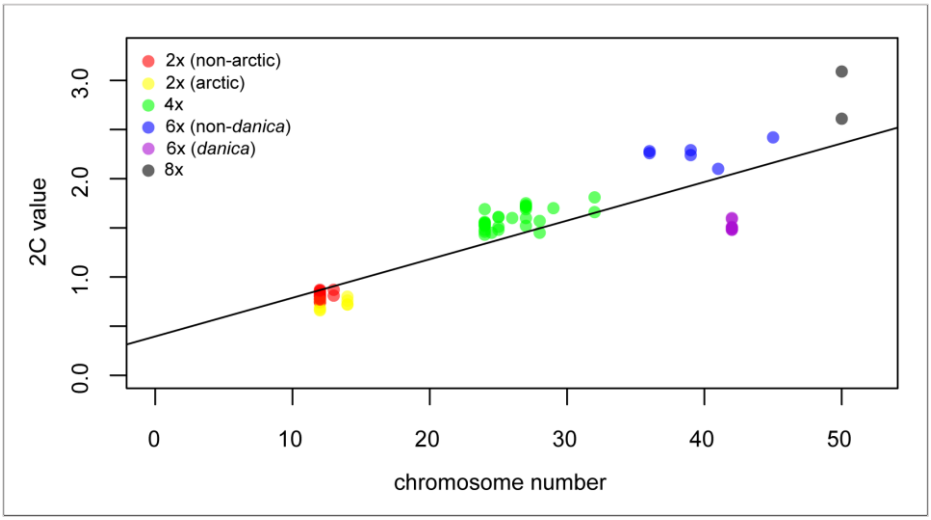

Figure I-8. Regression of increase in 2C values (in picograms (pg)) with higher ploidy levels and respective chromosome numbers in all analyzed *Cochlearia* samples (R^2 = 0.74; Supplemental Data Set I-4).

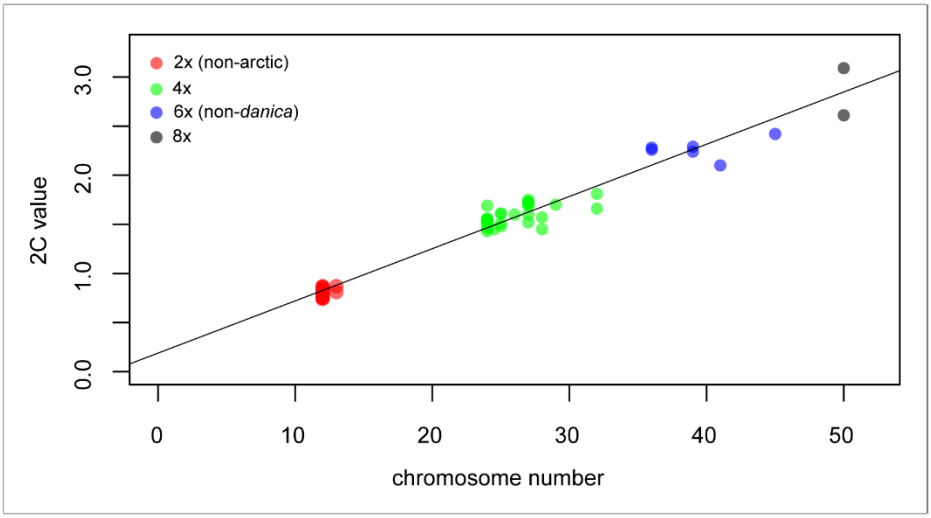

Figure I-9. Regression of increase in 2C values (in picograms (pg)) with higher ploidy levels and respective chromosome numbers in analyzed *Cochlearia* samples excluding arctic diploids and *C. danica* as putative outliers (R^2 = 0.97; Supplemental Data Set I-4).

Likewise, as illustrated in Figure I-10, per chromosome size significantly decreases with ploidy level (Supplemental Data Set I-1, sheet 4 'R input': *2C value per chromosome*, *ChrAll*). Here, the scatterplot clearly shows deviating values in arctic diploids and in *C. danica*. Thus again, the regression analysis was performed both including the putative outliers (arctic diploids and *C. danica*) and excluding them (Figure I-11), and again, the reduced data set produced better results both based on R-squared values (all data points: R^2 = 0.46; reduced data set: R^2 = 0.49) and respective QQ plots (Supplemental Data Set I-4). The predicted decrease in chromosome size for the reduced data set was 3.395e-04 pg DNA per additional chromosome which refers to ~0.26 megabase pairs (Mb) following Hohmann et al. (2014; *Arabidopsis thaliana*: 0.173 pg DNA = 135 Mb) .

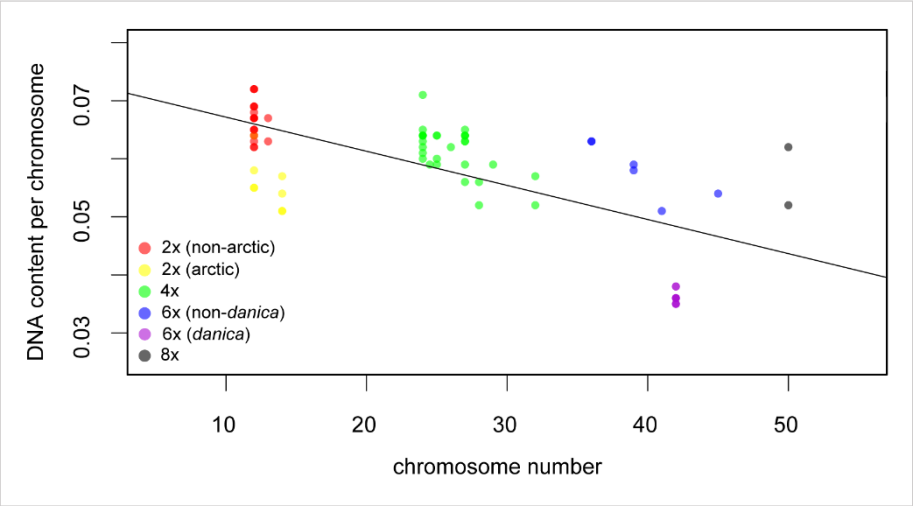

Figure I-10. DNA content per chromosome of all analyzed *Cochlearia* samples plotted against respective chromosome numbers. Simple linear regression analysis showed that chromosomal DNA content declines with increasing chromosome numbers (R^2 = 0.46; Supplemental Data Set I-4).

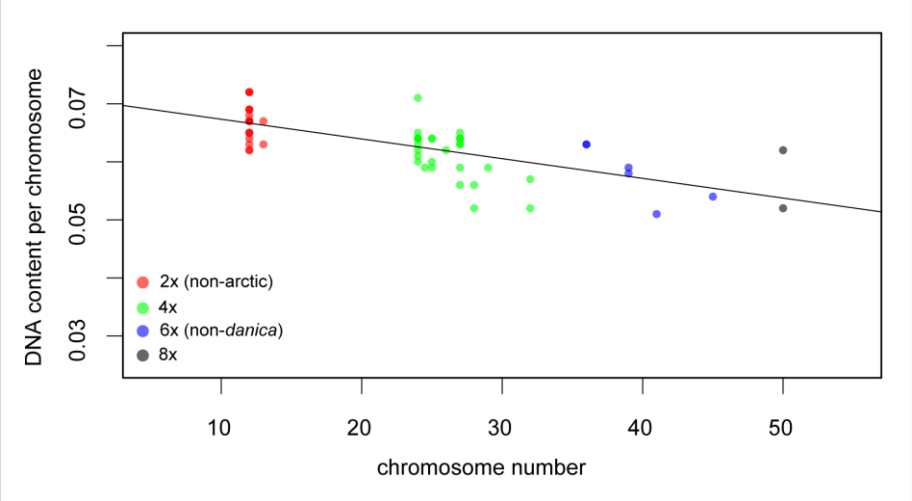

Figure I-11. DNA content per chromosome plotted against respective chromosome numbers for analyzed *Cochlearia* samples excluding arctic diploids and *C. danica* as putative outlier samples (R^2 = 0.49; Supplemental Data Set I-4).

Both correlation tests based on the complete dataset (including arctic diploids and *C. danica*) were highly significant and revealed a) a positive correlation between 2C value and chromosome number and b) a negative correlation between DNA content per chromosome and chromosome number (Table I-4 and Supplemental Data Set I-4).

Table I-4. Correlation Analyses. Correlations between 2C value / chromosome number and DNA content per chromosome / chromosome number based on the full dataset (including putative outliers). Detailed output given with Supplemental Data Set I-4.

	Kendall's rank correlation tau	**Spearman's rank correlation rho**
2C-value / chromosome number	0.603** (P = 2.2e-13)	0.782** (P < 2.2e-16)
DNA content per chromosome / chromosome number	-0.542** (P = 5.1e-11)	-0.681** (P = 3.5e-12)

1.5 Discussion

1.5.1 Two diploid karyotypes and the question of the ancestral base chromosome number

The literature review, complemented by own chromosome counts, nicely confirms a clear geographical distinction of arctic (2n=14) vs. central European (2n=12) diploid cytotypes (Figure I-2), which might be the result of an early split of the two different cytotypes and separation hereafter. Regarding this picture of the geographical karyotype distribution, one of the most interesting questions certainly concerns the ancestral base chromosome number in *Cochlearia*. This question has been addressed before and based on former cytological and hybridization studies (Gill, 1971a, 1973) revealing the formation of trivalents in F_1 hybrids between the two diploid cytotypes, it has been hypothesized that n=6 was the ancestral chromosome number, whereas n=7 represents the derived state that originated via primary tetrasomy. Thus, a diploid ancestor with 2n=12 was proposed at the base of the genus which implies a putative origin south of the glaciated regions in central Europe. An alternative scenario would start from the 2n=14 diploids that are nowadays restricted to arctic regions, which is why the second scenario would likely include arctic refuge areas during the Pleistocene as discussed and documented for other plant species before (e.g. Abbott et al., 2000; Abbott and Brochmann, 2003). The genus-wide phylogenomic analyses, presented in chapter 2, are more in favor of the second scenario with an ancestral base chromosome number of n=7 in ancestral arctic diploids and further reduction to n=6 in the European diploids. In this scenario, Iceland might have played a central role in the evolution of the two base chromosome numbers since it represents the only (recent) contact zone. Here, the transformation from n=7 to n=6 might have taken place via chromosome fusion and centromere loss which would partly explain the smaller calculated DNA content per chromosome in arctic samples. Reductions of base chromosome numbers have been described for other Brassicaceae taxa before (Mandáková and Lysak, 2008) and chromosome loss/fusion via reciprocal translocations and centromere loss is meanwhile well understood (e.g. Schranz et al., 2006; Lysak, 2014; Mandáková et al., 2016). However, the question of the ancestral base chromosome number in *Cochlearia* can't be finally answered based on the cytogenetic methods used herein. Phylogenetic methods can give further insights and both scenarios presented in here will be further discussed in chapter 2 in concert with the results gained from the analyses of all three plant genomes. In order to fully understand the processes leading to the karyotype constitutions we see today within the genus, more detailed studies are needed, including for example comparative chromosome painting and focusing e.g. on centromere and telomere placement and the mode of chromosomal rearrangements. A first step in this direction was achieved in a recent study on post-polyploid diploidization processes in mesopolyploid Brassicaceae taxa, including *Cochlearia pyrenaica*, by Mandakova et al. (2017). Here, a mesopolyploidy event was revealed for the tribe Cochlearieae and the authors point out that different base chromosome numbers are one common outcome of the processes following whole genome duplication events.

1.5.2 Stability, instability and chaos? – Diploidy, polyploidy and aneuploidy in *Cochlearia*

Besides the picture of a clear separation between arctic and European diploid cytotypes, one of the most interesting results from the cytogenetic analyses is the frequency and geographic distribution of aberrant chromosome numbers (Figures I-4 and I-5). Whereas rarely spotted in diploid species, the frequency of their occurrence is strikingly high throughout higher ploidy levels and especially along the coasts. However, the origin of these aberrations is not always completely apparent. Former studies have repeatedly described the occurrence of B-chromosomes or "accessory chromosomes" in different *Cochlearia* taxa (e.g. Gill, 1971b, 1973; Gill et al., 1978; Nordal, 1988) but with regard to the occurrence of satellite chromosomes as described before by Saunte (1955) for different ploidy levels and observed in here based on more precise chromosome pictures (Figures I-3 and I-7), and given the overall difficulty of defining different chromosome types based on their size (Schulz, 2015), these results should be handled with caution. Thus, rather than ascribing unexpected or odd chromosome numbers in *Cochlearia* to the presence of true B-chromosomes, these might often rather be classified either as result of mechanical disruption at fragile chromosomal regions during sample preparation or as a consequence of the taxa's polyploid and/or hybrid origin followed by independent ascending/descending dysploidies via chromosome rearrangements (see e.g. Lysak, 2014) or by mitotic/meiotic difficulties (reviewed e.g. by Comai, 2005). The formation of aberrant chromosome numbers in *Cochlearia* hybrid swarms has already been suggested before by several authors (e.g. Lövkvist, 1963a, b; Gill, 1971a). One tempting hypothesis on the origin of smaller sized B-chromosome-like additional chromosomes in intermediate and high ploidy levels would be past hybridization events with the hexaploid *C. danica* (Saunte, 1955) for which strong differences in chromosome morphology and size within its 2n=42 karyotype have been described (e.g. Saunte, 1955; Gill, 1976). In any case, the high frequency of aberrant chromosome numbers in the coastal populations of Central Europe, where intermediate (putatively) hybrid plants (e.g. Fearn, 1977; Pegtel, 1999) are frequently found, can be interpreted as a signature of highly dynamic and ongoing evolutionary processes in these habitats fueled by nearly absent interspecific fertility barriers (e.g. Gill, 1971a, 1973; Koch et al., 1998) and the dynamics of seed dispersal via sea-water (see e.g. Kadereit et al., 2005). Aside from (auto-)polyploidization, hybridization has long been recognized as an important origin of new genetic diversity, putatively bringing about new adaptive potential (e.g. Anderson, 1949; Arnold, 1997; Barton, 2001). As reviewed by Hoffmann and Sgrò (2011) this adaptive potential might also play a role in climate adaptation. According to Mandáková et al. (2016), the karyotypic variation and aneuploidy resulting from hybridization can be seen as an "adjustment" of the parental subgenomes within the allopolyploid taxa, putatively associated with the beginning diploidization. While interploidy hybridization in Angiosperms is often hindered, especially by endosperm-based barriers to hybridization (reviewed by Lafon-Placette and Köhler, 2016), the rates of hybridization across ploidy levels are remarkably high in *Cochlearia* and further studies on the resulting adaptive outcome as investigated by Pegtel (1999) might shed new light on the general patterns of hybridization and adaptation.

Regarding the strong tendency for (allo-)polyploidization in central Europe, it is rather surprising that the arctic regions are in contrast perfectly stable on the diploid level (Figure I-1). Likewise, polyploid taxa are found in alpine regions of Eastern Europe (*C. tatrae* from the High Tatra Mountains) and northern UK (*C. micacea*, *C. alpina*) but *C. excelsa*, the only high alpine taxon in the European Alps, is also a diploid species (2n = 12). This stands in contrast to the observation that polyploid plants are frequently found in more arctic/alpine environments than their diploid relatives (Otto and Whitton, 2000; Brochmann et al., 2004). Reasons for this phenomenon might be a broader range of metabolic tolerance to ecological conditions in polyploids and/or an increased occurrence of unreduced gametes in cold habitats (Otto and Whitton, 2000). These effects are either not present throughout the arctic regions and the European Alps or other factors prevent the appearance of polyploid *Cochlearia* taxa in these environments. Here, the high rate of endopolyploidy determined via flow cytometry, might serve as an alternative way of adjusting to the metabolic needs of the plants (see e.g. Comai, 2005). Based on the observation of a high morphological plasticity, a quick reaction to and thus strong tolerance for changing environments has been suggested for *C. groenlandica* (Zmudczyńska-Skarbek et al., 2013), which could be connected to the stability of the diploid level.
The described pattern of a steady diploid level (both 2n=12 and 2n=14 karyotypes) compared to the highly variable and instable polyploid karyotypes with frequently detected aneuploidies and intrapopulational variability in chromosome number and genome size, nicely matches the hypothesis of a necessary balance between genome variability or flexibility and a certain stability as suggested for the Brassicaceae family (Hohmann et al., 2015) and discussed recently by Schubert and Vu (2016) on a more general level. While strict genomic stability would inhibit adaptation to the continuously changing environment, moderate stability, retained by the described processes following WGD events, will allow for further adaptive evolution (Hohmann et al., 2015).

1.5.3 Genome duplication, genome size increase and subsequent parallel genome downsizing

As shown recently for the whole Brassicaceae family (Hohmann et al., 2015), polyploidization does not necessarily lead to seemingly ever increasing genome sizes, as revealed for example for *Fritillaria* from the Liliaceae family (Kelly et al., 2015). Instead, genome downsizing and diploidization seem to keep Brassicaceae genome sizes comparatively stable over long-term evolutionary timescales. However, regarding the positive correlation of chromosome numbers and genome size as revealed via both Kendall's tau and Spearman's rank correlation coefficient (rho) (Table I-4), and also via linear regression analysis (Figures I-8, I-9), these processes are not (yet) as apparent in the genus *Cochlearia* where several events of neopolyploidization are obviously too young to be withdrawn again on a genomic scale. Still, the slight but significant negative correlation of chromosome size with increasing chromosome numbers (Figures I-10, I-11), independent of inclusion or exclusion of putative outliers, shows a tendency of reducing genomic redundancy in the young polyploid taxa and thus supports the general trend observed for the Brassicaceae (Hohmann et al., 2015). This trend of genome downsizing in higher ploidy levels has

also been described by Kochjarová et al. (2006) for *C. tatrae* and *C. borzaeana* with 1.1–1.6 times smaller chromosomes compared to diploid *C. pyrenaica* analyzed in the same study.

The fact that the most prominent decrease of DNA content per chromosome is found in *C. danica* can most probably be attributed to a selfing syndrome and its annuality, bringing about a higher drought-resistance (e.g. Koch, 2012). Drastic DNA loss has been shown to have happened in other annuals (e.g. Jones and Brown, 1976; Jakob et al., 2004), the most prominent example being *Arabidopsis thaliana* (~125 Mb) that strongly reduced its genome size since the divergence from the perennial *A. lyrata* (~200 Mb) about 10 million years ago (Hu et al., 2011). It has been suggested that this correlation between smaller plant genome sizes and annuality is associated with the accordingly shorter cell cycles and the resulting benefits of a faster development (Bennett, 1972). Yet, Albach and Greilhuber (2004) showed a correlation of low genome sizes and selfing in *Veronica*, putatively due to less transposable elements in selfing plants, and thus argue that rather the connection between selfing and annuality might be the reason for the various described cases of genome size decrease in annual plants. Anyway, smaller chromosomes obtained via hybridization with *C. danica*, might also be another explanation for the decrease of DNA content per chromosome observed for populations/taxa with chromosome numbers between 2n=24 (*C. officinalis*) and 2n=42 (*C. danica*) as seen in Figure I-11.

1.6 References

Abbott, R.J., and Brochmann, C. (2003). History and evolution of the arctic flora: in the footsteps of Eric Hultén. Molecular ecology **12,** 299-313.

Abbott, R.J., Smith, L.C., Milne, R.I., Crawford, R.M., Wolff, K., and Balfour, J. (2000). Molecular analysis of plant migration and refugia in the Arctic. Science **289,** 1343-1346.

Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. Current opinion in plant biology **8,** 135-141.

Albach, D.C., and Greilhuber, J. (2004). Genome size variation and evolution in *Veronica*. Annals of Botany **94,** 897-911.

Albert, V.A., Barbazuk, W.B., Der, J.P., Leebens-Mack, J., Ma, H., Palmer, J.D., Rounsley, S., Sankoff, D., Schuster, S.C., and Soltis, D.E. (2013). The *Amborella* genome and the evolution of flowering plants. Science **342,** 1241089.

Anderson, E. (1949). Introgressive hybridization. (John Wiley And Sons, Inc.; New York; Chapman And Hall, Limited; London).

Appel, O., and Al-Shehbaz, I.A. (2002). Cruciferae. In The Families and Genera of Vascular Plants, Vol. V, K. Kubitzki, ed (Heidelberg, Germany: Springer Berlin)**,** pp. 75–174.

Arnold, M.L. (1997). Natural hybridization and evolution. (Oxford University Press).

Bajer, A. (1950). Cytological studies on *Cochlearia polonica*. Acta Societatis Botanicorum Poloniae **20,** 635-646.

Barker, M.S., Vogel, H., and Schranz, M.E. (2009). Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. Genome Biology and Evolution **1,** 391-399.

Barow, M. (2006). Endopolyploidy in seed plants. Bioessays **28,** 271-281.

Barow, M., and Meister, A. (2003). Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. Plant, Cell & Environment **26,** 571-584.

Barton, N. (2001). The role of hybridization in evolution. Molecular Ecology **10,** 551-568.

Bennett, M. (1972). Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London B: Biological Sciences **181,** 109-135.

Bennett, M., and Leitch, I. (2005). Plant DNA C-values database. Royal Botanic Gardens, Kew.

Blanc, G., and Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. The Plant Cell **16,** 1667-1678.

Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome research **13,** 137-144.

Böcher, T.W. (1938). Zur Zytologie einiger arktischen und borealen Blütenpflanzen. Svensk botanisk tidskrift **32,** 346-361.

Böcher, T.W., and Larsen, K. (1950). Chromosome Numbers of Some Arctic Or Boreal Flowering Plants: Den Botaniske Ekspedition Til Vestgrønland 1946. (Reitzel).

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422,** 433-438.

Brandrud, M.K. (2014). Polyploidy and ecotype variation in *Cochlearia officinalis* L. and related species. Master thesis. University of Oslo.

Brochmann, C., Brysting, A., Alsos, I., Borgen, L., Grundt, H., Scheen, A.C., and Elven, R. (2004). Polyploidy in arctic plants. Biological Journal of the Linnean Society **82,** 521-536.

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M.A., and Wang, X. (2013). Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. The Plant Cell **25,** 1541-1554.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nature reviews genetics **6,** 836-846.

Costich, D.E., Friebe, B., Sheehan, M.J., Casler, M.D., and Buckler, E.S. (2010). Genome-size variation in switchgrass (*Panicum virgatum*): flow cytometry and cytology reveal rampant aneuploidy. The Plant Genome **3,** 130-141.

Couvreur, T.L., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Molecular Biology and Evolution **27,** 55-71.

Crane, M.B., and Gairdner, A. (1923). Species-crosses in *Cochlearia*, with a preliminary account of their cytology. Journal of Genetics **13,** 187-200.

De Bodt, S., Maere, S., and Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. Trends in Ecology & Evolution **20,** 591-597.

Dodsworth, S., Leitch, A.R., and Leitch, I.J. (2015). Genome size diversity in angiosperms and its influence on gene space. Current opinion in genetics & development **35,** 73-78.

Dodsworth, S., Chase, M.W., and Leitch, A.R. (2016). Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? Botanical Journal of the Linnean Society **180,** 1-5.

Doležel, J., Binarová, P., and Lucretti, S. (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. Biologia plantarum **31,** 113-120.

Doležel, J., Greilhuber, J., and Suda, J. (2007a). Estimation of nuclear DNA content in plants using flow cytometry. Nature protocols **2,** 2233-2244.

Doležel, J., Greilhuber, J., and Suda, J. (2007b). Flow cytometry with plant cells: analysis of genes, chromosomes and genomes. (John Wiley & Sons).

Edger, P.P., and Pires, J.C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Research **17,** 699.

Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glöckner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., and Tang, M. (2015). The butterfly plant arms-race escalated by gene and genome duplications. Proceedings of the National Academy of Sciences **112,** 8362-8366.

Elkington, T. (1984). Cytogenetic variation in the British flora: Origins and significance. New Phytologist **98,** 101-118.

Fawcett, J.A., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proceedings of the National Academy of Sciences **106,** 5737-5742.

Fearn, G. (1977). A morphological and cytological investigation of *Cochlearia* populations on the Gower peninsula, Glamorgan. New Phytologist **79,** 455-458.

Fernandez Casas, J. (1975). Numeros cromosomicos de plantas espanolas: 2. Anales del Instituto Botánico A. J. Cavanilles **32,** 301-307.

Flovik, K. (1940). Chromosome numbers and polyploidy within the flora of Spitzbergen. Hereditas **26,** 430-440.

Franzke, A., Koch, M.A., and Mummenhoff, K. (2016). Turnip Time Travels: Age Estimates in Brassicaceae. Trends in plant science **21,** 554-561.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in plant science **16,** 108-116.

Geiser, C., Mandáková, T., Arrigo, N., Lysak, M.A., and Parisod, C. (2015). Repeated whole-genome duplication, karyotype reshuffling and biased retention of stress-responding genes in Buckler Mustards. The Plant Cell**,** TPC2015-00791-RA.

Gill, B., McAllister, H., and Fearn, G. (1978). Cytotaxonomic studies on the *Cochlearia officinalis* L. group from inland stations in Britain. Watsonia **8,** 395-396.

Gill, J. (1965). Diploids in the genus *Cochlearia*. Watsonia **6,** 188-189.

Gill, J. (1971a). Cytogenetic studies in *Cochlearia* L. The chromosomal homogeneity within both the 2n = 12 diploids and the 2n = 14 diploids and the cytogenetic relationship between the two chromosome levels. Annals of Botany **35,** 947-956.

Gill, J. (1971b). The cytology and transmission of accessory chromosomes in *Cochlearia pyrenaica* DC. (Cruciferae). Caryologia **24,** 173-181.

Gill, J. (1973). Cytogenetic studies in *Cochlearia* L. (Cruciferae). The origins of *C. officinalis* L. and *C. micacea* Marshall. Genetica **44,** 217-234.

Gill, J. (1975). *Cochlearia* L. In: Hybridisation and the Flora of the British Isles (Ed. by CA Stace). (London, New York, San Francisco.: Academic Press).

Gill, J. (1976). Cytogenetic studies in *Cochlearia* L. (Cruciferae). The chromosomal constitution of *C. danica* L. Genetica **46,** 115-127.

Guignard, M.S., Nichols, R.A., Knell, R.J., Macdonald, A., Romila, C.A., Trimmer, M., Leitch, I.J., and Leitch, A.R. (2016). Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. New Phytologist.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., and Wendel, J.F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome research **16,** 1252-1261.

Hedberg, O. (1966). Chromosome numbers of vascular plants from arctic and sub-arctic North America.

Heubl, G. (1996). Bemerkungen zur Karyologie der *Cochlearia pyrenaica*-Gruppe unter besonderer Berucksichtigung von *C. macrorrhiza* (Schur) Pobed. Berichte der Bayerischen Botanischen Gesellschaft**,** 66-67.

Hoffmann, A.A., and Sgrò, C.M. (2011). Climate change and evolutionary adaptation. Nature **470,** 479-485.

Hohmann, N., Wolf, E.M., Lysak, M.A., and Koch, M.A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. The Plant Cell **27,** 2770-2784.

Hohmann, N., Schmickl, R., Chiang, T.-Y., Lučanová, M., Kolář, F., Marhold, K., and Koch, M.A. (2014). Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. BMC evolutionary biology **14,** 224.

Holmen, K. (1952). Cytological studies in the flora of Peary Land, North Greenland. Meddelelser om Grønland **128,** 1-40.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., and Gundlach, H. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nature genetics **43,** 476-481.

Jakob, S.S., Meister, A., and Blattner, F.R. (2004). The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Molecular Biology and Evolution **21,** 860-869.

Jones, R.N., and Brown, L. (1976). Chromosome evolution and DNA variation in *Crepis*. Heredity **36,** 1-104.

Kadereit, J.W., Arafeh, R., Somogyi, G., and Westberg, E. (2005). Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. Taxon**,** 861-876.

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., and Links, M.G. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. The Plant Cell **26,** 2777-2791.

Kakes, P. (1973). The chromosome number of *Cochlearia pyrenaica* DC. near Moresnet (Belgium). Acta botanica neerlandica **22,** 206-208.

Karl, R., and Koch, M.A. (2013). A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. Annals of botany**,** mct165.

Kelly, L.J., Renny-Byfield, S., Pellicer, J., Macas, J., Novák, P., Neumann, P., Lysak, M.A., Day, P.D., Berger, M., and Fay, M.F. (2015). Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytologist **208,** 596-607.

Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., and Koch, M.A. (2013). BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Plant and Cell Physiology**,** pct158.

Koch, M., Hurka, H., and Mummenhoff, K. (1996). Chloroplast DNA restriction site variation and RAPD-analyses in *Cochlearia* (Brassicaceae): Biosystematics and speciation. Nordic Journal of Botany **16,** 585-603.

Koch, M., Huthmann, M., and Hurka, H. (1998). Isozymes, speciation and evolution in the polyploid complex *Cochlearia* L. (Brassicaceae). Botanica acta **111,** 411-425.

Koch, M., Mummenhoff, K., and Hurka, H. (1999). Molecular phylogenetics of *Cochlearia* (Brassicaceae) and allied genera based on nuclear ribosomal ITS DNA sequence analysis contradict traditional concepts of their evolutionary relationship. Plant systematics and evolution **216,** 207-230.

Koch, M.A. (2012). Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). Taxon**,** 76-92.

Kochjarová, J., Valachovič, M., Bureš, P., and MrÁZ, P. (2006). The genus *Cochlearia* L. (Brassicaceae) in the Eastern Carpathians and adjacent area. Botanical Journal of the Linnean Society **151,** 355-364.

Krisai, R., and Greilhuber, J. *Cochlearia pyrenaica* DC, das Löffelkraut, in Oberösterreich (mit Anmerkungen zur Karyologie und zur Genomgröße). Beiträge zur Naturkunde Oberösterreichs **5,** 151-160.

Lafon-Placette, C., and Köhler, C. (2016). Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. Molecular ecology **25.11,** 2620-2629.

Leitch, A., and Leitch, I. (2008). Genomic plasticity and the diversity of polyploid plants. Science **320,** 481-483.

Leitch, I., Soltis, D., Soltis, P., and Bennett, M. (2005). Evolution of DNA amounts across land plants (Embryophyta). Annals of Botany **95,** 207-217.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., Ma, J., Yu, J., and Huang, S. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nature communications **5**.

Loureiro, J., Rodriguez, E., Doležel, J., and Santos, C. (2006). Flow cytometric and microscopic analysis of the effect of tannic acid on plant nuclei and estimation of DNA content. Annals of Botany **98,** 515-527.

Löve, Á., and Löve, D. (1948). Chromosome numbers of northern plant species. (Ingólfsprent Reykjavík).

Lövkvist, B. (1963a). Något om de skånska *Cochlearia*-arterna. Bot. notiser **116,** 326-330.

Lövkvist, B. (1963b). Taxonomic problems in aneuploid complexes. Regnum vegetab. **27,** 51-57.

Ludwig, W. (1960). *Cochlearia officinalis* s. str. und *Cochlearia pyrenaica* in Hessen. Hessische floristische Briefe **10,** 51-53.

Lysak, M.A. (2014). Live and let die: centromere loss during evolution of plant chromosomes. New Phytologist **203,** 1082-1089.

Lysak, M.A., and Mandáková, T. (2013). Analysis of plant meiotic chromosomes by chromosome painting. Plant Meiosis: Methods and Protocols**,** 13-24.

Lysak, M.A., Mandakova, T., and Schranz, M.E. (2016). Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. Current Opinion in Plant Biology **30,** 108-115.

Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A., and Leitch, I.J. (2009). The dynamic ups and downs of genome size evolution in Brassicaceae. Molecular Biology and Evolution **26,** 85-98.

Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proceedings of the National Academy of Sciences of the United States of America **103,** 5224-5229.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist **207,** 437-453.

Mandakova, T., Li, Z., Barker, M.S., and Lysak, M.A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. The Plant Journal.

Mandáková, T., and Lysak, M.A. (2008). Chromosomal phylogeny and karyotype evolution in x= 7 crucifer species (Brassicaceae). The Plant Cell **20,** 2559-2570.

Mandáková, T., Gloss, A.D., Whiteman, N.K., and Lysak, M.A. (2016). How diploidization turned a tetraploid into a pseudotriploid. American journal of botany **103.7,** 1187-1196.

Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K., and Lysak, M.A. (2010). Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. The Plant cell **22,** 2277-2290.

Mandáková, T., Mummenhoff, K., Al-Shehbaz, I.A., Mucina, L., Mühlhausen, A., and Lysak, M.A. (2012). Whole-genome triplication and species radiation in the southern African tribe *Heliophileae* (Brassicaceae). Taxon **61,** 989-1000.

Matsuura, H., and Sutô, T. (1935). Contributions to the idiogram study in phanerogamous plants I. Journal of the Faculty of Science, Hokkaido Imperial University. Ser. 5, Botany **5,** 33-75.

Maude, P.F. (1939). The Merton catalogue a list of the chromosome numerals of species of British flowering plants. New Phytologist **38,** 1-31.

Mayfield-Jones, D., Washburn, J.D., Arias, T., Edger, P.P., Pires, J.C., and Conant, G.C. (2013). Watching the grin fade: Tracing the effects of polyploidy on different evolutionary time scales. In Seminars in cell & developmental biology (Elsevier), pp. 320-331.

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V., and Lewis, K.L. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature **452,** 991-996.

Mosquin, T., and Hayley, D.E. (1966). Chromosome numbers and taxonomy of some Canadian arctic plants. Canadian Journal of Botany **44,** 1209-1218.

Nordal, I. (1988). *Cochlearia pyrenaica* DC., a species new to Scotland. Watsonia **17,** 49-52.

Nordal, I., and Laane, M. (1990). Cytology and Reproduction in arctic *Cochlearia*. Sommerfeltia **11,** 147-158.

Nordal, I., Eriksen, A., Laane, M., and Solberg, Y. (1986). Biogeographic and biosystematic studies in the genus *Cochlearia* in Northern Scandinavia. Symbolae Botanicae Upsalienses **27,** 83-93.

Otto, S.P. (2007). The evolutionary consequences of polyploidy. Cell **131,** 452-462.

Otto, S.P., and Whitton, J. (2000). Polyploid incidence and evolution. Annual review of genetics **34,** 401-437.

Peer, W.A., Mamoudian, M., Lahner, B., Reeves, R.D., Murphy, A.S., and Salt, D.E. (2003). Identifying model metal hyperaccumulating plants: germplasm analysis of 20 Brassicaceae accessions from a wide geographical area. New Phytologist **159,** 421-430.

Pegtel, D.M. (1999). Effect of ploidy level on fruit morphology, seed germination and juvenile growth in scurvy grass (*Cochlearia officinalis* L. s.l., Brassicaceae). Plant Species Biology **14,** 201-215.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

Saunte, L.H. (1955). Cytogenetical studies in the *Cochlearia officinalis* complex. Hereditas **41,** 499-515.

Schranz, M.E., and Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. Plant Cell **18,** 1152-1165.

Schranz, M.E., Lysak, M.A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends in Plant Science **11,** 535-542.

Schranz, M.E., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Current Opinion in Plant Biology **15,** 147-153.

Schubert, I., and Vu, G.T. (2016). Genome stability and evolution: attempting a holistic view. Trends in plant science **21,** 749-757.

Schulz, R. (2015). Evolutionary insights from cytogenetic analyses: the case of *Cochlearia* L. Master thesis. Universität Heidelberg.

Soltis, D.E., and Soltis, P.S. (1999). Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution **14,** 348-352.

Soltis, D.E., Soltis, P.S., and Tate, J.A. (2004). Advances in the study of polyploidy since plant speciation. New phytologist **161,** 173-191.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. American journal of botany **96,** 336-348.

Soltis, P.S., Liu, X., Marchant, D.B., Visger, C.J., and Soltis, D.E. (2014). Polyploidy and novelty: Gottlieb's legacy. Philosophical Transactions of the Royal Society B **369,** 20130351.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., and Harmon, L.J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytologist **207,** 454-467.

Thompson, J.N., Nuismer, S.L., and Merg, K. (2004). Plant polyploidy and the evolutionary ecology of plant/animal interactions. Biological Journal of the Linnean Society **82,** 511-519.

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. Genome research **24,** 1334-1347.

Vision, T.J., Brown, D.G., and Tanksley, S.D. (2000). The origins of genomic duplications in *Arabidopsis*. Science **290,** 2114-2117.

Vogt, R. (1985). Die *Cochlearia pyrenaica*-Gruppe in Zentraleuropa. Berichte der Bayerischen Botanischen Gesellschaft **56,** 5-52.

Vogt, R. (1987). Die Gattung *Cochlearia* L. (Cruciferae) auf der Iberischen Halbinsel. Mitteilungen der Botanischen Staatssammlung München **23,** 393-421.

Warwick, S., Francis, A., and Al-Shehbaz, I. (2006). Brassicaceae: species checklist and database on CD-Rom. Plant Systematics and Evolution **259,** 249-258.

Wendel, J.F. (2000). Genome evolution in polyploids. In Plant molecular evolution (Springer), pp. 225-249.

Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. Nature Reviews Genetics **2,** 333-341.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., and Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. Proceedings of the national Academy of sciences **106,** 13875-13879.

Zhukova, P.G. (1965). Kariologicheskaya kharakteristika nekotorykh rasteniy Chukotskogo poluostrova. Bot. Zhur. **50,** 1001-1004.

Zhukova, P.G. (1967). Chromosome numbers in some species of plants of the north-eastern part of the USSR II. Bot. Zhur. **52,** 982-987.

Zmudczyńska-Skarbek, K., Barcikowski, M., Zwolicki, A., Iliszko, L., and Stempniewicz, L. (2013). Variability of polar scurvygrass *Cochlearia groenlandica* individual traits along a seabird influenced gradient across Spitsbergen tundra. Polar Biology **36,** 1659-1669.

Zuccolo, A., Sebastian, A., Talag, J., Yu, Y., Kim, H., Collura, K., Kudrna, D., and Wing, R.A. (2007). Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evolutionary Biology **7,** 1.

Chapter 2 – *Cochlearia* Phylogenomics

2.1 Abstract

Chapter 2 provides comprehensive phylogenomic analyses based on all three plant genomes (plastid, mitochondrial, nuclear), covering the whole distribution range of the genus *Cochlearia* and including all ploidy levels and ecotypes. Organellar phylogenies are largely in congruence and reveal major evolutionary lineages. Here, highly interesting results are the putatively parallel colonization of and thus adaptation to the different central European high alpine regions and the fact that the most basal evolutionary lineages within the trees are found in arctic regions which implies a glacial survival of the whole genus in the arctic. Based on whole chloroplast genomes, divergence time estimates are generated, revealing a diversification of the genus within the last ~700,000 years. Moreover, divergence times of the six major chloroplast lineages nicely match high glacial periods, indicating a strong impact of these past climatic fluctuations on the evolution of the genus *Cochlearia*.

Nuclear sequence data supports 1) the clear separation of arctic and Central European gene pools as revealed before by the distribution pattern of diploid karyotypes in chapter 1, as well as 2) the basal phylogenetic position of the artic 2n=14 taxon *C. tridactylites* from eastern Canada as also revealed by organellar phylogenies. This supports an early colonization of and survival in arctic regions and hints at n=7 as the ancestral base chromosome number. Based on genetic assignment tests, several cases of putative admixture between the two major gene pools are detected. However, based on the presented data, a final conclusion on whether the "admixture" between the Central European and the arctic gene pool is a result of repeated southward migration from arctic areas or rather a genetic legacy from the suggested arctic origin, carried by recent *Cochlearia* in varying degrees, cannot be made.

2.2 Introduction

2.2.1 NGS sequencing in phylogenetics and phylogeography

A nicely resolved phylogeny is a key requisite to answer many of the questions that arise in evolutionary biology as it enables for example to ascribe evolutionary important events to certain nodes of a tree, to reveal the ancestral state of a trait or to give estimates on timing and speed of different evolutionary processes (Schranz and Mitchell-Olds, 2006; Schranz et al., 2006). Likewise, a sound phylogenetic and biogeographic background of the study system is expedient in order to interpret the genetic patterns and signatures of plant adaptation. In the past this has often been challenging due to limited amounts of sequence information combined with mechanisms covering the genomic traces of phylogenetic evolution such as incomplete lineage sorting, hybridization or hidden paralogy (Templeton et al., 1992; Maddison and Knowles, 2006). Thanks to the new technologies of next generation sequencing (NGS) that have been rapidly growing over the course of the last decades (for review see Shendure and Ji, 2008), collecting large amounts of genome-wide sequencing data points and unravelling the evolutionary histories of model as well as non-model taxa with it, has become a feasible task and meanwhile there exists a wealth of different methods and strategies applied for different research questions (e.g. Moore et al., 2007; Egan et al., 2012; Cronn et al., 2012; Weitemier et al., 2014).

Besides large scale phylogenetic histories, that have taken place over the course of millions of years (e.g. Brassicaceae phylogeny and divergence times (e.g. Hohmann et al., 2015; Huang et al., 2015)), it has now come within reach to analyze evolution acting on younger and shorter time-scales thanks to the immense amount of sequence information generated via NGS (e.g. Whittall et al., 2010). The Pleistocene (~2.58 - 0.0117 mya) as a time of drastic climatic fluctuations had a dramatic impact on dimensions and shape of the biodiversity in the northern hemisphere (Hewitt, 1996; Comes and Kadereit, 1998; Hewitt, 2000; Hewitt, 2004; Alsos et al., 2005). Recurrent ice formation and retreat brought about massive extinction events, with survival of most taxa only possible in geographically restricted ice-free refuge areas, but at the same time, it putatively accelerated speciation and reticulation via secondary contact zones especially during recolonizing in interglacial periods or post-glacial times (Abbott et al., 2000; Abbott and Brochmann, 2003; Jordon-Thaden and Koch, 2008). The Iberian Peninsula, Southern Italy and the Balkan have been described early on as main refuge areas where European plant and animal species survived glacial maxima (Taberlet et al., 1998) but besides these, a list of other, so called cryptic refugia north of these main centers of glacial survival has been suggested (e.g. Provan and Bennett, 2008; dicussed in Stewart and Lister, 2001) making the phylogeographic history of survival and recolonization more complex. Thus, these dynamic processes often resulted in intricate distribution patterns and cryptic phylogenetic relationships (e.g. Koch et al., 2006; Skrede et al., 2006; Allen et al., 2012). Recovering these phylogenetic and phylogeographic histories that have happened over the comparatively short intervals of glaciation and deglaciation is challenging but highly interesting as it might improve our knowledge on evolution acting on small evolutionary time-scales in reaction to drastic environmental changes and likely including phenomena like the parallel adaptation to similar, geographically separated habitat types.

Chloroplast DNA, which is maternally inherited in most Angiosperms (reviewed by Reboud and Zeyl, 1994), has long been utilized for phylogeographic studies owing to several characteristics (seed dispersal only, low mutation rate, limited gene flow, small effective population size) that facilitate the reconstruction of e.g. historic migration routes (e.g. Comes and Kadereit, 1998; Petit et al., 2002). The first complete chloroplast genomes have already been sequenced in 1986 with *Nicotiana tabacum* (Shinozaki et al., 1986) and the liverwort *Marchantia polymorpha* (Ohyama et al., 1986). While this meant a huge technical and analytical effort at that time, and even in 1999, more than 10 years later, when the sequence of the *Arabidopsis thaliana* chloroplast genome was revealed (Sato et al., 1999), it meanwhile became a viable task, thanks to the advent of massively parallel sequencing in combination with the plastome being present in many copies per cell and showing a comparatively stable genome architecture (Wicke et al., 2011). Therefore, whole chloroplast genome sequencing is nowadays among the most popular tools for reconstructing large or complex plant phylogenies (e.g. Parks et al., 2009; Njuguna et al., 2013; Ruhfel et al., 2014; Stull et al., 2015; reviewed by Gao et al., 2010). Another benefit of whole chloroplast genome data besides these highly resolved phylogenies lies in the possibility to generate fine-scale divergence time estimates based thereon as shown recently for the Brassicaceae phylogeny by Hohmann et al. (2015).

In contrast to chloroplast data, mitochondrial data has been rather neglected by phylogenetic studies so far, due to the generally lower variability combined with much higher complexity of mitochondrial genomes and the dynamics of their evolution (Burger et al., 2003; Knoop, 2004; Knoop et al., 2011; Steele and Pires, 2011; Straub et al., 2012). Yet, thanks to modern-day NGS strategies, sequence information originating from the mitochondrial genome is readily generated e.g. via genome-skimming approaches (Straub et al., 2011; Straub et al., 2012; Bock et al., 2014) and although mitochondrial data analysis is still challenging, depending on the complexity of the respective mitochondrial genome, its incorporation into phylogenomic analyses can add valuable information to the evolutionary history of the study system (Straub et al., 2012).

Besides the organellar genomes, NGS of total genomic DNA usually comes along with a wealth of sequencing reads originating from the nuclear genome, even in shallow sequencing approaches (Straub et al., 2012). This information can be analyzed and interpreted in many ways, even in the absence of a reference genome (e.g. Davey et al., 2011). Yet, SNP (single nucleotide polymorphism) calling strategies clearly benefit from the availability of such a reference as it facilitates and improves the data analysis and it extends the possibilities for data interpretation, e.g. by making use of gene annotations.

2.2.2 *Cochlearia* phylogenetics

As stated above in the general introduction, several aspects of the evolutionary history of the study genus *Cochlearia* have already been unraveled (e.g. Gill, 1965; Nordal et al., 1986; Koch et al., 1996; Koch et al., 1998; Koch, 2002; Kochjarová et al., 2006; Koch, 2012), yet many aspects remained cryptic so far. Initial studies mainly focused on the complex cytology of the genus and taxonomic considerations combined with first evolutionary hypotheses were based on these and morphological as well as ecological characters (e.g. Crane and Gairdner, 1923; Saunte, 1955; Gill,

1971, 1973, 1976; Gill et al., 1978). Later on, these studies were complemented by glucosinolate, isozyme, RFLP, RAPD, AFLP and DNA sequencing data leading to novel insights, taxonomic revisions and first comprehensive evolutionary concepts for the whole genus (e.g. Nordal et al., 1986; Koch et al., 1996; Koch et al., 1998; Koch et al., 2003; Koch, 2012) as outlined in the general introduction.

A central discovery concerning the genus as a whole was the deep split between *Cochlearia* and its Mediterranean sister-clade *Ionopsidium* which was dated to ~13.8 mya by Koch (2012) based on the two nuclear-encoded internal transcribed spacer regions ITS1 and ITS2 and the chloroplast trnL/trnF intergenic spacer. This split was apparently followed by a long period of evolutionary stasis in the genus *Cochlearia* since further diversification of recent taxa did not start before mid- or late Pleistocene (Koch, 2012). Hence, as already stated above, in this era of drastic environmental changes and periodically extremely harsh conditions (Hewitt, 2004), a series of cold-adapted species evolved, some of which are nowadays withstanding the extremes of arctic and alpine habitats. However, owing to the young age of the genus, previous marker systems failed to provide a highly resolved phylogeny (e.g. Koch et al., 1996; Koch et al., 1999), thus leaving several questions unanswered so far. Some of the pending questions regarding this evolutionary background are: 1) Where did the genus originate and where did it survive glacial periods? 2) When and at what speed did the diversification in the genus *Cochlearia* take place? 3) Was the adaptation to extreme cold conditions of arctic and (high) alpine regions accomplished once or several times in parallel? 4) What factors enabled and triggered the metamorphosis to a cold-adapted genus?

Based on NGS data analyses of all three plant genomes (chloroplast/mitochondrial/nuclear genome) generated via a genome-skimming approach (Straub et al., 2011; Straub et al., 2012), this chapter aims at providing novel, genus-wide insights in order to finally answer some of these questions, thereby paving the way for further in-depth studies e.g. as a model-system for the evolution of cold adaptation.

2.3 Material and methods

2.3.1 Plant material and taxon sampling

A total of 65 *Cochlearia* samples (Table II-1) were selected for phylogenomic analyses in order to cover the whole distribution range of the genus from northern Spain to circum-arctic regions and to include all different ploidy-levels and ecotypes (Figure II-1). With *C. islandica* treated as a separate species, tetraploid populations of British *C. pyrenaica* referred to as *C. pyrenaica* subsp. *alpina* (hereafter abbreviated as *C. alpina* or Calp) and eliding sub-species levels in *C. officinalis* samples from Scandinavia, 19 taxa were included in the study, comprising all accepted species as listed in the *BrassiBase* database (Kiefer et al., 2013). One sample per population/accession was

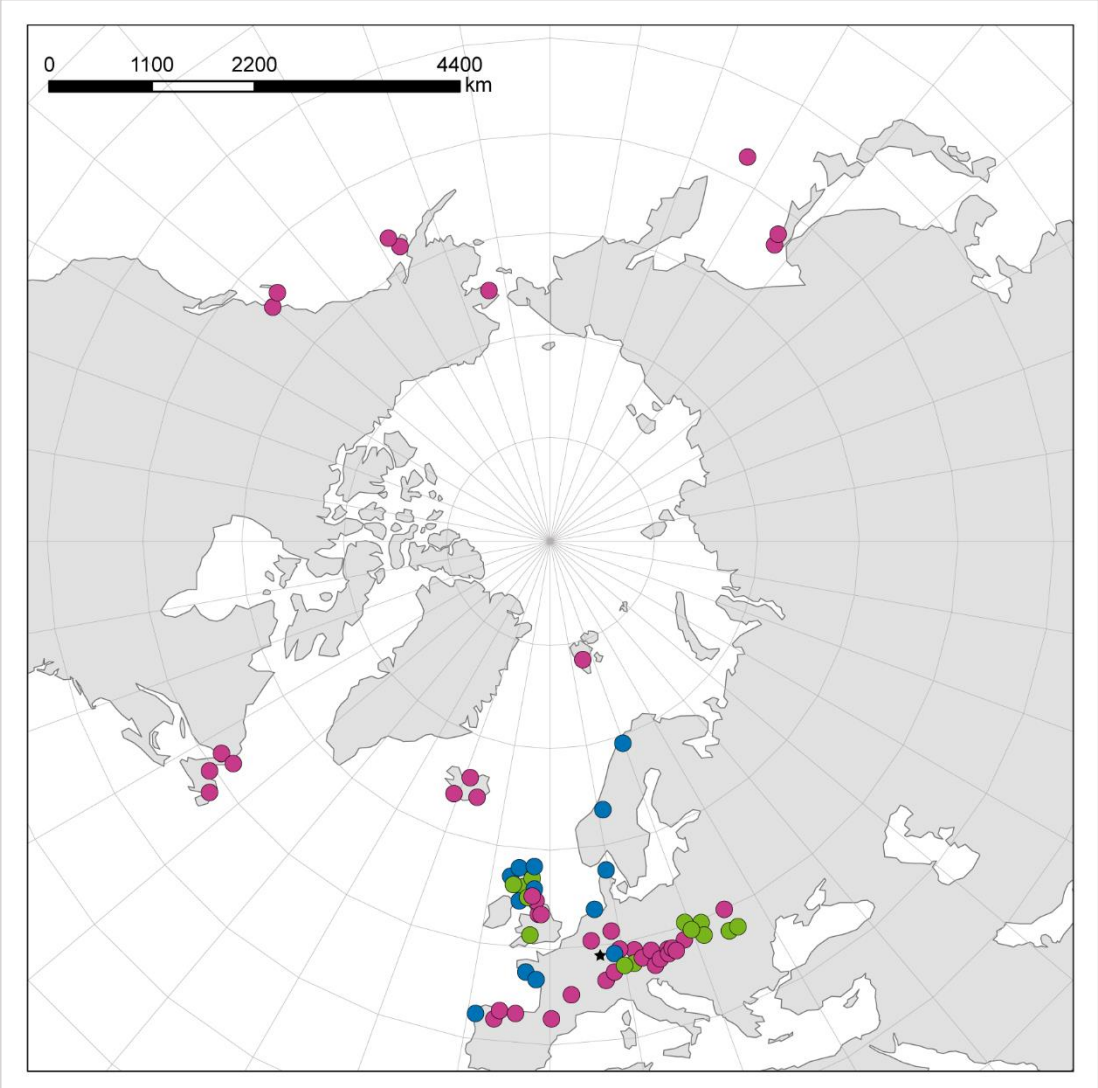

Figure II-1. Map showing the distribution of *Cochlearia* samples included in the study. Colors indicate different ploidy levels/ecotypes. Pink indicates diploid samples, green polyploid inland samples and blue polyploid coastal samples. The asterisk marks an inland sample of *C. danica* which is a coastal species rapidly dispersed inland along highways.

included in the analyses and three samples representing three species from the sister genus *Ionopsidium* were included as outgroup.

Table II-1. Samples included in this study with information on (putative) ploidy levels and origin. Where possible, missing coordinates were extracted (c.e.) based on information of the respective locality, otherwise coordinates are marked as n/a.

Species	Lab-ID	2n	Locality		lat	lon	alt	
C. aestuaria	Caes_0741	2x	ES	Asturias, Pesués	43.4	-4.5	0 m	
C. aestuaria	Caes_0160	2x	ES	Asturias, Ría del Nalón	43.5	-6.1	54 m	
C. pyr. subsp. *alpina*	Calp_0165	4x	UK	England, Somerset, Mendip Hills, Cheddar Gorge	51.3	-2.8	138 m	
C. pyr. subsp. *alpina*	Calp_1127	4x	UK	Scotland, Moffat, Hartfell Rig	55.4	-3.4	714 m	
C. pyr. subsp. *alpina*	Calp_0828	4x	UK	Scotland, Cairngorm Mountains	57.1	-3.7	1134 m	
C. anglica	Cang_1023	8x	FR	Dep. Morbihan, La Trinité-sur-Mer, "Le Men Du"	47.6	-3.1	6 m	c.e.
C. anglica	Cang_1157	8x	UK	Scotland, north of Creetown	54.9	-4.4	11 m	
C. bavarica	Cbav_1207	6x	DE	Bavaria, Glonn, river Glonn	48.0	11.9	556 m	
C. bavarica	Cbav_1208	6x	DE	Bavaria, Immenthal	47.8	10.4	811 m	
C. borzaeana	Cbor_1015	8x	RO	Carpathians, Distr. Maramures, Muntii	47.7	25.0	1226 m	
C. borzaeana	Cbor_1063	8x	RO	Carpathians, Distr. Maramures, Valea Salhoi	47.7	25.0	1226 m	
C. danica	Cdan_0654	6x	UK	Scotland, Moray, Lossiemouth	57.7	-3.3	7 m	
C. danica	Cdan_1067	6x	FR	Dep. Vendée, Barbâtre en Noirmoutier, "La Fosse"	46.9	-2.2	11 m	c.e.
C. danica	Cdan_1261	6x	DE	Lower Saxony, Borkum, western island	53.9	6.4	-	
C. danica	Cdan_1766	6x	DE	BW, Rheinstetten, along highway B36	49.0	8.3	118 m	c.e.
C. danica	Cdan_1778	6x	ES	Galicia, Carnota, boulders behind Stol Sea Farm	42.8	-9.2	1 m	
C. excelsa	Cexc_0297	2x	AT	Styria, Gurktal Alps, Mt. Eisenhut, Dieslingsee	47.0	13.9	2012 m	
C. excelsa	Cexc_1253	2x	AT	Styria, Seckau Tauern, Mt. Seckauer Zinken	47.3	14.7	2295 m	
C. groenlandica	Cgro_0474	2x	ZN	Svalbard, Bjørndalen	78.2	15.3	31 m	
C. groenlandica	Cgro_1038	2x	Alas.	Nome Census Area, Wooly Lagoon	64.9	-166.4	42 m	
C. groenlandica	Cgro_1764	2x	IS	Gilsbakkifjall, Skagafjardarssýsla	65.4	-19.0	1032 m	
C. islandica	Cisla_1233	2x	IS	Stokkseyri (South Coast)	63.8	-21.0	1 m	
C. islandica	Cisla_1765	2x	IS	Ingòlfshöfði, Austre-Skaftafellssýsla	63.8	-16.6	42 m	
C. macrorrhiza	Cmac_1022	2x	AT	Lower Austria, Moosbrunn	48.0	16.4	185 m	
C. micacea	Cmica_0979	4x	UK	Scotland, Beinn an Dothaidh	56.5	-4.7	743 m	
C. micacea	Cmica_0983	4x	UK	Scotland, Mt. Ben Lawers	56.5	-4.3	590 m	
C. officinalis	Coff_1024	4x	NO	Storfjord, Skibotn	69.4	20.2	289 m	
C. officinalis	Coff_1012	4x	UK	Scotland, N-Littleferry	56.2	-2.8	5 m	
C. officinalis	Coff_1073	4x	DK	Nordjylland, Hirtshals	57.6	10.0	16 m	c.e.
C. officinalis	Coff_1289	4x	NO	Storfjord, Stordal, salt marsh	63.5	10.9	0 m	
C. officinalis	Coff_1357	4x	RU	Kuril Archipelago, Ushishir Group, Yankicha Island	47.5	152.8	32 m	

C. officinalis	Coff_1366	4x	RU	Northeast Sakhalin, east of Piltun Village	52.7	-143.3	3 m	
C. officinalis	Coff_1367	4x	RU	Northeast Sakhalin, south of Okha	52.7	-143.3	1 m	
C. officinalis	Coff_1364	4x	CA	British Columbia, Along Sangan, Graham Island	54.0	-132.0	32 m	
C. officinalis	Coff_1365	4x	CA	British Columbia, Galloway Rapids	54.3	-130.3	98 m	
C. polonica	Cpol_1016	6x	PL	Olkusz, Pustynia Bledowska	50.3	19.6	332 m	
C. pyrenaica	Cpyr_0506	2x	DE	BW,Hohenlohe, Dörzbach, St. Wendel chapel	49.4	9.7	250 m	
C. pyrenaica	Cpyr_0699	2x	UK	England, Mid-West Yorkshire, Gordale Scar	54.1	-2.1	300 m	
C. pyrenaica	Cpyr_0021	2x	FR	Dep. Cantal, Mauriac, Ruisseau de la Bouige	45.1	2.6	1527 m	
C. pyrenaica	Cpyr_1053	2x	FR	Dep. Hautes-Pyrénées, Bagnères-de-Bigorre, Chiroulet	43.0	0.1	1110 m	
C. pyrenaica	Cpyr_1058	2x	UA	Verkhobuzh	49.9	25.1	298 m	
C. pyrenaica	Cpyr_0106	2x	CH	Canton Bern, Kandersteg, nature reserve Filfalle	46.5	7.7	1198 m	
C. pyrenaica	Cpyr_0260	2x	BE	Wallonia, province of Lüttich, Kelmis	50.7	6.0	207 m	
C. pyrenaica	Cpyr_0252a	2x	DE	NRW, Brilon, Alme spring	51.5	8.6	345 m	
C. pyrenaica	Cpyr_0456	2x	SK	Carpathians, Veľká Fatra National Park, Jazierce	49.0	19.3	555 m	
C. pyrenaica	Cpyr_0310	2x	AT	Lower Austria, Mostviertel, Traisen valley, Türnitz	47.9	15.5	525 m	
C. pyrenaica	Cpyr_1211	2x	ES	Asturias, between Villar de Vildas and La Pornacal	43.1	-6.3	1000 m	
C. pyrenaica	Cpyr_0560	2x	DE	BW, Allgäu, east of Isny, nature reserve Sägeweiher	47.7	10.0	701 m	
C. pyrenaica	Cpyr_0198	2x	DE	Bavaria, Taching am See, calcareous fen	48.0	12.7	480 m	
C. pyrenaica	Cpyr_1442	2x	UK	England, Malham, Malham Tarn fen	54.1	-2.2	379 m	
C. pyrenaica	Cpyr_0396	2x	AT	Styria, Leoben, Hieflau	47.7	14.8	530 m	
C. pyrenaica	Cpyr_0759	2x	UK	Scotland, Teesdale, Burnhope Seat	54.7	-2.4	621 m	
C. pyrenaica	Cpyr_1552	2x	UK	England, Alston, Nenthead	54.8	-2.3	449 m	
C. pyrenaica	Cpyr_1537	2x	AT	Lower Austria, Scheibbs, Lunz, Lehen	47.8	15.0	576 m	
C. scotica	Cscot_0963	4x	UK	Scotland, Ballachulish	56.7	-5.2	2 m	
C. scotica	Cscot_1009	4x	UK	Scotland, Achmelvich	58.2	-5.3	8 m	
C. sessilifolia	Cses_1285	2x	Alas.	Kodiak Island, Olds River, head of Kalsin Bay	57.5	-152.4	202 m	
C. sessilifolia	Cses_1286	2x	Alas.	Kodiak Island, Buskin River mouth and lagoon	57.8	-152.5	4 m	
C. tatrae	Ctat_1017	6x	SK	Carpathians, Tatra Mts., Velicka Dolina	49.2	20.2	1765 m	c.e.
C. tatrae	Ctat_1290	6x	SK	Carpathians, High Tatras, Hlinska Dolina	49.2	20.0	1802 m	
C. tatrae	Ctat_1306	6x	PL	Carpathians, Tatra Mts.	49.2	20.1	2317 m	c.e.
C. tridactylites	Ctri_1287	2x	CA	Labrador Straits, L'Anse Amour	51.5	-56.9	< 15 m	
C. tridactylites	Ctri_1288	2x	CA	Labrador Straits, Belle Isle, Wreck Cove	52.0	-55.4	6 m	
C. tridactylites	Ctri_1370	2x	CA	Newfoundland, Notre Dame Bay	49.5	-55.9	0 m	
C. tridactylites	Ctri_1371	2x	CA	Newfoundland, Bonavista South Dist., Overs Islands	48.5	-53.8	0 m	
I. abulense	Iabu_1074	2n=28	ES	Province of Avila	n/a	n/a	n/a	

I. acaule	Iacau_1072	2n=24	ES	n/a (Gomez Campo Collection Madrid)	n/a	n/a	n/a	
I. megalospermum	Imega_1776	2n=38	ES	Province of Granada, Sierra Nevada, Rio Laroles valley	37.0	-3.0	1155 m	c.e.

Samples were either collected in the wild and stored as herbarium vouchers, silica-dried leaf material and/or as living plants in the Botanical Garden Heidelberg, or they were received from different herbaria in form of herbarium specimen, silica-dried samples and/or seed material which was grown at the Botanical Garden Heidelberg. Table II-1 lists all samples/populations included in this study together with information on the respective sample locations. Information on the type of the respective leaf material is given in Table II-2.

Table II-2. Information on the leaf material type of 68 analyzed *Cochlearia* (65) and *Ionopsidium* (3) samples.

Fresh leaf material	Cpyr_0506, Ctat_1017, Cmac_1022, Cexc_0297, Cbor_1015, Cang_1023, Coff_1024, Cpol_1016, Calp_0828, Cbor_1063, Cdan_1067, Cscot_0963, Coff_1012, Cmica_0979, Iacau_1072
Silica-dried leaf material	Cpyr_0699, Cpyr_0021, Cgro_0474, Cdan_0654, Caes_0741, Cpyr_1053, Cpyr_1058, Cpyr_0106, Cpyr_0260, Cpyr_0252a, Cpyr_0456, Cgro_1038, Cpyr_0310, Calp_1127, Cpyr_1211, Cbav_1207, Cbav_1208, Cpyr_0759, Cmica_0983, Cang_1157, Cisla_1233, Caes_0160, Cscot_1009, Cpyr_0204, Cpyr_0560, Cpyr_0198, Ctat_1306, Cisla_1765, Calp_0165, Cgro_1764, Cpyr_1442, Cpyr_0396 , Cdan_1766, Cpyr_1552, Cpyr_1537, Cdan_1778
Herbarium specimen	Iabu_1074, Coff_1073, Cexc_1253, Cdan_1261, Ctat_1290, Cses_1285, Cses_1286, Ctri_1287, Ctri_1288, Coff_1289, Coff_1357, Coff_1366, Coff_1367, Ctri_1370, Coff_1364, Coff_1365, Ctri_1371, Imega_1776

2.3.2 DNA extraction and NGS sequencing

DNA was extracted from leaves of either herbarium material, silica-dried material or fresh material frozen in liquid nitrogen and grinded using a pistil (Table II-2). Grinding of dried leaf material was done in 2 ml tubes with five 2.5 mm glass beads, using a Precellys®24 homogenizer (Bertin Technologies, Montigny-le-Brettonneux, France) for 10 seconds at 5,000 rounds per minute. If necessary, the grinding step was repeated once. DNA extractions were performed using the Invisorb Spin Plant Mini Kit (STRATEC Biomedical AG, Birkenfeld, Germany) according to manufacturer's instructions, followed by an RNA digestion step with 2 units of RNase A for 30 min at 37°C. For five samples (Lab-IDs: Cpyr_0506, Ctat_1017, Cmac_1022, Cexc_0297, Cbor_1015), cpDNA was enriched prior to extraction by using a Percoll step-gradient centrifugation. Therefore, living plants were kept in the dark for 2 days in order to reduce starch accumulation in the chloroplasts. For the chloroplast isolation/enrichment, 2.5 g of fresh, young leaf material were homogenized on ice, with precooled mortar and pestle, in 30 ml of ice-cold extraction buffer (modified from Jansen et al. (2005): 0.35 M Sorbitol, 50 mM Tris-HCL (pH 8.0), 5 mM EDTA, 1% BSA) with 10 mM β-mercaptoethanol added right before use. The resulting suspension was filtered through a double layer of Miracloth, followed by a centrifugation of the flow-through for 15 min at 2000 g and 4°C. The pellet was resuspended in 1.5 ml of the extraction buffer and loaded onto a precooled discontinuous 3-step Percoll gradient (Percoll concentrations: 1.5 ml 80%, 1.5 ml 40%, 7 ml 35%). The gradients were centrifuged for 15 min at 5000g and 4°C

and the band of intact chloroplast that should have accumulated in between the 50% and the 35% Percoll layers was taken off, washed twice with the extraction buffer, followed by centrifugation, and finally resuspended in 1 ml of the extraction buffer. Hereafter, DNA was extracted following a CTAB protocol by (Doyle and Doyle, 1987). Therefore, samples were incubated for 30 min at 60°C with 800 µl of CTAB buffer (β-mercaptoethanol added). After adding 700 µl CI, samples were kept on a shaker for 20 min and centrifuged for 15 min at 5.5g. The supernatant was transferred to a new tube, 600 µl 2-Propanol were added and samples were kept on ice for 15 min and on -20°C for another 10 min. After centrifugation for 15 min at 7.6 rpm at 4°C, the supernatant was discarded and the pellet was washed twice with 70% ethanol, centrifuged for 15 min at 7.6 rpm and 4°C. Dried pellets were dissolved in 50 µl of TE buffer with 1 µl RNase A and incubated for 30 min at 37°C. In order to increase the DNA yield, CTAB extraction was followed by a rolling circle amplification of the purified genomic DNA using the REPLI-g Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Quality and quantity of the DNA extractions (both Invisorb Spin Plant Mini Kit and CTAB) were controlled prior to library preparation using 1% agarose gels as well as fluorescence spectroscopy on a Qubit® dsDNA HS Assay (Thermo Fisher Scientific, Waltham, Massachusetts, USA) using a high-sensitivity, double-stranded DNA specific dye.

For all samples, libraries of total genomic DNA (or enriched for cpDNA), with fragment sizes of 200 to 400 bp, were prepared at the CellNetworks Deep Sequencing Core Facility (Heidelberg) using either the TruSeq Kit (Illumina Inc., San Diego, California, U.S.) or the NEBNext Ultra DNA Library Prep Kit for Illumina (formerly NEBNext DNA Library Prep Kit for Illumina; New England Biolabs Inc., Ipswich, Massachusetts, U.S.) with the NEBNext Multiplex Oligos for Illumina. Per sample, about 100 to 500 ng of starting DNA was used (for earlier preparations more DNA (up to 5 µg) was needed) and fragmented on a Covaris S2 instrument. Six to twelve samples were multiplexed per lane and sequencing was performed in paired-end mode (100 bp) on an Illumina HiSeq 2000 system.

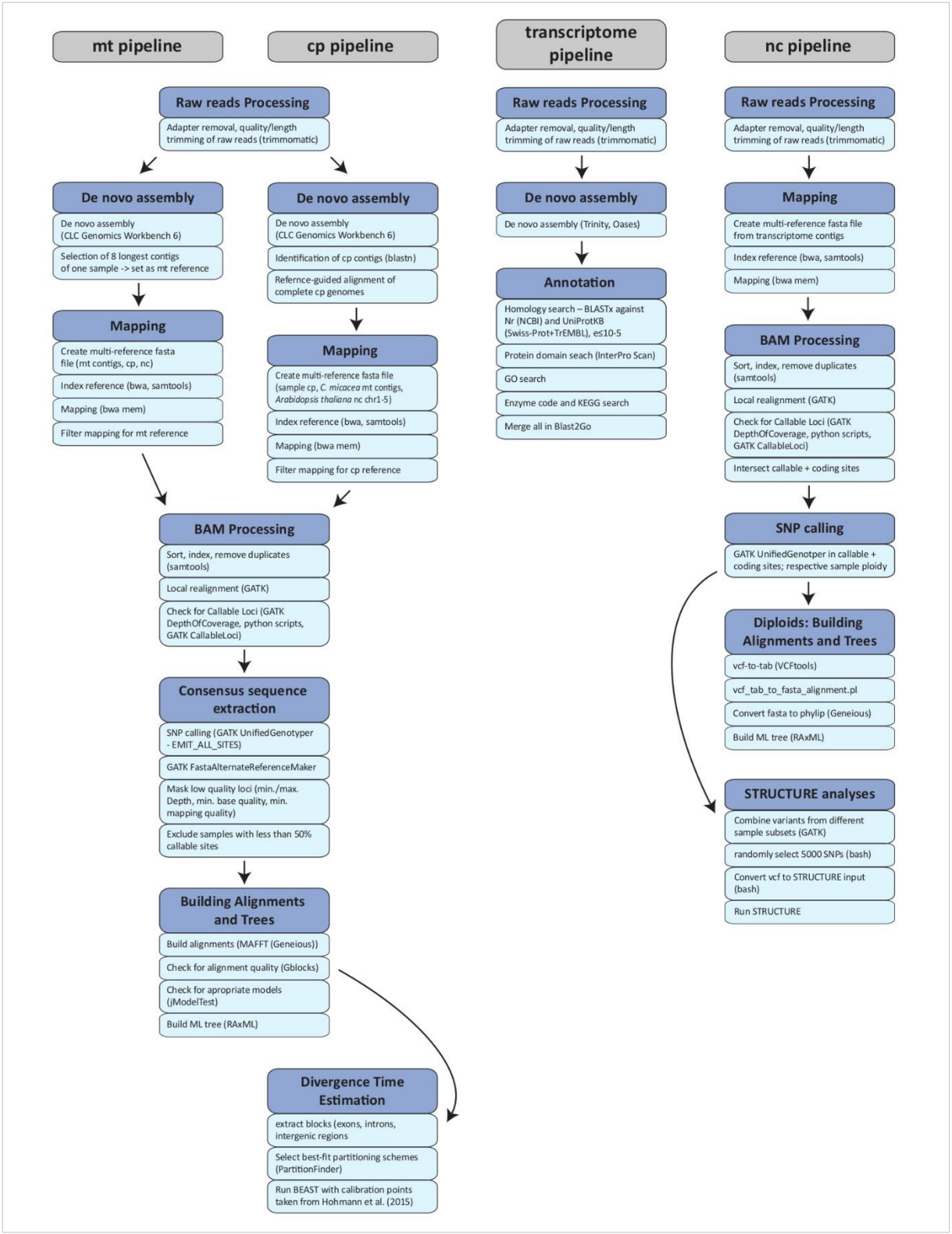

Figure II-2. Data analysis pipelines for chloroplast, mitochondrial, transcriptome and nuclear datasets. Transcriptome assembly and annotation were performed by Dr. Lúa López (Lopez et al., submitted).

2.3.3 Chloroplast genome data analysis

De novo assembly and reference-based control mappings

Whole chloroplast genome assembly was done using a combination of *de novo* assembly and reference-based mapping strategies (see Hohmann et al., 2015). As a first step, a quality and adapter trimming was performed on the raw data using the CLC Genomics Workbench version 6.0.4 (CLC Bio, Aarhus, Denmark) with a quality score limit of 0.001 (equal to a phred score of 30) and a minimum read length of 50 bp. *De novo* assemblies were carried out for all of the 68 samples with the CLC de novo algorithm, using only trimmed paired reads and with length and similarity fractions set to 0.8. A preselection of putative chloroplast *de novo* contigs was made based on average coverage levels and visual inspection. Hereafter, contig identity was verified using BLASTn (with default settings) as implemented in the browser version of the NCBI nucleotide blast, and chloroplast contigs were aligned manually using PhyDE v0.9971 (Muller et al., 2010). Contig alignment was guided by the chloroplast genome of one of the samples (Lab-ID Cbor_1015), which had been assembled into two overlapping *de novo* contigs. Remaining gaps between chloroplast contigs were closed by copying bases from the reference sequence. In order to get full genomic sequences and as a quality-check, reference-based mappings to the so-created pseudo-references were performed using the CLC *Map Reads to Reference* tool (length and similarity fraction set to 0.9) followed by a visual control of the mapping quality and running the CLC variant detection with a variant probability of 0.1. Remaining mapping errors were adjusted manually and as an additional quality control, another reference-guided assembly was performed using the bwa-mem algorithm (Li, 2013). Therefore, a multi-fasta reference file was created containing the respective chloroplast genome (containing only one copy of the inverted repeat region, namely the IRA) as well as eight *Cochlearia* mitochondrial *de novo* contigs (see 4.2.3.2 for mitochondrial genome assembly) and 5 published *Arabidopsis thaliana* chromosomes (NC_003070, NC_003071, NC_003074, NC_003075, NC_003076). Mitochondrial and nuclear references were included to avoid mismapping of reads originating from mitochondrial or nuclear pseudo-genes to the chloroplast reference. Adapter sequences and low quality bases were removed from the raw data using trimmomatic version 0.32 (Bolger et al., 2014) (with settings LEADING:20, TRAILING:20, SLIDINGWINDOW:4:15, MINLEN:50). Then, the BWA-MEM (Li et al., unpublished) as implemented in BWA version 0.7.8 (Li et al., 2009) was called with default setting for matching score (1), mismatch penalty (4), gap open penalty (6), gap extension penalty (1) and clipping penalty (5) and with the penalty for an unpaired read pair set to 15. The resulting bam files were further processed using SAMtools version 0.1.19 (Li et al., 2009; Li, 2011), thereby sorting the bam files, filtering them for the chloroplast reference, removing ambiguously mapped reads and putative PCR duplicates (*rmdup*). The Genome Analysis Toolkit (GATK; McKenna et al., 2010) was used for a local realignment around indels (using GATK *RealignerTargetCreator* and *IndelRealigner*) and an evaluation of per base mapping quality (using GATK *CallableLoci*). The GATK tools *UnifiedGenotyper* with the output mode set to EMIT_ALL_SITES as well as the *FastaAlternateReferenceMaker* were used to extract the consensus sequences and the resulting fasta files were finally masked based on the output of the

CallableLoci analysis and using the *maskfasta* application implemented in BEDtools version 2.19 (Quinlan and Hall, 2010; Quinlan, 2014).

Annotation of whole or nearly complete chloroplast genomes was done using the *Transfer Annotation* tool implemented in Geneious version 7.1.7 (Biomatters Ltd., Auckland, New Zealand). The annotated chloroplast genome of *Arabidopsis thaliana* (NC_000923) served as an initial reference and was aligned to the respective sample chloroplast genome in Geneious using MAFFT v7.017 (Katoh et al., 2002; Katoh et al., 2005; Katoh and Standley, 2013) under the FFT-NS-I x1000 algorithm with a 200PAM / k=2 scoring matrix, a gap open penalty of 1.53 and offset value of 0.123. Required similarity for annotation transfer was set to 65%. Transferred annotations were checked manually for completeness (presence of start and stop codons) and, if necessary, adjusted to preserve reading frames or to remove internal stop codons. Hereafter, remaining chloroplast samples were aligned to the most closely related and annotated *Cochlearia* chloroplast genome for annotation transfer.

Alignments and phylogenetic trees

The Geneious MAFFT plugin v7.017 was also used to generate an alignment of the 68 annotated and masked whole chloroplast genome sequences with the same settings as noted above. The resulting alignment had a length of 131,681 bp. As a first step, the alignment was split into blocks of exons, introns and intergenic spacers, resulting in a total of 264 blocks. These blocks were then subjected to an automatic alignment quality control using Gblocks v0.91b (Castresana, 2000) with a minimum block length of 5 and allowing for positions with a gap in up to 50% of the samples. Hereby, 2 small blocks were completely removed and in total 97.1% of positions were retained, resulting in an alignment length of 127,907 bp. From this alignment, 4 blocks (trnH, ndhK, rrn16S, rrn23S) were removed as they were suspected to be pseudo-genes in at least some of the samples. Also, three poorly aligned AT-rich regions (in the trnE-trnT intron, the rpl16 intron and the trnH-psbA intergenic region respectively) that had passed the Gblocks quality trimming were removed manually. After this, the final selection of 258 trimmed blocks summed up to an alignment length of 122,798 bp (Supplemental Data Set II-2). PartitionFinder v1.1.1 (Lanfear et al., 2012; Lanfear et al., 2014) with unlinked branch lengths was applied to search for the best partitioning schemes and substitution models of the respective sequence blocks, evolving at similar rates, which were then used for the construction of phylogenetic trees.

A ML tree search was performed in RAxML version 8.1.16 (Stamatakis, 2014) using the best partitioning schemes detected with PartitionFinder (Supplemental Data Set II-7), all assuming a GTR+Γ evolutionary model, and using the rapid bootstrap algorithm with 1000 replicates. Given the results from a tribal-wide analysis of Koch (2012), the three sequenced *Ionopsidium* samples served as an outgroup.

Divergence time estimation using BEAST

As summarized recently by Franzke et al. (2016), age estimation within the Brassicaceae family is a problematic task, given the sparse fossil record for the family. Estimates therefore largely depend on the chosen primary and/or secondary calibration points. To circumvent this obstacle, for divergence time estimation within Cochlearieae, five *Cochlearia* samples and one representative of the genus *Ionopsidium* were chosen from the dataset and included in a large-scale age estimation analysis using 73 conserved chloroplast genes (51 protein-coding genes, 19 tRNAs, three rRNAs) across the whole Superrosidae performed by Nora Hohmann (published in Hohmann et al., 2015). For details on taxon sampling and data analysis see Hohmann et al. (2015). This way, four fossil records/constrains could be placed along the tree for a more reliable dating analysis within Brassicaceae. These four selected primary calibration points, were based on reliable angiosperm fossil records that have also been used in other large-scale age estimation analyses within Angiosperms (e.g. Njuguna et al., 2013; Magallón et al., 2015). In detail, these were minimum ages of 48.4 mya for the *Prunus*/*Malus* split (Benedict et al., 2011), 84 mya for the split between *Castanea* and *Cucumis* (Sims et al., 1999), 65 mya for the *Mangifera*/*Citrus* split (Knobloch and Mai, 1986) and 88.2 mya for *Oenothera*/*Eucalyptus* (Takahashi et al., 1999). In accordance with Magallón et al. (2015), the root age was set to 92-125 mya (with a uniform distribution) and the debatable Brassicaceae fossil *Thlaspi primaerum* (discussed in Franzke et al., 2011; Franzke et al., 2016) was not included in the calibration.

The results of this foregoing analysis are summarized in Figure II-3. Based on the age estimates received herein, secondary calibration points were extracted for the generation of a temporal framework within the whole Cochlearieae dataset.

BEAST version 1.7.5 (Drummond et al., 2012) was used for divergence time estimation based on the whole chloroplast genome alignment (122,798 bp), partitioned into alignment blocks as described above, using the partitioning schemes received via PartitionFinder (Supplemental Data Set II-7) and choosing independent site and clock models among the two resulting partitions (both with substitution schemes GTR+Γ). The uncorrelated log normal relaxed-clock (UCLN) model with estimated rates was applied, this way accounting for varying rates among branches (Drummond et al., 2006). The chosen tree prior was Speciation: Birth-Death Incomplete Sampling (Stadler, 2009). Prior distributions for node ages, taken from the results in Hohmann et al. (2015), were defined as follows. The *Ionopsidium*/*Cochlearia* split was set to 10.81 mya and the *Cochlearia* crown age was set to 0.71 mya. Normal distributions were chosen for both secondary calibration points. The BEAST analysis was performed in two independent MCMC runs each with 100 million generations with samples taken every 10,000 generations. Effective sample size values (>200) were checked using Tracer version v1.6.0 (Rambaut et al., 2014) and trees were combined using LogCombiner version 1.5.5 (Drummond et al., 2012), discarding the first 50,000,000 generations as burn-in. Hereafter, treeAnnotator version 1.7.5 (Drummond et al., 2012) was used to create a maximum clade credibility tree from the 18,000 generated trees. Finally, the resulting tree was visualized using FigTree version 1.4.1 (Drummond et al., 2012).

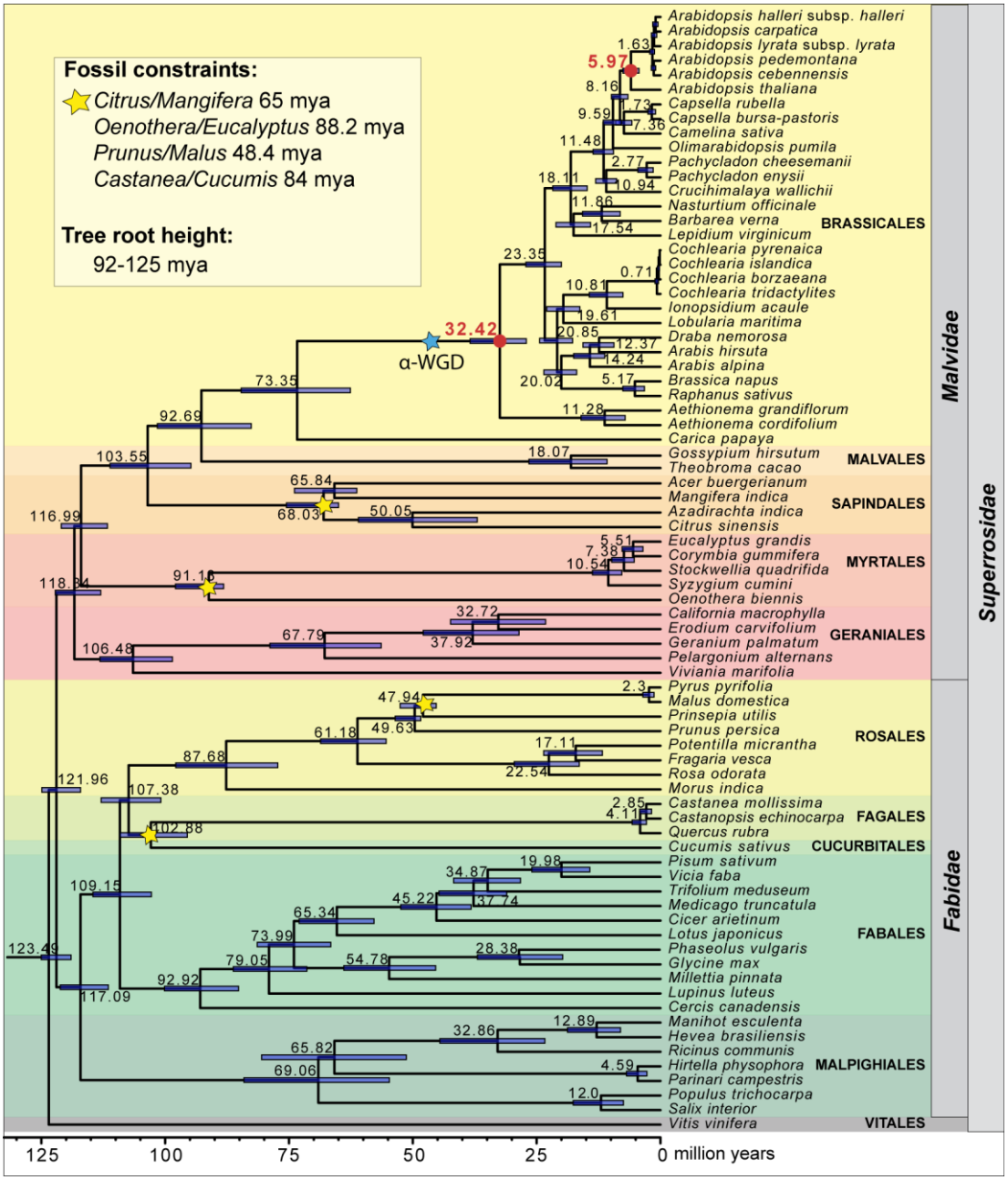

Figure II-3. Result of a BEAST analysis performed by N. Hohmann, based on 73 chloroplast genes from samples collected throughout the Superrosidae clade including six representatives of the tribe Cochlearieae. In accordance with Kagale et al. (2014), the maximum age of the Brassicaceae-specific At-α WGD (blue star) was set to 47 mya. Chosen fossil constraints are indicated with yellow stars and age estimates are given with 95% confidence intervals. Figure made by N. Hohmann, redrawn from Hohmann et al. (2015).

2.3.4 Mitochondrial genome data analysis

De novo assembly and reference-based mappings

Compared to the chloroplast genome, *de novo* assemblies of complete mitochondrial genomes are expected to be more challenging, due to the generally lower sequencing depth of mitochondrial genomes, combined with a high amount of repetitive regions which are difficult to assemble (Schatz et al., 2010; Straub et al., 2012). Therefore, following an approach presented by Straub et al. (2011), instead of assembling whole mitochondrial genomes *de novo* for every sample, the eight longest mitochondrial consensus sequences obtained from a single *de novo* assembly of one of the samples, namely Cmica_0979 (*C. micacea*; annotated mt contigs are given with Supplemental Data Set II-4), performed with the CLC de novo assembly algorithm (and using the same settings as described under 4.2.3.1.1 cp *de novo* assembly) served as a reference in reference-based mappings of the remaining samples. The chosen sample was selected based on a comparatively high mitochondrial genome coverage and accordingly long mitochondrial *de novo* contigs. These contigs were identified using BLASTn (with default settings) implemented in the NCBI nucleotide blast (browser version), and consensus sequences were retrieved using the CLC *Extract Consensus Sequence* tool with a minimum coverage threshold of 10x. The annotation (genes and coding sequences) of the eight Cmica_0979 consensus sequences was done using the Mitofy Webserver (Alverson et al., 2010) with default settings.

Both a *Cochlearia* (Cmica_0979) chloroplast genome and the published nuclear genome of *A. thaliana* (NC_003070, NC_003071, NC_003074, NC_003075, NC_003076) were included as references in the reference-based mappings in order to filter reads putatively originating from pseudogenes. Sequencing reads of each sample, trimmed using trimmomatic version 0.32 (for trimming details see 2.3.3 Chloroplast genome data analysis), were then mapped to the indexed mitochondrial pseudo-reference (including chloroplast and nuclear references) using the BWA-MEM algorithm with settings as specified for reference-based chloroplast genome mappings (under 2.3.3). Further processing of the resulting mapping files included the same steps as chosen for chloroplast control mappings (see 2.3.3), namely a removal of duplicate reads using SAMtools *rmdup* and a local realignment around indels using the GATK tools *RealignerTargetCreator* and *IndelRealigner*. Hereafter, GATK *CallableLoci* was used to detect high quality regions within the mitochondrial mappings by setting the minimum coverage to 20x and the minimum mapping quality to 30. Mitochondrial consensus sequences in fasta format were retrieved from the mapping files via the GATK tools *UnifiedGenotyper*, with the output mode set to EMIT_ALL_SITES and calling both SNPs and Indels, and *FastaAlternateReferenceMaker*. Finally, BEDtools *maskfasta* was used to mask those positions in the fasta files that had not passed the required quality filters.

Alignments and phylogenetic trees

The resulting masked mitochondrial consensus sequences in fasta format were aligned manually in PhyDE to account for the large amount of positions masked as 'N'. A total of 19 small inversions were reverse-complemented in PhyDE and hereafter, the eight contig alignments were concatenated in Geneious, resulting in a total alignment length of 307,602 bp. Out of these alignment positions, one region of suspiciously high sequence divergence (32 variants within 368

bp, position 203,076 – 203,442) was removed as a putative pseudo-gene, and about 7% of the alignment were removed in a check for alignment quality performed via Gblocks (with a removal of gap positions, the minimum number of sequences for a conserved position set to 31 and for a flanking position to 51, the minimum number of contiguous nonconserved positions set to 8 and a minimum block length of 10). From the resulting alignment with a length of 288,828 bp, regions that were masked in all samples were removed, so the final alignment had a length of 232,036 bp (Supplemental Data Set II-3).

The GTR+Γ+ I model that accounts for a proportion of invariant sites, was detected to be the best-fitting substitution model using jModelTest version 2.1.7 (Darriba et al., 2012; Posada, 2008). RAxML version 8.1.16 (Stamatakis, 2014) was used to produce a ML tree with 1000 rapid bootstrap replicates under the GTR+Γ+ I model.

2.3.5 Nuclear genome data analysis

Reference-based mappings and SNP calling

In order to analyze the nuclear genome, reference-based mappings of reads trimmed via trimmomatic version 0.32 (for trimming details see 2.3.3 Chloroplast genome data analysis) were performed using a previously assembled transcriptome of *Cochlearia pyrenaica* (transcriptome assembly and annotation generated by Dr. Lua Lopez (Lopez et al., submitted) as a reference. After indexing the transcriptome reference using the BWA *index* tool, the *CreateSequenceDictionary.jar* application embedded in Picard tools v. 1.91 (http://broadinstitute.github.io/picard) and the samtools *faidx* command, the BWA-MEM algorithm was applied for read mapping (settings as described under 2.3.3 Chloroplast genome data analysis) and SAMtools version 0.1.19 was used to remove ambiguously mapped reads and putative PCR duplicates. Alignment quality was further controlled and improved by performing a local realignment around indels with the GATK tools *RealignerTargetCreator* and *IndelRealigner* with default settings. Hereafter, sequencing depth was determined for each position using the *DepthOfCoverage* tool implemented in GATK (including reference sites with Ns, omitting intervals and locus table). The upper 2% cutoff coverage for every sample was retrieved using own python scripts and the highest respective sample coverage, when excluding the upper 2% sites, was then taken as upper limit for a GATK *CallableLoci* run, thereby removing organellar and rDNA transcripts. Minimum coverage for a callable site was set to 4x, minimum base quality to 25 and minimum mapping quality to 10. Transcriptome-wide nuclear SNP callings (coding and non-coding) were performed on different subsets of the complete data set. Sampling details of the different subsets are given in Supplemental Data Set II-6. For each of the subsets (‘all samples’, ‘all diploids’, ‘European diploids’, ‘*C. bavarica* hybrid origin’) *multiIntersectBed*, as implemented in BEDtools version 2.19 was used to combine the callable sites of all samples included in the respective analysis and the following SNP calling was restricted to those sites that passed quality requirements in all samples. Variant detections were performed using GATK’s *UnifiedGenotyper* with the sample ploidy settings adjusted to the respective dataset. For the SNP calling on the whole dataset, all samples were treated as diploids since tetra-, hexa-, or octoploid SNP callings resulted in a slightly reduced power of calling heterozygous sites compared

to diploid SNP callings for the same set of samples. This rather conservative approach, despite the putatively non-exhaustive SNP calling in polyploid samples, produced very robust phylogenetic signals over different analyses. A clustering of individuals of the same ploidy level could not be observed and alternative SNP callings treating all samples (including the diploids) as tetraploids didn't change the final outcome (data not shown). An overview of the different SNP calling runs, with respective numbers on combined callable sites, is given in Table II-3.

Genetic assignment tests using STRUCTURE

The Bayesian clustering program STRUCTURE version 2.3.4 (Pritchard et al., 2000; Falush et al., 2003) was used to perform genetic assignment tests and infer genetic structuring within the dataset. Based on the vcf files resulting from the different SNP callings, STRUCTURE input files were created for different subsets of samples. An overview of the different STRUCTURE runs is given in Table II-3 and respective STRUCTURE input files are given with Supplemental Data Set II-8.

Table II-3. List of STRUCTURE runs performed on different sample subsets together with information on the respective numbers of intersected callable bp and variants found within. All given subsets were analyzed twice, using both *correlated* and *independent allele frequencies*.

	Dataset	Ind.-no.	Callable Sites	Variant-no.	Calling ploidy
1)	all samples	62	1,425,819	101,386	diploid
2)	all diploids	38	2,829,028	119,629	diploid
3)	European diploids	22	9,654,391	310,202	diploid
4)	*C. bavarica* hybrid origin	8	3,341,378	118,488	hexaploid

In order to detect and select the optimal *K* (the number of populations/clusters), STRUCTURE runs were performed for *K* = 1 to *K* = 12 (all samples, diploids only, diploids and tetraploids) or *K* = 1 to *K* = 6, depending on the dataset, with a burn-in of 5,000 MCMC (Markov Chain Monte Carlo) iterations followed by a run for 5,000 iterations. In order to check for consistency, each K level was replicated ten times. The datasets were analyzed under the *admixture model* and each STRUCTURE analysis was performed twice, to test both of the two available allele frequency models – *correlated* and *independent allele frequencies*. The *correlated allele frequencies* option as described by Falush et al. (2003) is used to detect discrete population structure between closely related populations based on a correlation of allele frequencies, theoretically without affecting the result if no such correlation exists (Porras-Hurtado et al., 2013). Yet, according to Pritchard et al. (2010) there is a small risk of overestimating *K* under this model, which is why both models were tested and compared here. The R script *structure-sum* (Ehrich, 2006) was used to select the optimal *K* for each analysis. Here, the likelihood of every analyzed *K* is determined by considering the respective likelihood values of each run (Rosenberg et al., 2002), the similarity coefficient between the runs as well as values of delta *K* (Evanno et al., 2005). Finally, the optimal *K* of each analysis was selected according to the Evanno method. Only *K*s excluding empty groups were finally considered. The result files gained from the different runs were further processed using the Python program *structureHarvester.py* version 0.6.94 (Earl, 2012) and replicate runs of the optimal *K* were summarized using CLUMPP version 1.1 (CLUster Matching and Permutation Program; Jakobsson and Rosenberg, 2007).

For the analysis of the hexaploid *C. bavarica* (2n = 6x = 36), that was suggested to be of allopolyploid origin, existing hypotheses were taken into account for the composition of the separate data set (Heubl and Vogt, 1985; Koch et al., 1996; Koch et al., 1998). Hence, *C. bavarica* was analyzed in a data set composed of two *C. bavarica* samples, and three samples from both of the two putative parental taxa, the diploid *C. pyrenaica* (2n = 2x = 12) and the tetraploid *C. officinalis* (2n = 4x = 24).

Phylogenetic trees

Furthermore, for the diploid data set, consisting of 38 *Cochlearia* samples, a tree based approach was used to analyze the phylogenetic relationships. Again, this was based on the output of the respective variant callings. To get a better idea of the phylogenetic relationship with the sister genus, trees were produced for *Cochlearia* only, as well as including one of the *Ionopsidium* samples (Imega_1776), treated as a diploid, in the analysis. The *Ionopsidium* sample was selected based on sequence and alignment quality in order to maximize the overlap of high quality sequence information among all samples.

Input files for the tree building software were created from the respective transcriptome mapping files generated before and described under 2.3.5 (Nuclear genome data analysis - Reference-based mappings). The output files of the GATK *CallableLoci* analysis, with a minimum coverage of 4x and an upper 2% coverage cutoff, were reused and BEDtools (version 2.19) application *multiIntersectBed* was utilized to identify shared callable sites between the selected sample subsets. Again, variant calling was performed in callable sites via the GATK *UnifiedGenotyper* tool with sample ploidy set to diploid for both data sets. The resulting vcf files were converted to fasta files by using the VCFtools (v.0.1.11) *vcf-to-tab* function (Danecek et al., 2011), followed by the Perl script *vcf-tab-to-fasta-alignment.pl* (Bergey, 2012). Singletons were removed before the analysis using the VCFtools application *--singletons*, followed by VCFtools *--remove/exclude-positions* based on the list of singleton positions. This way, putative sequencing errors were also further reduced. JModelTest version 2.1.10 was used to select the best fitting models of nucleotide substitution for subsequent ML analyses according to Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC).

Maximum likelihood phylogenetic analyses for both datasets were conducted using RAxML version 8.1.16 (Stamatakis, 2014). A rapid bootstrap search with 1,000 replicates and with a final ML search was performed under the GTR+Γ substitution model (the general time reversible model with a discrete gamma distribution) for both datasets. RAxML input files are given with Supplemental Data Sets II-9, II-10. FigTree version 1.4.1 (Drummond et al., 2012) was used to visualize the best output tree.

SplitsTree analysis

In addition to the phylogenetic trees of diploid samples, phylogenetic networks of all samples were produced using SplitsTree version 4.14.2 (Huson, 1998; Huson and Bryant, 2006) since this approach is often better suited than a unique tree approach to detect conflicting or reticulate phylogenetic patterns.

The analysis was performed with all *Cochlearia* samples that have been selected for nuclear data analysis. SplitsTree input files were based on regions of transcriptome mappings that had passed the GATK search for *CallableLoci* with a minimum coverage of 4x and an upper 2% coverage cutoff. Again, BEDtools *multiIntersectBed* was used to select shared callable sites between all samples and GATK's *UnifiedGenotyper*, with a minimum phred-scaled confidence threshold of 30, was used for SNP calling in the selected regions and treating all samples as diploids, since higher ploidy levels had turned out to filter ambiguous sites more strictly in polyploids than diploids, which was putatively due to similar coverage values over all samples irrespectively of a sample's ploidy level. Since multiallelic sites can be indicative of noisy alignment regions, the *SelectVariants* tool implemented in GATK was used to select only biallelic sites from the generated variant calls. Yet, both datasets, 'all variants' and 'biallelic only', were used for SplitsTree analyses. The different variant calls were then converted into fasta files via the GATK *FastaAlternateReferenceMaker* using IUPAC ambiguity codes for heterozygous SNPs. Split decompositions for all datasets were calculated and plotted using SplitsTree version 4.14.2 (SplitsTree input files are given with Supplemental Data Set II-11).

2.4 Results

2.4.1 Illumina sequencing data

Raw sequencing reads for all samples, generated on an Illumina HiSeq 2000 system, have been deposited in the European Nucleotide Archive (ENA) under ENA accession numbers ERS1803485 to ERS1803552. Sample specific ENA accession numbers are given with Supplemental Data Set II-1.

The read number generated per sample from multiplexed Illumina HiSeq 2000 sequencing ranged from 21,864,610 (for sample Cpyr_0699) to 73,738,692 (for sample Cses_1286), thus showing comparatively strong variation putatively resulting from both variation in DNA quality and especially the multiplexing step. Detailed information on the sequencing outcome are documented in Supplemental Data Set II-1.

Via quality and adapter trimming of raw sequencing reads, on average 90.1% of the reads were retained, ranging from 69.5% to 96.7%. Hence, overall quality of sequencing data was good for most of the samples with only few samples showing slightly reduced sequencing read quality putatively due to highly fragmented DNA extractions. Yet, no clear connection between the age of leaf material and the sequencing outcome could be observed with the two oldest samples (Coff_1364, 1957; Coff_1365, 1954) yielding average read counts before and after the trimming step (Supplemental Data Set II-1).

2.4.2 Chloroplast dataset

Chloroplast genome assembly and alignments

Whole or nearly complete chloroplast genomes were successfully assembled and annotated for all of the 65 ingroup and the 3 outgroup samples included in the study. Annotated whole chloroplast genome sequences have been deposited under Genbank accession numbers LT629868 - LT629930 and LN866844 - LN866848 (the latter were first published in Hohmann et al., 2015).

As expected, all chloroplast genomes showed a highly conserved gene content and structure with lengths ranging from 154,000 bp (Iacau_1072) to 155,353 bp (Imega_1776) over the whole dataset and from 154,518 bp (Cisla_1765) to 154,775 bp (Ctat_1306) within *Cochlearia*. The average read depth of chloroplast assemblies ranged from 22x (Iabu_1074) to 9,126x (Cbor_1015) with a mean of 2,093x. Approximately 0.07% (Iabu_1074) to 32.19% (Cbor_1015), on average 9.57%, of all trimmed reads mapped to the respective chloroplast genome reference. As expected, samples with prior chloroplast enrichment had a chloroplast sequencing depth higher than average and accordingly showed a reduced nuclear and especially mitochondrial read depth. Detailed information on per sample chloroplast length, read no., read depth, Genbank accession numbers, et cetera is given in Supplemental Data Set II-1.

Within the final chloroplast alignment (122,798 bp; Supplemental Data Set II-2), a total of 5,292 SNPs were detected, thereof 1,003 within the genus *Cochlearia*, the remaining 4,289 SNPs separating the two sister genera. 2,118 of the SNPs within the full alignment were parsimony

informative, thereof 569 within *Cochlearia*. Indels were not taken into account for phylogenetic analyses.

Phylogenetic analyses

The phylogenetic tree based on (nearly) complete chloroplast genome sequences and produced via maximum likelihood analysis using RAxML had a good overall bootstrap support (Figure II-4). Figure II-4 shows the ML tree with the *Ionopsidium* outgroup being collapsed and condensed for a better representation of phylogenetic relationships within *Cochlearia* (the original tree is given in Supplementary Data Set II-12). Six major phylogenetic lineages were revealed and are represented by colored bars which will be referred to in the following. Resolution of the ML tree reconstruction is generally high, yet several unresolved polytomies remain, especially along basal nodes of the green lineage.

Pink and green lineages (Figures II-4 and II-5) show a wider geographical distribution range whereas the remaining phylogenetic lineages have a more restricted distribution. The most basal lineage detected based on cp data is the yellow lineage containing only two samples of *C. tridactylites* (Ctri_1288, Ctri_1371) from the Canadian east coast of the North American mainland. The second most basal lineage, the pink lineage shows a much wider distribution with a first deep split separating a western Spanish *C. danica* (2n=6x) from Romanian *C. borzaeana* (2n=8x) and an arctic group of *C. sessilifolia* (2n=2x) and *C. groenlandica* (2n=2x) from Alaska and British Columbia respectively. The blue lineage contains coastal taxa of different ploidy levels, distributed along the coast of Western France up till Northern Norway. The inland sample Cdan_1766 is a sample of *C. danica* found along a federal highway of Southern Germany. The green lineage contains most of the Central European inland samples but also three arctic subgroups including the arctic species *C. groenlandica*, *C. tridactylites* and Icelandic *C. islandica*. The purple lineage contains diploid samples from Northern Spain, most samples from the UK and few Central European samples.

Interestingly, the basal yellow, pink and orange lineages almost exclusively contain taxa with a base chromosome number of n=7. Moreover, the blue coastal lineage also contains two samples with n=7, and several basal lineages within the green lineage are composed of n=7 taxa as well.

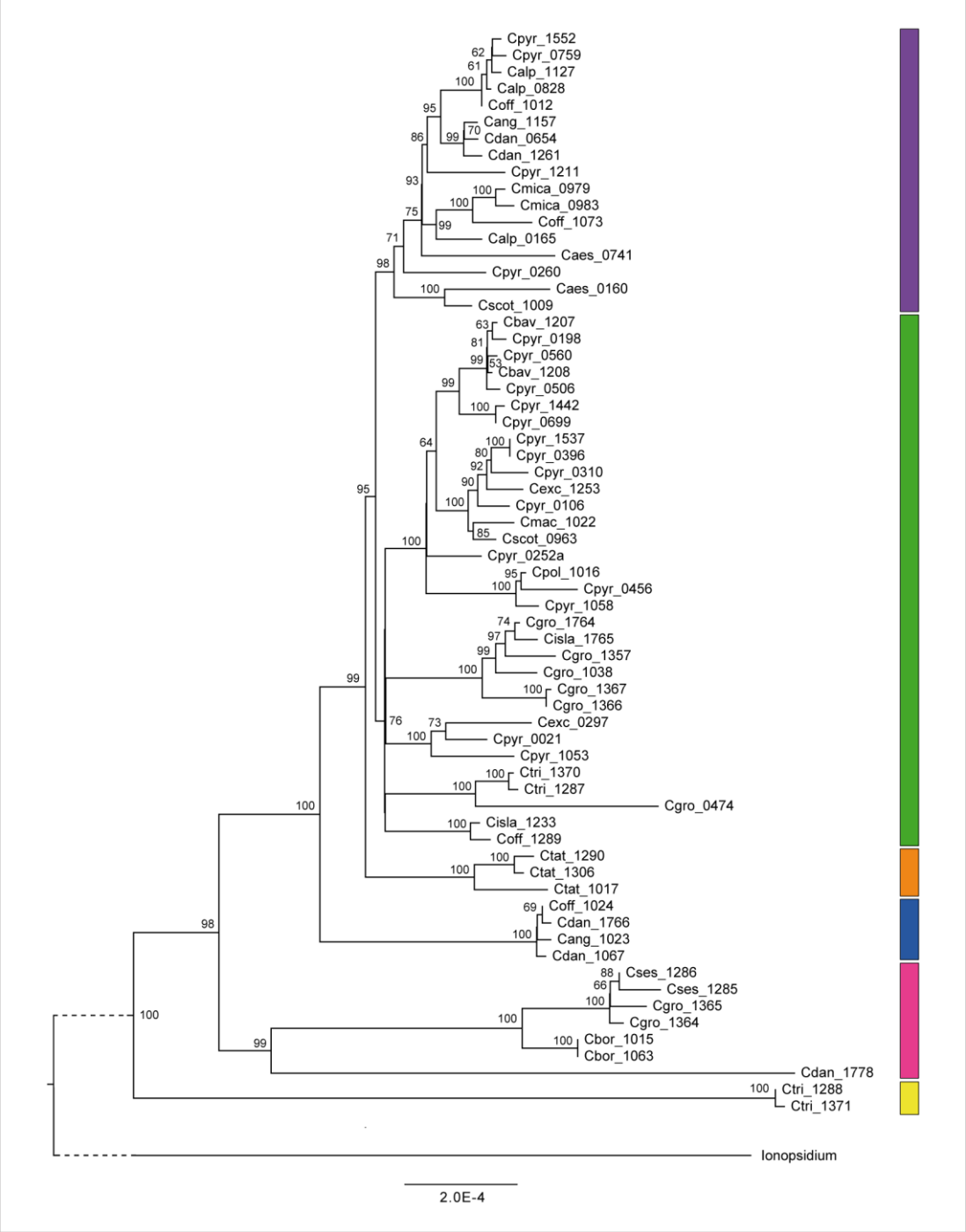

Figure II-4. Maximum likelihood phylogram based on complete chloroplast genomes from the tribe Cochlearieae produced with RAxML. Bootstrap support (1000 replicates) above 50% is shown near the respective nodes. For a better illustration of *Cochlearia* samples, the outgroup lineage of *Ionopsidium* is collapsed and condensed (full tree is given with Supplemental Data Set II-12). To the right, six evolutionary lineages within *Cochlearia* are displayed as colored bars.

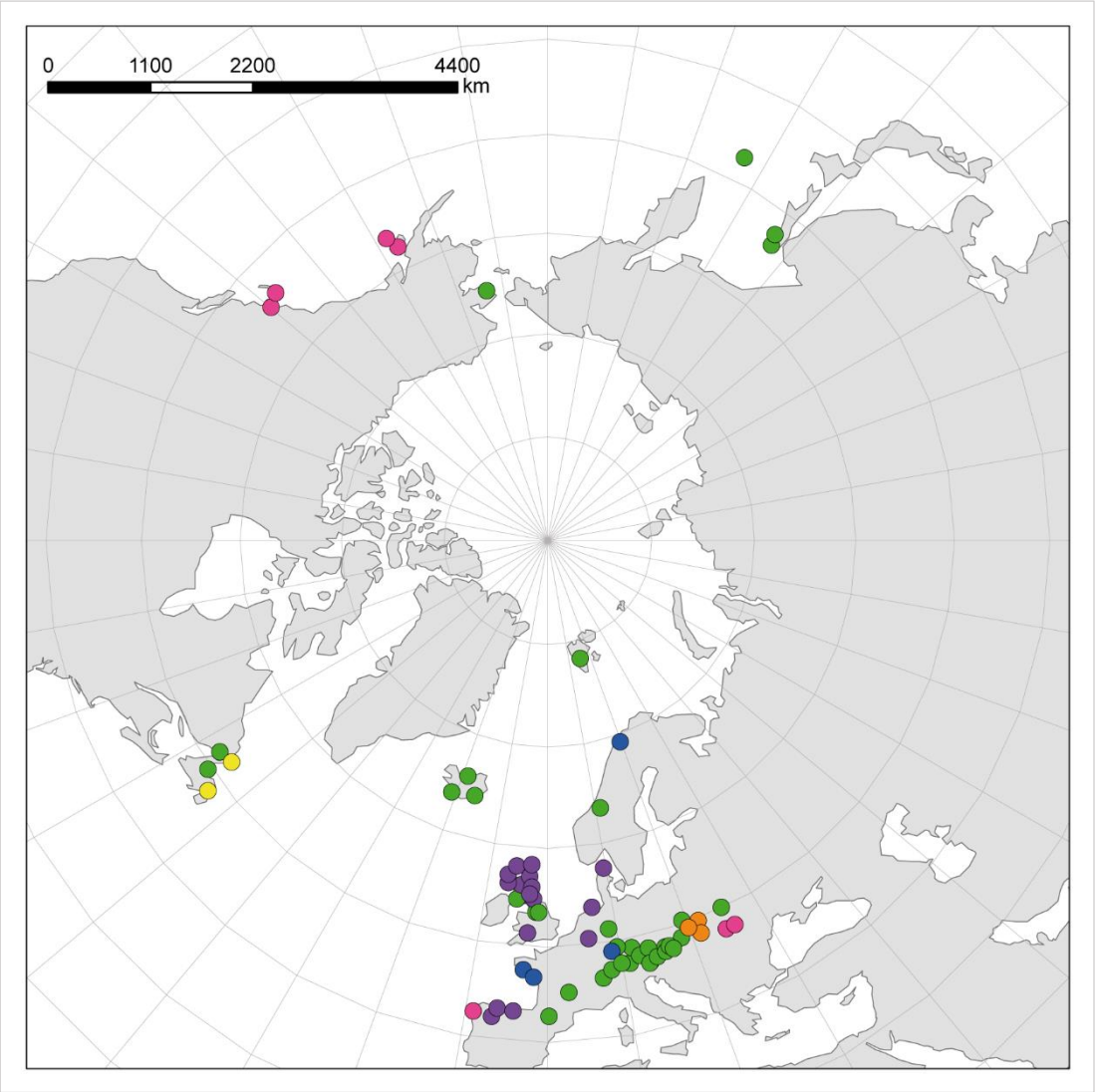

Figure II-5. Geographical distribution pattern of the six *Cochlearia* chloroplast lineages as revealed from maximum likelihood analysis (Figure II-4). Colors are corresponding to those given in Figure II-4.

Figure II-6 shows in more detail that both the purple and the green chloroplast lineage are connected to Great Britain, indicating a repeated colonization from continental Europe. On the other hand, several samples from continental Europe can be found within the British subgroups of the purple lineage, indicating dispersal from the UK back to Central Europe. Both *C. aestuaria* samples from Northern Spain (Caes_0160, Caes_0741) and a *C. pyrenaica* sample (Cpyr_0260) from Belgium, thus coastal and inland diploid taxa, occur at basal positions of the purple lineage.

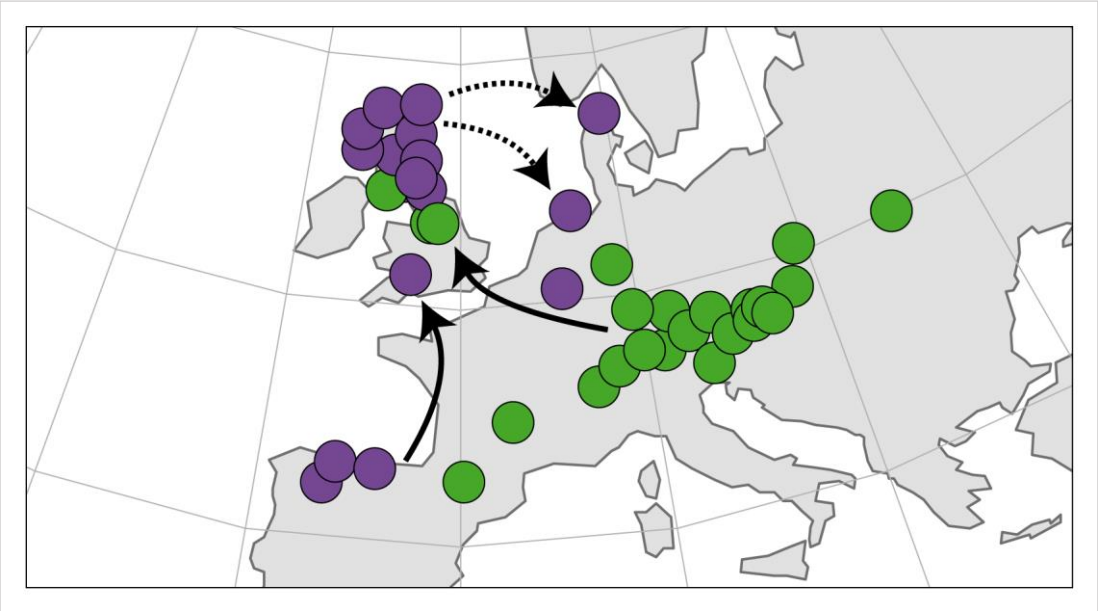

Figure II-6. Putative colonization routes (arrows with solid lines) into the UK from Northern Spain (purple lineage, Figure II-4) and Central Europe (green lineage, Figure II-4) as revealed from maximum likelihood analysis of whole chloroplast genomes. Putative repeated dispersal from the UK back to continental Europe is indicated with dashed lines.

The two samples of *C. bavarica* (Cbav_1207, Cbav_1208), representing the two distribution areas of this hexaploid species, do not form a monophyletic group but show closer relationships with neighboring populations of *C. pyrenaica* (Cpyr_0198, Cpyr_0560; Figure II-7).

Likewise, monophyly was not supported for the two samples of *C. excelsa*, collected from the two distribution areas of this species in the Gurktal Alps (Cexc_0297) and the Seckau Tauern (Cexc_1253). Whereas Cexc_0297 was grouped together with two populations of *C. pyrenaica* from the French Pyrenees (Cpyr_1053) and the Massif Central (Cpyr_0021), Cexc_1253 was found within a phylogenetic group of Swiss and Austrian samples of *C. pyrenaica*.

Several polyploid taxa were also not arranged into monophyletic groups as shown e.g. for *C. alpina* which occurs in two different subgroups of the purple lineage. These groups also contain samples of tetraploid *C. officinalis* and *C. micacea* as well as diploid *C. pyrenaica*.

Also, none of the coastal taxa shows monophyly within the chloroplast phylogeny. The hexaploid *C. danica* for example occurs in four out of the six bigger chloroplast lineages, the tetraploid *C. officinalis* in three of them and octoploid *C. anglica* as well as tetraploid *C. scotica* are found in two lineages. A rather peculiar result was the placement of Cscot_0963, a sample of *C. scotica* from Ballachulish in Scotland, at the base of a subgroup from the green lineage, which apart from that only includes diploid samples from Swiss and Austrian *C. pyrenaica*, *C. macrorrhiza* and *C. excelsa*.

The hexaploid *C. polonica* is most closely related to geographically close *C. pyrenaica* samples Cpyr_0456 and Cpyr_1058 from Slovakia and the Ukraine respectively, whereas *C. tatrae* the other eastern hexaploid inland species forms its own monophyletic lineage.

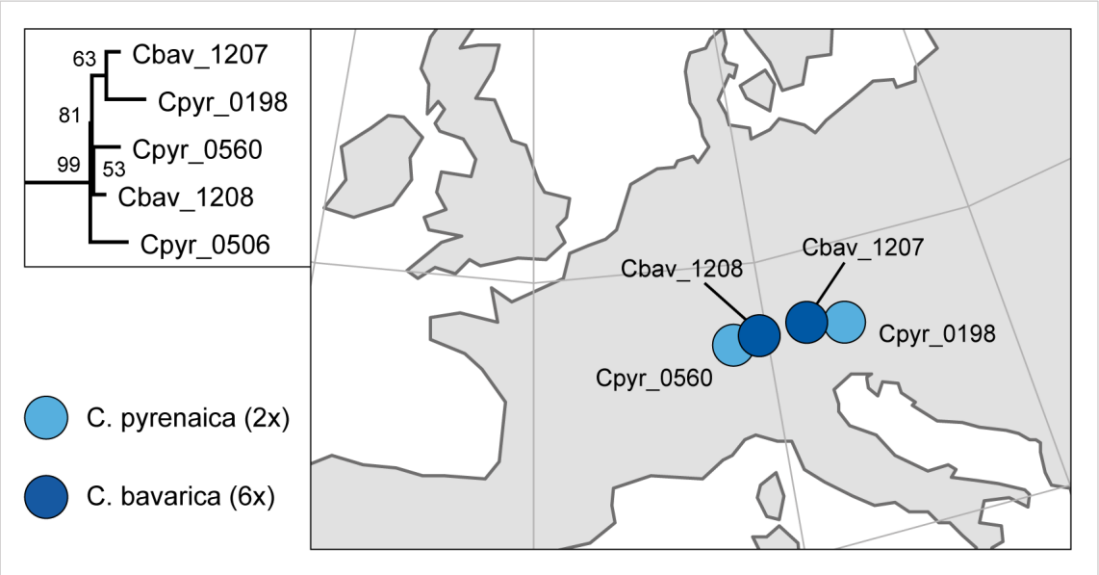

Figure II-7. Phylogenetic placement of hexaploid *C. bavarica* samples with geographically close populations of *C. pyrenaica* as putative maternal plants in past allopolyploidization processes as revealed from the maximum likelihood chloroplast phylogeny (Figure II-4).

Divergence time estimates

The tree topology obtained via BEAST analysis of whole chloroplast genome data (Figures II-8 and II-9) based on secondary calibration points is nearly identical to the generated ML phylogeny (Figure II-4). Only some subgroups along the basal polytomy of the green lineage in the ML tree are swapped in the strictly bifurcating BEAST tree.

The revealed estimates of divergence times indicate that the split between *Cochlearia* and *Ionopsidium* dates back ~9.25 mya and radiation within the tribe *Cochlearia* started at ~0.66 mya (Figure II-8). These estimates are well in agreement with the ones provided by Hohmann et al. (2015) as well as with earlier estimates for the *Cochlearia*/*Ionopsidium* split by Koch (2012). The detected mean rate (ucld.mean) of evolutionary change estimated under the birth-death model was 1.46×10^{-9} substitutions/site/year when averaged over partition 1 (ucld.mean: 0.65×10^{-9} over 62.1% (76,263 bp) of the total alignment) and partition 2 (ucld.mean: 2.79×10^{-9} over 37.9% (46,535 bp) of the total alignment). This result is comparable to previously published estimates of chloroplast sequence evolution, like the rate of 1.35×10^{-9} estimated for chloroplast genomes of *Fragaria* (Wolfe et al., 1987; Njuguna et al., 2013).

All major phylogenetic lineages originated before the Würm glacial, the last glaciation in the European Alps at about 30-80 kya and split times within the genus correlate with high glacial periods (Figure II-9) as revealed e.g. from Antarctic ice cores such as Vostok (Petit et al., 1999) and Dome C (Augustin et al., 2004). The deepest split within *Cochlearia*, separating the yellow arctic lineage, dates to 659 kya, close to the end of the Günz glacial at about 640 kya. Hereafter, divergence of the pink chloroplast lineage, including arctic taxa, *C. borzaeana* and Spanish *C. danica*, seems to have taken place about the beginning Mindel glacial around 450 kya. Near the end of Mindel glacial, the next phylogenetic lineage, the blue coastal lineage separated and further divergence within this lineage is of very recent origin. Further diversification within the widely

distributed green lineage and the mainly Western European purple lineage mainly occurred during Riss (150-200 kya) and Würm glaciations (30-80 kya).

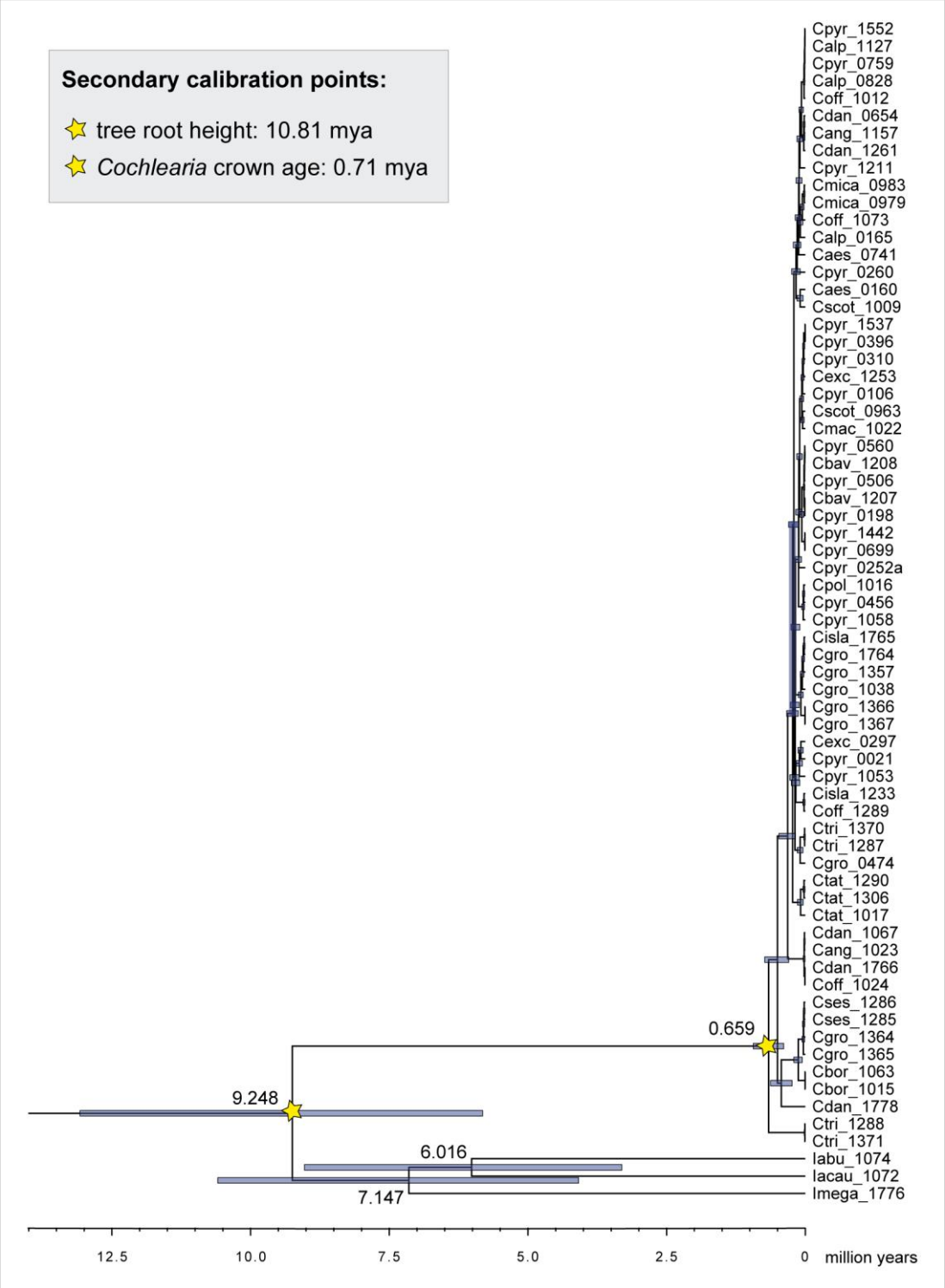

Figure II-8. Cochlearieae chronogram based on whole chloroplast genome sequence data and inferred using BEAST. Secondary calibration points are indicated accordingly. Divergence times for the *Ionopsidium*/*Cochlearia* split, the genus *Ionopsidium* and the *Cochlearia* crown age are indicated with 95% confidence intervals. Divergence times within the genus Cochlearia are given in Figure II-9.

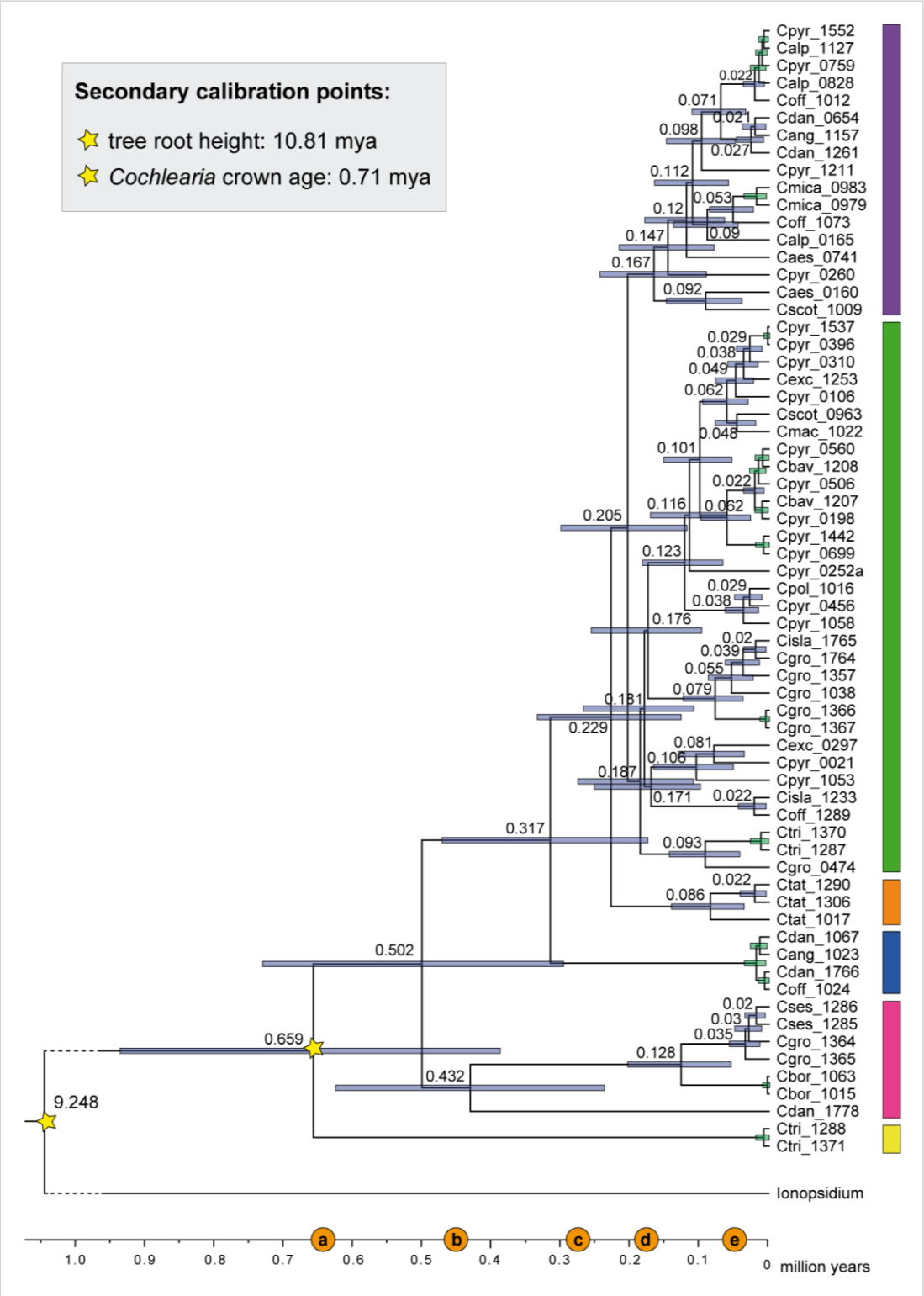

Figure II-9. BEAST chronogram based on whole chloroplast genome sequence data from the tribe Cochlearieae. For a better illustration of *Cochlearia* samples, the outgroup lineage of *Ionopsidium* is collapsed and condensed. Secondary calibration points are indicated accordingly and divergence times older than 0.02 mya age are indicated with 95% confidence intervals. Divergence times younger than 0.02 mya are indicated by green node bars. The six major lineages within *Cochlearia* are displayed as colored bars (corresponding to the ML phylogeny generated with RAxML, Figure II-4). Letters displayed on the timeline indicate high glacial periods as follows: a) 640 ky, end of Günz glacial; b) 450 ky, begin of Mindel glacial; c) 250-300 ky, Mindel-Riss inter-glacial; d) 150-200 ky, Riss glacial; e) 30-80 ky, Würm glacial.

2.4.3 Mitochondrial dataset

Mitochondrial contig assembly and alignments

Table II-4 lists the eight longest mitochondrial contigs generated for sample Cmica_0979 (*C. micacea*) and hereafter serving as a partial mitochondrial genome reference (Supplemental Data Set II-4), spanning a total of 307,510 bp and containing 32 protein-coding genes (most of them complete) as well as 15 tRNA sequences representing 13 different tRNAs. Considering the length as well as the gene content of other published mitochondrial genomes within the Brassicaceae (e.g. *Arabidopsis thaliana* (NC_001284, 366,924 bp, 35 protein-coding genes; Unseld et al., 1997), *Brassica oleracea* (NC_016118, 360,271 bp, 56 protein-coding genes; Chang et al., 2011), *Brassica napus* (NC_008285, 221,853 bp, 34 protein-coding genes; Handa, 2003)) these contigs seem to represent a big part of the complete mitochondrial genome.

Table II-4. Eight mitochondrial *de novo* consensus sequences (*C. micacea*, Cmica_0979) chosen as reference sequences for mitochondrial genome mappings with information on contig length and mitochondrial gene content.

contig name	contig length	annotated mt gene content
contig 3	90,418 bp	ccmB, ccmFC, cox3, matR, orf161, nad1 (exon 4, 5), nad5, nad7, orfB (atp8), atp1, orf25 (atp4), nad4L; tRNA-Gly, tRNA-Tyr, tRNA-fMet, tRNA-Asp, tRNA-Cys, tRNA-Pro, tRNA-Glu
contig 4	68,240 bp	rps3 (exon 1), ccmC, cox2, cob, rpl5, rpl16, cox1, atp6; tRNA-Ser, tRNA-Tyr, tRNA-Lys
contig 8	24,860 bp	ccmFN1, nad6; tRNA-Trp
contig 11	11,520 bp	nad1 (exon 1)
contig 18	44,257 bp	ccb203 (partial), nad5 (exon 3), nad2 (exon 3, 4, 5), nad9; tRNA-Gln, tRNA-Met
contig 33	11,474 bp	atp9, nad1 (exon 2, 3); tRNA-fMet, tRNA-Ile
contig 45	46,512 bp	nad2 (exon 1, 2), rps4, rps7, rps12 (partial), rps3 (exon 1), rpl2, orfX (partial), nad4, orf116
contig 199	10,229 bp	nad5 (exon 4, 5)

In contrast to the chloroplast genomes, mitochondrial genomes were found to be structurally much less conserved, just as one could expect from what is known on plant mitochondrial evolution (Burger et al., 2003; Knoop, 2004; Knoop et al., 2011; Steele and Pires, 2011; Straub et al., 2012). Also, due to lower copy numbers per cell, the sequencing depth of mitochondrial genomes was generally much lower, ranging from 4.44x (Ctat_1017) to 484.64x (Cpyr_0560) with about 0.04% (Cbor_1015) to 3.28% (Cpyr_0106) of the trimmed reads mapping to the mitochondrial genome reference contigs. Detailed per-sample information on coverage, callable sites and percentage of masked regions is given in Supplemental Data Set II-1. Nine samples (Caes_0160, Cbor_1015, Cexc_0297, Cgro_0474, Cmac_1022, Coff_1289, Ctat_1017, Iabu_1072 and Iacau_1072) were finally removed from phylogenetic analyses since more than 50% of the total 307,510 bp were masked. Four of these samples, namely Cbor_1015, Cexc_0297, Cmac_1022 and Ctat_1017, had been enriched for chloroplasts before sequencing, which explains the very low mitochondrial sequencing depth in these cases. For the *Ionopsidium* samples, divergence from the *Cochlearia*

reference sequences putatively caused additional problems during reference-based mappings. Imega_1776 (~57% masked) also failed the 50% filter but was kept in order to have an outgroup sample for the phylogenetic analyses. Given a total of ~8.4% masked bases in the final alignment (Supplemental Data Set II-3) with a length of 232,036 bp (after excluding pure 'N' positions) including 58 *Cochlearia* samples and one *Ionopsidium* sample, the portion of positions masked due to quality constraints was comparatively high. Yet, a total of 1,417 high quality SNPs, thereof 306 parsimony informative, were detected in these remaining alignment positions. Out of these SNPs, 854 were found between *Cochlearia* and *Ionopsidium*, the remaining 563 within *Cochlearia*.

Phylogenetic analyses and cytoplasmic (in-)congruence

Bootstrap support of the mitochondrial ML phylogeny (Figure II-10) generated with RAxML was high for several nodes but overall support and resolution were much lower compared to the ML tree based on whole chloroplast genomes (Figure II-4; the original mitochondrial tree is given with Supplemental Data Set II-13). The two organellar phylogenies show some incongruences at first sight as several of the bigger lineages detected for the chloroplast phylogeny are split up in the mitochondrial ML tree.

Especially the green chloroplast lineage, the biggest lineage detected within the chloroplast phylogeny, was split up into five smaller lineages, namely a lineage containing only *C. tridactylites*, one French Pyrenees/Massif Central lineage, an arctic lineage with *C. islandica* and *C. groenlandica* samples and a Central European lineage. Another single sample from the green chloroplast lineage, a *C. islandica* sample (Cisl_1233), was placed basal to a mitochondrial subgroup formed by the orange, blue and pink chloroplast lineages. However, almost all basal nodes within the mitochondrial phylogeny have very low bootstrap support (<50%) indicating that the resolution at deeper nodes is simply not high enough for a clear grouping of lineages. The phylogenetic relationships within groups/lineages are often better supported and reveal almost exactly the same picture as obtained in the chloroplast phylogeny.

Some of the samples/lineages show notably long branches. Whilst this might be explained by correspondingly long branches (old age) in the chloroplast phylogeny for members of the split up pink lineage, it should rather be interpreted as a result of low coverage of the mitochondrial genome and/or low mapping quality in other samples, e.g. Cdan_0654.

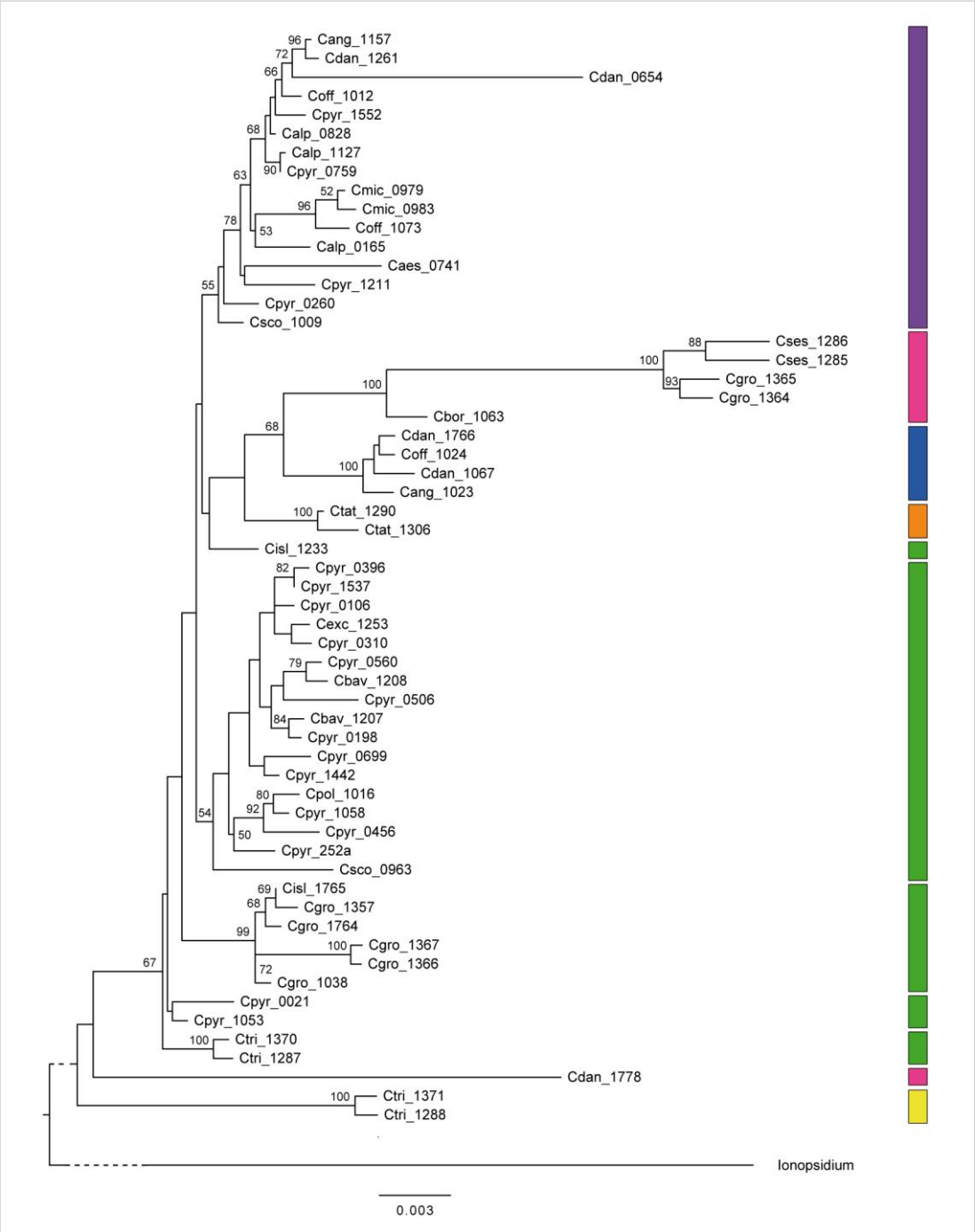

Figure II-10. Maximum likelihood tree based on 58 partial mitochondrial genomes of genus *Cochlearia* and one sample of the genus Ionopsidium (Imega_1776) set as outgroup. For a better illustration of *Cochlearia* samples, the outgroup lineage of *Ionopsidium* is condensed (original tree is given with Supplemental Data Set II-13). Bootstrap support (1000 replicates) above 50% is shown near the respective nodes. Colored blocks to the right correspond to respective chloroplast lineages as displayed in Figure II-4.

2.4.4 Nuclear dataset

Reference-based mappings

Within *Cochlearia* samples, the percentage of reads that mapped to the transcriptome reference ranged from 5.2% (Caes_0160) to 40.0% (Cmica_0979), with an average of 32.1%. For *Ionopsidium*, the percentage of mapped reads was generally lower, with 2.8% (Iabu_1074), 18.4% (Iacau_1072) and 25.1% (Imega_1776) of reads that mapped to the transcriptome. This is partly explained by higher sequence divergence between the *Ionopsidium* samples and the *Cochlearia* transcriptome reference, but also due to a generally low DNA and sequence quality at least in the case of sample Iabu_1074.

As expected, the overall coverage in the transcriptome mappings was comparatively low. Yet, out of the full transcriptome length of 58,236,171 bp, the amount of callable regions with a minimum coverage of 4x ranged between 722,503 bp (Caes_0160) and 28,604,135 bp (Cang_1023). Three *Cochlearia* samples, namely Caes_0160, Ctat_1017 and Coff_1289, were finally excluded from the nc genome analyses due to low amounts of callable regions. Likewise, only Imega_1776 was included in the respective analyses whereas Iabu_1074 and Iacau_1072 were removed.

Genetic assignment tests using STRUCTURE

The STRUCTURE (Pritchard et al., 2000) analyses based on different variant calls from nuclear transcriptome mapping data (both coding and non-coding) revealed new insights into different aspects of the evolutionary history of the genus *Cochlearia*. The informative value of the genetic assignment tests was increased by the subdivision of the dataset into several subsets.

The respective results of deltaK analyses are given with Supplementary Data Set II-16.

1) All samples

For the STRUCTURE analysis of all samples, based on 101,386 variants from 1,425,819 bp, the best supported *K* under the *uncorrelated allele frequencies* model was determined to be $K = 3$, following the Evanno method (Evanno et al., 2005) for detecting the optimal number of clusters. Two big gene pools were revealed, separating European samples from arctic samples with varying degrees of admixture between the two gene pools, especially in Canadian *C. tridactylites* samples, in Iceland and Svalbard but also in northern European populations (UK, Scandinavia) and in Eastern European polyploids (Figure II-11). Putative admixture was found to be most prominent in Canada and Iceland with the Icelandic n = 6 samples (*C. islandica*) showing a higher degree of admixture with European samples than the n = 7 sample (*C. groenlandica*) from the North of Iceland.

Besides the two main gene pools, *C. danica* individuals showed clear distinctiveness from the remaining samples but without forming a completely separated group. Parts of these *C. danica* alleles were also found in *C. anglica*, *C. tatrae* and *C. borzaeana*.

The same analysis under the *correlated allele frequencies* model resulted in $K = 5$ as the optimal *K* and revealed a similar but less clear clustering, therefore the *uncorrelated* model was preferred for this dataset. However, using the *correlated* model, a distinctive pattern was detected for

C. tridactylites samples, this way separating the taxon from the remaining samples (results given with Supplemental Data Set II-14).

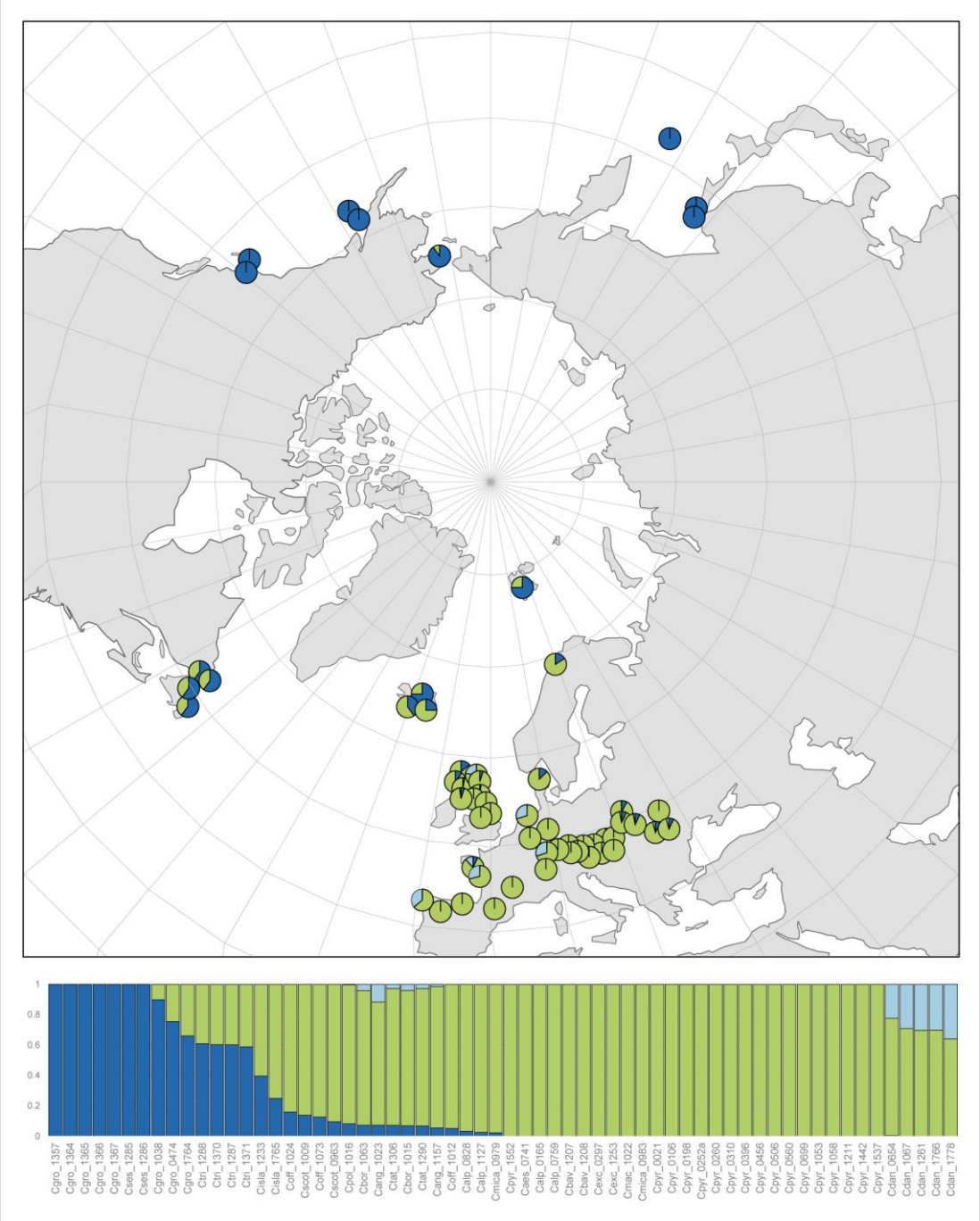

Figure II-11. Genetic assignment test in the genus *Cochlearia* with clustering results for K=3 (*uncorrelated allele frequencies*) and respective geographical representation. STRUCTURE analysis for 62 *Cochlearia* samples ('*all samples*' dataset) was based on 101,386 SNPs detected within 1,425,819 callable loci throughout the transcriptome (coding and non-coding). Every bar represents an individual and sample names are given below respective bars.

The *all samples* STRUCTURE analysis did not result in definite information on the origin of European polyploid taxa. However, several conclusions/interpretations can be drawn from the

results shown in here (see discussion). Moreover, for the Central European hexaploid taxon *C. bavarica* that was purely grouped within the green cluster in the *all samples* analysis, a separate analysis was performed in order to test the existing hypothesis on its allopolyploid origin from diploid *C. pyrenaica* and former inland relatives of tetraploid *C. officinalis* as parental taxa (Heubl and Vogt, 1985; Koch et al., 1996; Koch et al., 1998).

2) and 3) All diploids vs. European diploids

The *all diploids* dataset based on 119,629 SNPs from 2,829,028 bp callable sites and best analyzed with *K*=3) under *correlated allele frequencies*, suggests genetic distinctiveness of *C. tridactylites* and supports the admixture of European and circum-arctic gene pools in the geographically intermediate Icelandic populations of *C. islandica* and *C. groenlandica* (Figure II-12). In contrast to the *all samples* dataset under an *uncorrelated* model, no admixture with the Central European gene pool was found in Canadian diploids. Yet, the results for the *all diploids* dataset under *uncorrelated allele frequencies* with an optimal K of *K*=2 (Supplemental Data Set II-17) resembled the picture revealed from the uncorrelated *all samples* analysis with *C. tridactylites* showing strongly admixed genotypes. Shown in Figure II-12 are the results from the *correlated* analysis since these illustrate the special position of *C. tridactylites* as also suggested before based on whole chloroplast genome data (Figure II-4) and later on based on SplitsTree analyses and phylogenetic trees based on nuclear transcriptome SNP data (Figures II-15, II-16, II-17). Another difference between the *all samples* and the *all diploids* dataset was found in the degree of admixture within Central European diploids. Whereas traces of the arctic gene pool were restricted to European polyploids in the *all samples* dataset, reducing the dataset to diploid samples revealed slight admixture in Northern, Western and Eastern European diploids as well. A very similar pattern was revealed under the *uncorrelated allele frequencies* model. A reason for this difference between the two datasets might be the slightly higher number of variant sites detected for the *all diploids* dataset (101,386 SNPs vs. 119,629 SNPs) and the putatively higher sensitivity based hereon.
Among Central European diploids no strong structuring was detected, which is why the analysis was repeated excluding arctic diploids. In the *European diploids* analysis (*correlated allele frequencies*), the best supported *K* was *K*=2 and the results revealed two major gene pools, one western European gene pool in Northern Spain and one Central European gene pool centered in southern Germany, Switzerland and Austria (Figure II-13). Admixture was found in the putative geographical contact zones of these two gene pools in France, Belgium, Northern Germany and Great Britain but also to some degree in the easternmost populations in the Ukraine, Slovakia and some eastern Austrian populations. The *uncorrelated frequencies* analysis also revealed the best *K* to be *K*=2 but the structuring was less clear (Supplemental Data Set II-18) which is why the correlated model was preferred here.

4) C. bavarica

The STRUCTURE analysis including *C. bavarica, C. pyrenaica* and *C. officinalis* samples was based on 118,488 variants detected in 3,341,378 callable sites (all samples treated as hexaploids). Under both *correlated* and *uncorrelated allele frequencies* the dataset was best analyzed with *K*=2, and the *correlated* model, which is supposed to be more sensitive to discrete structuring (Falush et

al., 2003; Porras-Hurtado et al., 2013) strongly supports the hybrid origin of *C. bavarica*, whereas the pattern is also visible but less clear under the *uncorrelated allele frequencies* model (Figure II-14).

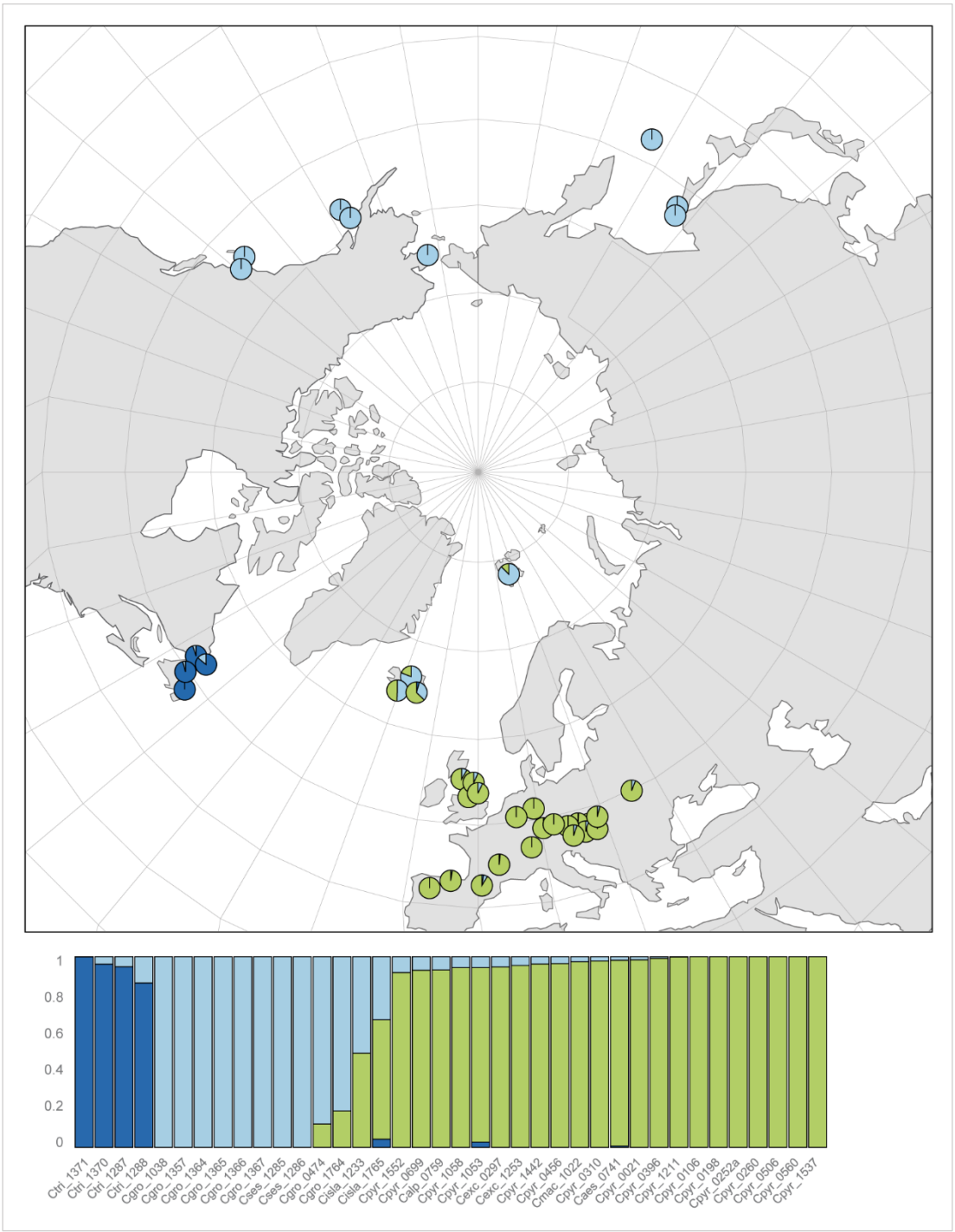

Figure II-12. Genetic assignment of all diploid *Cochlearia* samples (‘*all diploids*’). Results from STRUCTURE analysis (and geographical distribution) based on 119,629 SNPs (2,829,028 bp callable sites) with *K* = 3 under the correlated allele frequencies option.

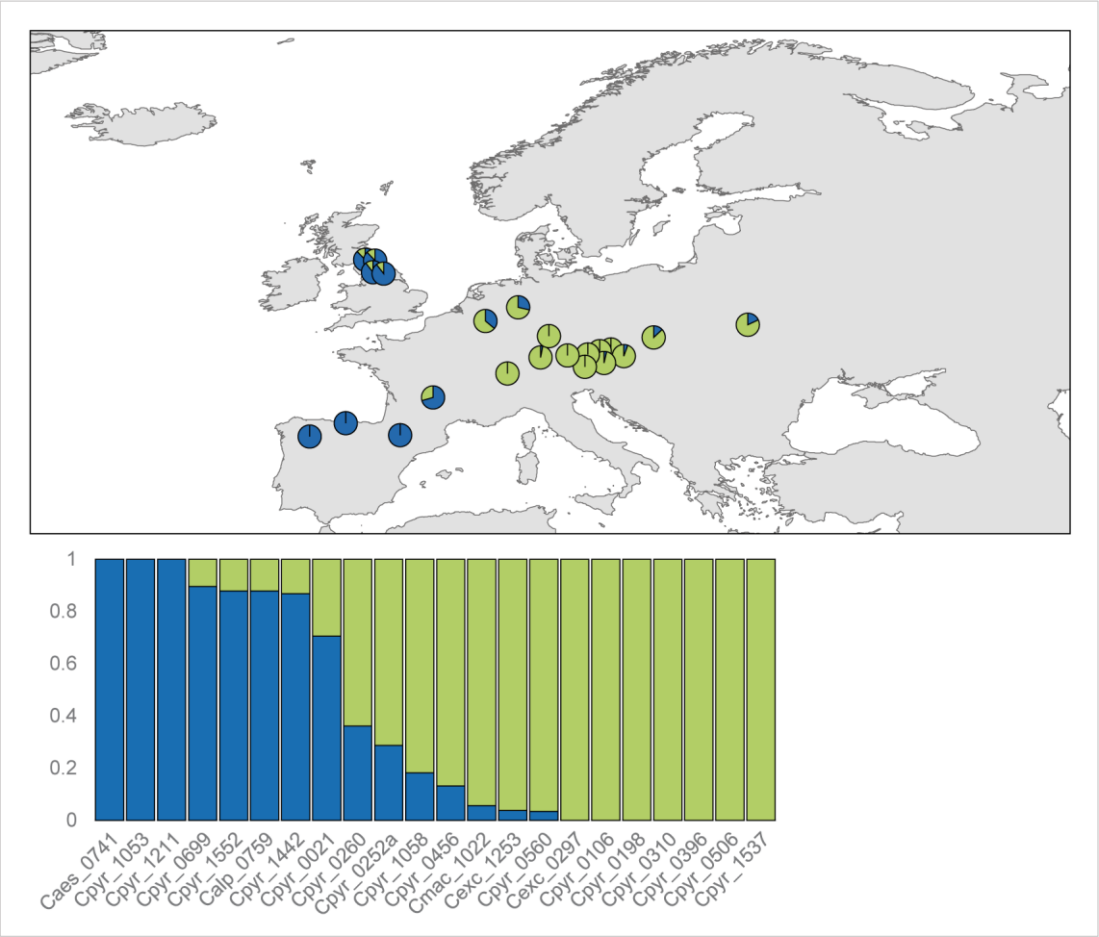

Figure II-13. Genetic assignment analysis of European diploid *Cochlearia* samples (‘*European diploids’*) based on nuclear transcriptome data (coding and non-coding). Results from STRUCTURE analysis (with geographical distribution) based on 310,202 SNPs with optimal detected K = 2 under the *correlated allele frequencies* option.

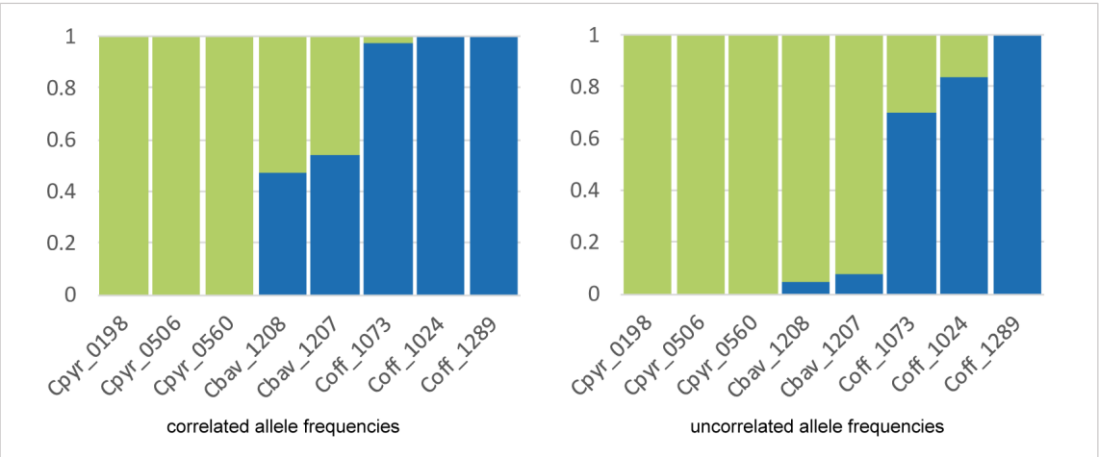

Figure II-14. STRUCTURE analyses investigating the allopolyploid origin of the hexaploid species *C. bavarica* (2n=6x=36) based on nuclear transcriptome data (coding and non-coding; 118,488 variants from 3,341,378 callable sites). Results shown for *K*=2 under the *correlated allele frequencies* model and the *uncorrelated allele frequencies* model.

SplitsTree analysis

Intersecting callable regions from transcriptome mappings for all *Cochlearia* samples analyzed via SplitsTree, resulted in a total overlap of 1,425,819 bp. Within these, variant calling revealed 113,721 SNPs and 107,269 of these remained after filtering for biallelic sites only (input files given with Supplemental Data Set II-11).

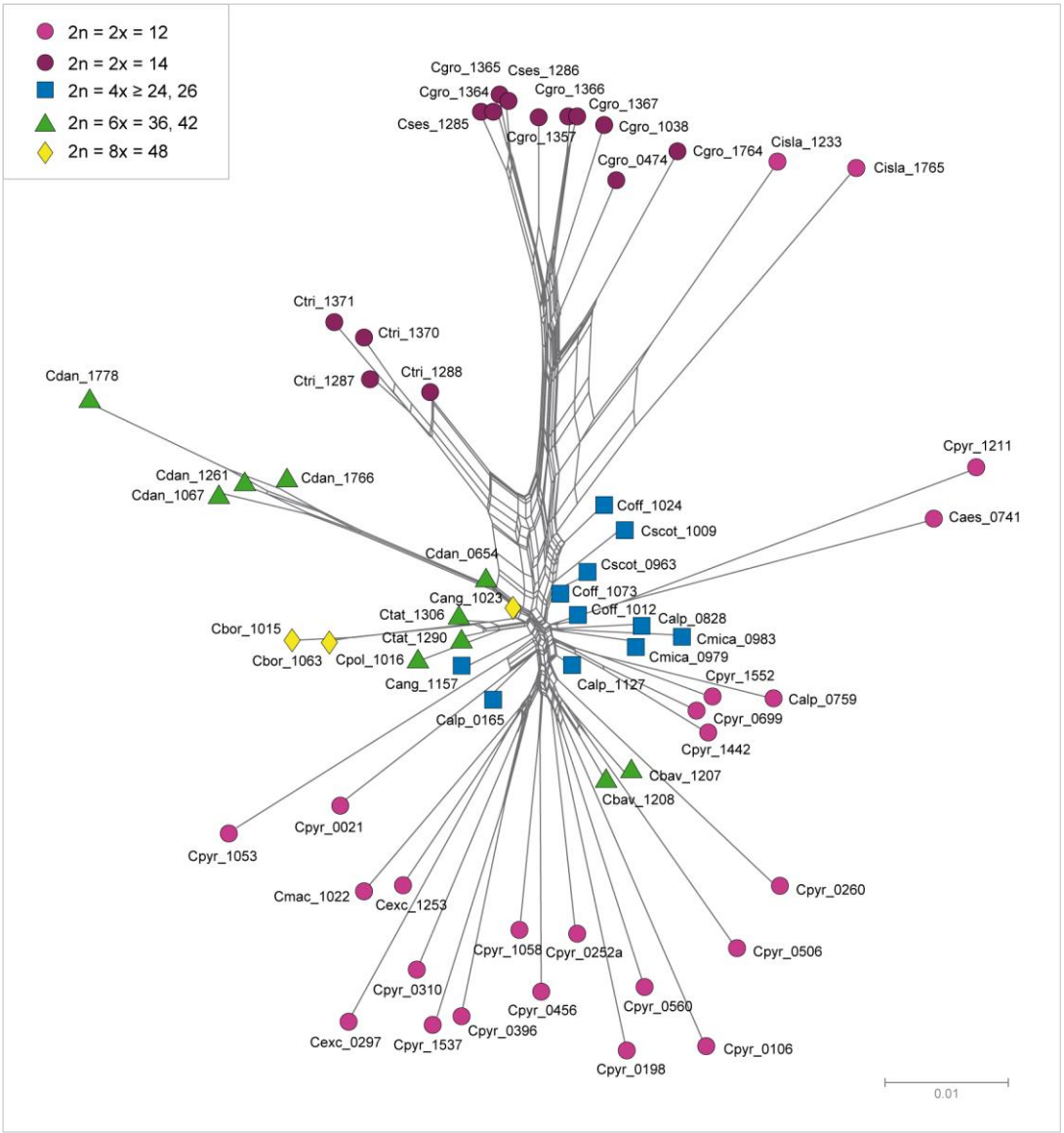

Figure II-15. SplitsTree analysis of a nuclear transcriptomic SNP dataset (diploid SNP calling) for 62 *Cochlearia* samples based on 113,721 variants including multiallelic sites detected across 1,045,252 bp of sequence information. Respective ploidy levels are represented by colors/symbols.

Figure II-15 shows the SplitsTree output from the dataset including multiallelic sites. The produced network contains a series of quadrangles especially in the central region, where all tetraploid samples as well as several higher polyploids are found. Diploid samples form several clusters that

are anchored at the center of the network. Central European diploids form a cluster with close association to hexaploid *C. bavarica*. The two diploid samples from Northern Spain (Cpyr_1211 and Caes_0741) show closer affiliation with tetraploid coastal and inland samples from the UK as already seen from chloroplast and mitochondrial phylogenies. Also, two French samples of *C. pyrenaica* (Cpyr_0021 and Cpyr_1053) are closely associated with tetraploid Calp_0165 from southern UK. Inland diploid samples from the UK fall within tetraploid UK inland samples. Arctic diploids are recognized as a distinct group, showing a deep split between *C. tridactylites* and the remaining arctic species. Whereas *C. tridactylites* stands close to the separated group of hexaploid *C. danica*, *C. islandica* shows closer affinities with tetraploid coastal taxa. The remaining Western European polyploid inland samples, namely *C. polonica*, *C. tatrae* and *C. borzaeana*, form another group anchored at the center of the network.

The SplitsTree network generated for biallelic sites produced a very similar result (Supplemental Data Set II-15). The main difference is a deeper split between *C. tridactylites* samples and the remaining arctic *Cochlearia* and the close association of *C. tridactylites* with *C. danica*. Yet both networks support the distinctiveness of *C. tridactylites* from the remaining arctic samples.

Phylogenetic trees

After excluding singletons from the called variants, the RAxML analysis for the 'diploid *Cochlearia* only' dataset of 38 samples was based on 119,629 SNPs, whereas a total of 73,087 SNPs remained for the 'diploids plus *Ionopsidium megalospermum*' dataset with 39 samples (input files given with Supplemental Data Sets II-9 and II-10).

The overall bootstrap support for both trees (Figures II-16 and II-17) is very high with most of the nodes being supported with 100%.

Both trees show strong geographical structuring and a clear separation of circum-arctic and European samples. The main difference between the two ML trees shown in Figures II-16 and II-17 is that including the *Ionopsidium* outgroup in the analysis results in a separation of *C. tridactylites* from the remaining arctic samples, forming a separate Eastern Canadian lineage at the basis of the *Cochlearia* diploids. Both trees match well with results from STRUCTURE and SplitsTree analyses (Figures II-11, II-12, II-15) as well as with the cytogenetic separation of arctic and European base chromosome numbers in diploid taxa (Figure I-2).

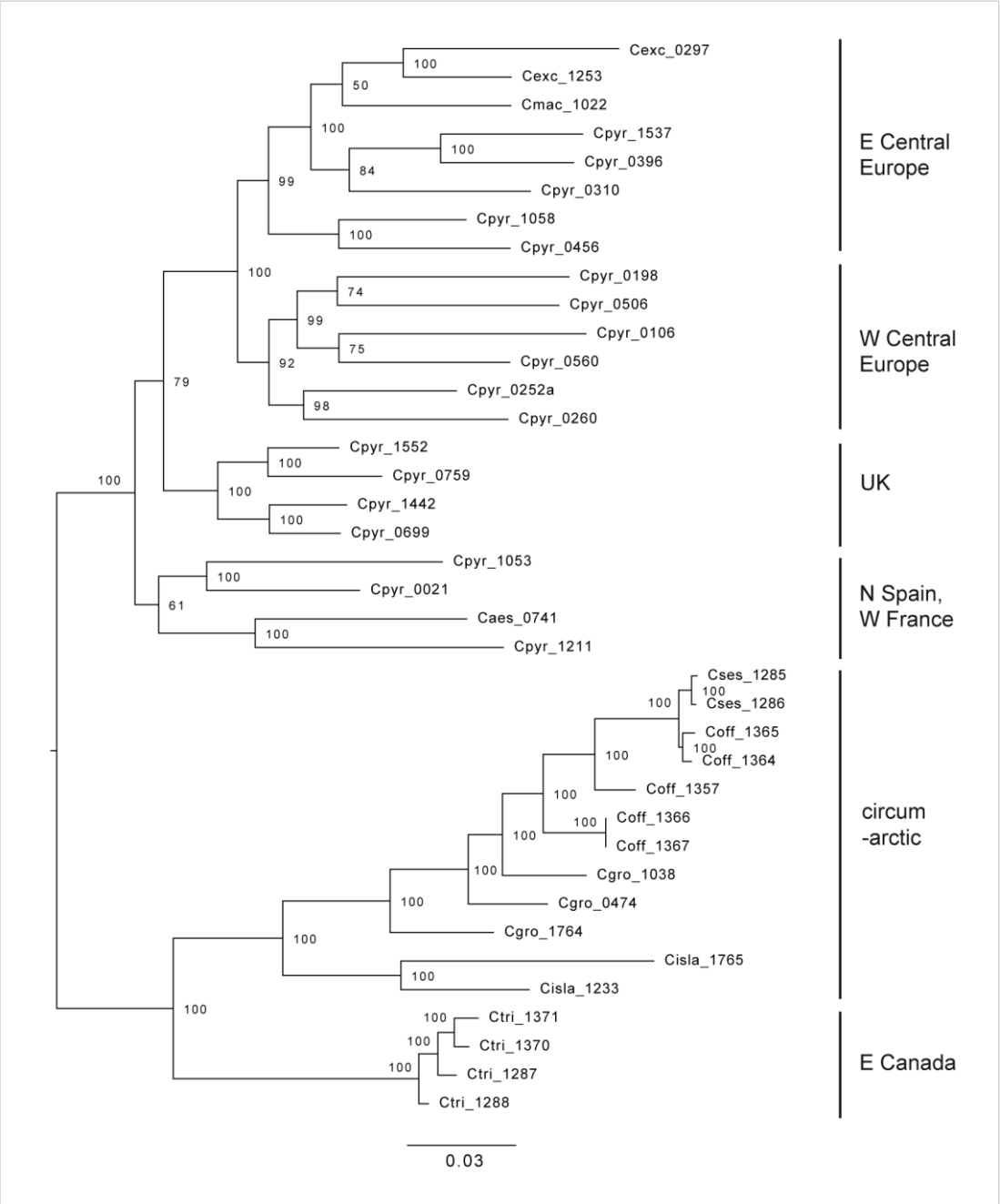

Figure II-16. Midpoint-rooted maximum likelihood tree of 38 diploid *Cochlearia* samples based on 73,033 transcriptome-wide nuclear SNPs (coding and non-coding; singletons excluded). Bootstrap support (1000 replicates) is shown near the respective nodes. Geographical distribution of groups and subgroups is indicated to the right.

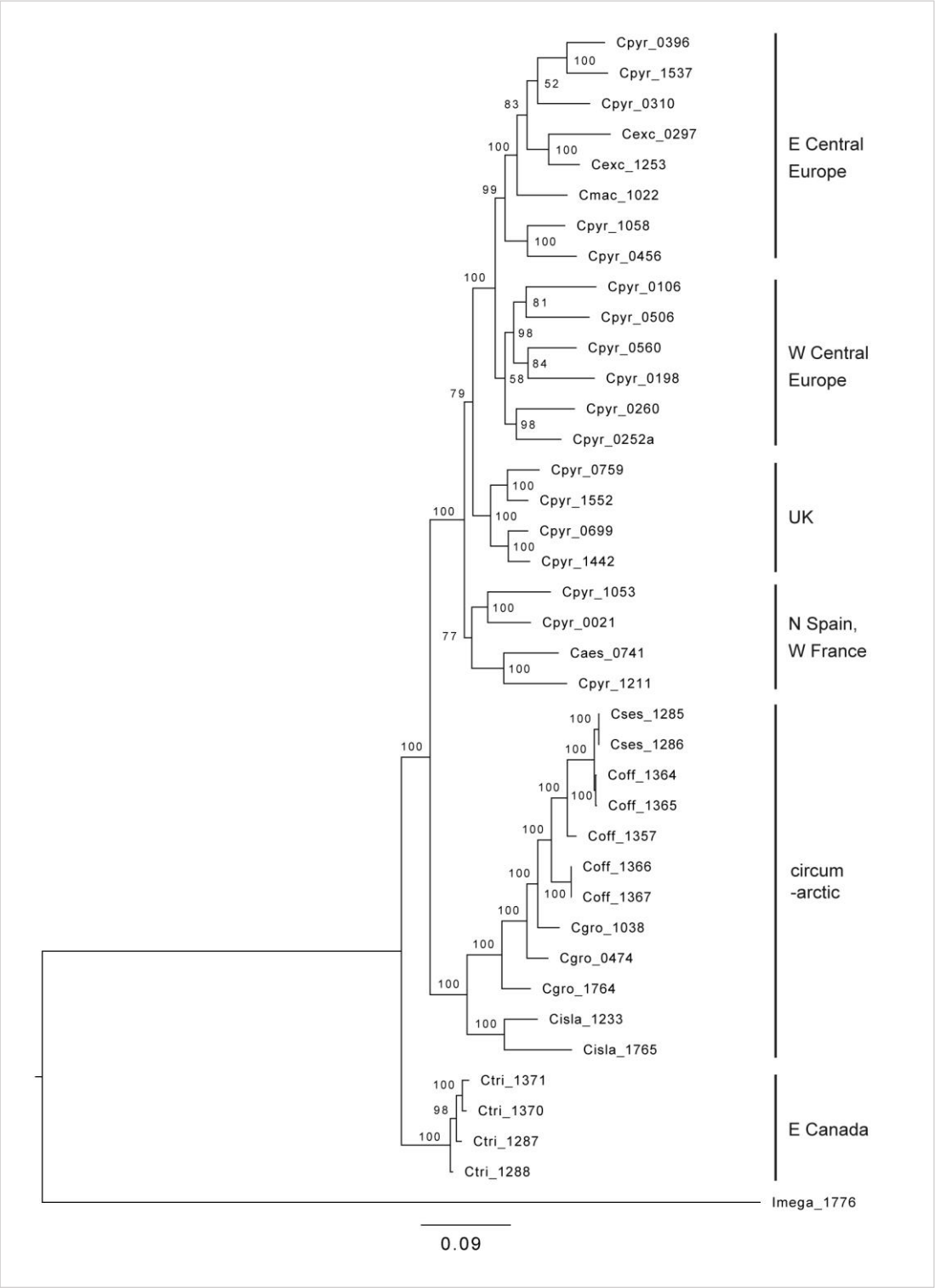

Figure II-17. Maximum likelihood tree of 38 diploid *Cochlearia* samples and *Ionopsidium megalospermum* set as outgroup based on 73,087 transcriptome-wide nuclear SNPs (coding and non-coding; singletons excluded). Bootstrap support (1000 replicates) is shown near the respective nodes. Geographical distribution of groups and subgroups is indicated to the right.

2.5 Discussion

In this study, complete plastome sequences, partial mitochondrial genome sequences and large nuclear SNP calling datasets were generated using a cost-efficient genome-skimming approach and *de novo* as well as reference-based mapping strategies. This approach is similar to the one presented by Straub et al. (2012) but taking advantage of the newly assembled and annotated *Cochlearia pyrenaica* transcriptome reference (Lopez et al., submitted) that enabled the high quality of nuclear data analysis.

The produced sequencing read quality was high for most of the samples although several DNA samples were extracted from comparatively old herbarium vouchers (e.g. 1954, 1957; see Supplementary Data Set II-1) and were thus highly fragmented and yielded in low quantities. As shown by the variation in the produced read number per sample, the multiplexing step caused a rather high degree of heterogeneity regarding the sequencing data output. This issue needs to be taken into consideration, especially for projects aiming at higher sample numbers per sequencing lane. However, thanks to the small number of individuals sequenced per lane in this study, the given variation in read number did not pose any bigger problems for the presented data analyses. Given the high copy number of chloroplast genomes per cell, and the consequently high sequencing depth, a cpDNA enrichment prior to sequencing is not necessary for the generation of whole plastome sequences and was therefore skipped in this study after the first sequencing lane.

SNP calling in polyploids, especially in higher ploidy levels, is still problematic and available strategies usually rely on high sequencing depths (e.g. Lynch, 2009; see Morris et al., 2011). Moreover, since all of the study samples were sequenced with the same or a similar sequencing depth, irrespectively of ploidy levels, a reduced power of calling heterozygous sites under a tetra-, hexa-, or octoploid SNP calling was observed compared to diploid SNP callings for the same samples. A more exhaustive detection of polyploid alleles could be achieved with a higher coverage of the nuclear genome, but it has been shown, that for the purpose of the phylogenomic analyses presented herein, this was not necessary.

Overall, these results nicely demonstrate once more the suitability of multiplexed Illumina sequencing for producing large-scale high-quality phylogenomic data sets at reasonable costs.

2.5.1 Congruence and incongruence in organellar phylogenies

Regarding the very low bootstrap support (< 50%) for most of the conflicting basal nodes in the generated mitochondrial phylogeny (Figure II-10), the detected incongruences along basal nodes of chloroplast and mitochondrial phylogenies most likely can be attributed to the difficulties in the analysis of mitochondrial sequencing data within the genus *Cochlearia* that can be easily explained by the generally complex mitochondrial genome evolution in plants which is characterized by a highly dynamic structural organization due to frequent structural rearrangements as well as an exchange of DNA with chloroplast and nuclear genomes (e.g. Burger et al., 2003; reviewed by Knoop, 2004). Likewise, the comparatively low number of detected variants and the high amount of polytomies revealed from mitochondrial data reflect the generally slow sequence evolution of

plant mitochondrial genomes in comparison to chloroplast and nuclear genome evolution (Knoop, 2004). Despite the present incongruences between chloroplast and mitochondrial phylogenies, the phylogenetic insights gained via complete chloroplast genomes are in large parts supported by mitochondrial phylogenies although the latter lack the high resolution and strong bootstrap support, indicating that the picture drawn by organellar genomes, and especially by chloroplast genomes, is very stable and reliable.

2.5.2 Organellar genome phylogenies, divergence time estimates and nuclear SNP data analyses – rewriting the evolutionary history of the genus *Cochlearia*

One of the main goals of the presented sequencing approach was the generation of complete chloroplast genomes and a highly resolved organellar phylogeny based heron. The insights gained from the latter (Figure II-4), in combination with the respective divergence estimates (Figures II-8 and II-9) and the results of nuclear SNP data analyses (Figure II-11 till Figure II-17) revealed several surprising and highly interesting findings and rewrite big parts of previous evolutionary concepts for the genus *Cochlearia*.

Early colonization of and Pleistocene survival in arctic regions

As stated above, the present-day distribution of species and genetic diversity was strongly affected by the drastic climatic fluctuations during the Pleistocene (Hewitt, 2000; Hewitt, 2004). Main refuge areas for European species have been detected in the Iberian Peninsula, Southern Italy and the Balkan (Taberlet et al., 1998) but so called "cryptic" northern refugia have been suggested as well (e.g. Provan and Bennett, 2008; dicussed in Stewart and Lister, 2001). For the genus *Cochlearia*, cytological studies by Gill (e.g. Gill, 1971; Gill et al., 1978) early on indicated an ancestral base chromosome number of n=6, as found today in European diploid taxa, that way implying a survival of the whole genus south of the European glaciers. Based on cpDNA restriction site variation (RFLP) data, Koch et al. (1996) came to a similar conclusion, also regarding the phylogenetic placement of *C. aestuaria* at a basal position of Central European taxa. Moreover, the low levels of chloroplast divergence as revealed by RFLP data indicated a recent origin of the *Cochlearia* species complex and postglacial colonization of the arctic regions. Now, the fact that the oldest chloroplast haplotypes revealed in the present study via phylogenetic analyses of both (nearly) complete chloroplast genomes and partial mitochondrial genomes were found in western Canadian *C. tridactylites* (2n=14), contradicts the former hypothesis of a post-glacial colonization of circum-arctic regions and instead hints at a very early colonization of and Pleistocene survival in the arctic, possibly followed by (repeated) recolonization of Northern and Central Europe from there. Also, inconsistent with previous evolutionary concepts (Koch et al., 1996; Koch et al., 1998), in both organellar phylogenies *C. aestuaria* does not have a basal position in respect to Central European coastal and inland taxa but only within the purple, western European lineage, including several coastal polyploids from the UK. The basal position of *C. tridactylites* in ML trees generated for diploid samples from transcriptome-wide nuclear SNP callings (Figures II-16, II-17) also strongly supports the scenario of an arctic colonization during the early evolution of the genus. Likewise, the high cpDNA divergence detected in arctic regions, stands in contrast to a strongly

reduced genetic diversity in these northern areas as revealed for several examples of glacial survival in southern areas and post-glacial extension to northern regions (e.g. *Ranunculus glacialis* (Schönswetter et al., 2003); *Ranunculus pygmaeus* (Schönswetter et al., 2006); (Hewitt, 2004)). Thus, as suggested before by Hultén (1971) for *C. groenlandica*, ice-free arctic refuge areas can be assumed for the genus *Cochlearia* as also described for other plant taxa before (e.g. *Saxifraga oppositifolia* (Abbott et al., 2000; Abbott and Comes, 2004); *Oxyria digyna* (Allen et al., 2012)). The distribution area of *C. tridactylites* along the western Canadian coast is not among the classical regions that were suggested to have served as Pleistocene refugia, like the ice-free areas in northwestern America and eastern Siberia, also called Beringia (reviewed by Abbott and Brochmann, 2003). However, ice-free areas have also been considered e.g. for the Canadian Arctic Archipelago and the continental shelf of the North American east coast (Hultén, 1937; reviewed by Abbott and Brochmann, 2003; Holland, 1981). The fact that arctic *Cochlearia* taxa are also represented in the basal pink chloroplast lineage, as well as in basal subgroups of the green lineage suggests additional arctic refugia and putative colonization of Iceland and Northern Europe out of these. It has been shown that colonization throughout arctic regions via recurring and frequent long-distance dispersal, putatively enabled by e.g. drifting sea ice, wind or birds, has been and putatively still is a common process in arctic plant species, especially if the species are pre-adapted to the environmental conditions at the destination locality (Alsos et al., 2007). Although seeds of *Cochlearia* taxa are not particularly equipped for long distance dispersal, seed dispersal among coastal habitats is expected to be generally dynamic (see e.g. Kadereit et al., 2005) due to ocean currents or dispersal via sea birds. A high abundance of *Cochlearia groenlandica* close-by seabird colonies with significantly higher biomass production in vicinity of the colony has been described for *C. groenlandica* populations from south-western Svalbard (Zmudczyńska-Skarbek et al., 2013), implying an important role of the dispersal by birds. Regarding these patterns, a dynamic colonization history of *Cochlearia* throughout arctic regions seems likely. These results are highly interesting as they are somehow shifting the evolutionary origin of this cold-adapted genus towards the arctic regions.

Putative additional refuge areas in Central Europe and the impact of recurrent environmental changes during the Pleistocene

The onset of diversification within *Cochlearia* was dated to about 0.659 mya and thus seems to have started in concert with the beginning of the Pleistocene's major climatic fluctuations which was dated to 0.7 mya (Webb III and Bartlein, 1992; Comes and Kadereit, 1998). This and the observed correlation of high glacial periods (Petit et al., 1999; Augustin et al., 2004) with diversification times in the six major chloroplast lineages (Figure II-9), implies a boosting impact of these climatic fluctuations on the diversification within the cold-adapted genus *Cochlearia* after a period of evolutionary stasis during the climatically more stable late Miocene (also described in Hohmann et al., 2015). The lack of resolution in several nodes, especially regarding the green and the purple lineage, most likely reflects periods of rapid, reticulate radiations, coinciding with high glacial periods of the Riss glacial (~150-200 kya) and the Würm glacial (~30-80 kya). Here, frequent events of polyploidization and hybridization likely might have facilitated adaptation to

the new environmental conditions as also suggested recently for the whole Brassicaceae family (Hohmann et al., 2015).

The detected diversification over the course of several glacials and interglacials and the revealed split times also indicate that besides the arctic, other regions in Central Europe putatively served as refuge areas during periods of glaciation. This is in accordance with former hypotheses (Koch, 2002; Koch et al., 2003) suggesting additional European refuge areas for *Cochlearia*, for example for *C. excelsa* in unglaciated alpine areas of the south-eastern Alps where refugia have been detected for other plant taxa as well (Tribsch et al., 2002; Tribsch and Schönswetter, 2003). Likewise, south-western England was suggested as a Pleistocene refuge for *Cochlearia*, at least for the time of the last glaciation, based on the discovery of *C. officinalis*/*C. pyrenaica*-like plant material found in autochthonous deposits of the respective period (Lang, 1994; Koch et al., 1996). Moreover, the basal phylogenetic position of *C. danica* from northern Spain, as well as the early split of the blue coastal lineage dated to 0.317 mya, indicate the existence of other glacial refuge areas along Atlantic shorelines in south-western Europe. A rapid northward migration along the coasts latest after the final retreat of the ice sheets is likely, as indicated for *C. officinalis* and *C. danica* by former studies that showed a decrease in genetic variation with higher latitudes (Koch et al., 1998). Likewise, glacial survival must be assumed for the easternmost areas of the distribution range, given 1) the phylogenetic placement of Romanian *C. borzaeana* in the basal pink lineage with divergence from the remaining arctic samples dated to 0.128 mya, as well as 2) the early separation of the orange *C. tatrae*-lineage, dated to 0.229 mya, with further diversification within *C. tatrae* starting at about 86 kya.

Separation of arctic and European gene pools

An early separation of an arctic and a European gene pool, followed by further diversification in the two areas over the course of several glacial periods, is also supported by the results gained via SplitsTree (Figure II-15) and STRUCTURE analyses (Figures II-11, II-12). This distinction is also reflected in the cytogenetic pattern revealed in chapter 1 for n=6 and n=7 diploid taxa by the literature review of documented chromosome counts (Figure I-2). As already outlined in chapter 1, the suggested basal position of arctic 2n=14 taxa in the evolution of the genus *Cochlearia*, as revealed from the combined results of organellar phylogenies and nuclear data analyses, implies an ancestral base chromosome number of n=7, this way contradicting previous hypotheses (e.g. Gill, 1971; Gill et al., 1978). The dominance of n=7 taxa in basal lineages of the chloroplast phylogeny also supports this idea, yet further comparative cytogenetic studies are needed in order to finally unravel the karyotypic evolution of the genus.

A prominent feature revealed from the different STRUCTURE analyses was the admixture between the northern, arctic gene pool and the southern, European gene pool especially in Western Canada, Iceland and Svalbard but also in Northern Europe and some areas in Eastern and Western Europe (Figures II-11, II-12). Repeated inter- and postglacial admixture and hybridization in contact zones that emerged with migration following the retreating ice have been postulated for other plant taxa before (e.g. *Clausia aprica*, Franzke et al., 2004; Stebbins, 1984; Hewitt, 1996) but based on the analyses done so far, no final conclusions can be made on whether the picture seen in *Cochlearia* is a result of repeated migration/gene flow and thus real admixture from arctic

areas into parts of the central European gene pool (and vice versa?) or rather the genetic legacy of an early migration southwards.

A complex and reticulate evolutionary history in Great Britain

Regarding the colonization history of Great Britain, the patterns gained from organellar genome data analysis suggest at least two independent colonization events from two phylogenetic lineages, namely the purple lineage putatively originating from northern Spain and the green, Central European lineage (Figure II-6). Northward migration from the Iberian peninsula along the Atlantic coast of France and up to Great Britain has already been shown to have happened in postglacial times, e.g. in white oaks (Petit et al., 2002), however, as revealed from the generated divergence time estimates, the colonization history of *Cochlearia* started much earlier, in accordance with a suggested Pleistocene refuge area in south-western England as stated above.

The polyphyletic grouping of several taxa and the admixture of inland and coastal taxa as well as different ploidy levels in subgroups of the chloroplast phylogeny (Figure II-4) likely reflect a highly dynamic and reticulate evolutionary past of the genus *Cochlearia* in Great Britain. A repeated origin of both coastal and upland ecotypes/phenotypes was also suggested by Gill (2008) who detected neither geographical clustering nor significant genetic divergence between different taxonomic groups, cytotypes or habitats based on an AFLP data analysis of British populations. Admixture between the two chloroplast lineages has been revealed from genetic assignment tests (Figure II-13) and besides this, putative admixture with the arctic gene pool was also apparent, especially in northern samples from both coastal and inland locations (Figures II-11, II-12). Overall, the results are displaying Great Britain as a contact zone of several evolutionary lineages where gene flow between coastal and inland/upland populations, different species and across ploidy levels putatively happened frequently.

2.5.3 New insights into the complex evolutionary histories of polyploid taxa

The origin of the different polyploid taxa within the genus has been studied intensively and various hypotheses have been made so far (e.g. Gill, 1973, 1976; Gill et al., 1978; Koch et al., 1996; Nordal and Laane, 1996; Koch et al., 1998; Koch, 2002; Kochjarová et al., 2006). Infraspecific variation for chloroplast genomes, as shown in here by the lack of monophyly in all coastal polyploid species, has been revealed before for the coastal taxa *C. officinalis*, *C. danica* and *C. anglica* based on cpDNA restriction site variation analysis (RFLP) by Koch et al. (1996). If a former incorrect taxonomic placement of plant populations due to the previously mentioned weak morphological differentiation can be ruled out, this observation can be best explained by reticulate evolution and multiple origin (Koch et al., 1996).

The information gained in this study, based on cytogenetic and phylogenomic data, provide additional insights on the various polyploidization processes as discussed below.

C. officinalis (2n=4x=24), C. scotica (2n=4x=24)

Since all of the three samples of *C. officinalis* that were included in the STRUCTURE analyses (Figure II-11) show signatures of admixture between the artic and the Central European gene pool, a simple autopolyploidization from *C. aestuaria* relatives, as suggested in earlier studies (Gill et al., 1978; Elkington, 1984; Koch et al., 1998) does not seem to be very likely, as *C. aestuaria* is purely grouped within the Central European gene pool in the *all samples* dataset. In the *all diploids* STRUCTURE analysis (Figure II-12), minor signs of admixture with the arctic gene pool are seen in *C. aestuaria* as well, however, this can hardly account for the strong admixture seen in the *C. officinalis* samples, especially in Coff_1024 and Coff_1073. In the SplitsTree analysis (Figure II-15), the *C. officinalis* samples are found in between the Icelandic diploids from the arctic complex, and the Spanish diploids including the coastal *C. aestuaria*. Especially Coff_1012 is closely connected with *C. aestuaria* whereas the other two samples stand closer to *C. islandica*. Whereas Coff_1024 from northern Norway is found in an old lineage of coastal polyploids in the chloroplast tree/BEAST analysis (Figures II-4, II-9), the other two samples were estimated to be of younger origin within the purple, mainly British lineage that also includes *C. aestuaria*. Altogether, the results are in favor of a repeated formation of the tetraploid *C. officinalis* with two likely scenarios: 1) autopolyploidization from former northern coastal diploids showing strong arctic impact or 2) hybridization/allopolyploidization between northern diploids and *aestuaria*-like diploids.

Regarding the peculiar phylogenetic placement of *C. scotica* in the green lineage of both organellar phylogenies (Figures II-4, II-10), close to the Austrian diploid *C. macrorrhiza*, a mix-up or wrong labeling of samples can be ruled out based on the results gained from nuclear SNP data analyses. Therefore, other reasons for the unexpected phylogenetic placement, such as chloroplast capture or occasional long-distance dispersal events in the organellar genomes must be expected. In general, no clear distinction can be made between analyzed *C. officinalis* and *C. scotica* samples regarding STRUCTURE and SplitsTree results. Thus, as suggested before (Gill, 1973), *C. scotica* might rather be treated as a small growing morphological variant of *C. officinalis*.

C. alpina (2n=4x=24) and C. micacea (2n=4x=26)

The various results presented herein, did not reveal clear insights into the evolutionary background of the British tetraploid species *C. alpina* and *C. micacea*, yet several assumptions can be made.

Whereas the two *C. micacea* samples from the Scottish mountains Beinn an Dòthaidh (Cmica_0979) and Ben Lawers (Cmica_0983) respectively, were grouped together in both organellar phylogenies (Figures II-4, II-10), monophyly could not be revealed for *C. alpina*, indicating a repeated formation of this tetraploid inland species. The dispersed occurrence of *C. alpina* samples in the SplitsTree network (Figure II-15) further supports this scenario.

Regarding *C. micacea*, Gill (1973) suggested a formation via primary tetrasomy from autotetraploid *C. officinalis*. However, this hypothesis is based on a distinct karyotype of 2n=26 as described by Gill (1973) which was later on questioned by Nordal and Stabbetorp (1990). As shown in chapter 1, the species-specific chromosome number of 2n=26 was not revealed by own chromosome counts of *C. micacea* (Table I-3) indicating the need for further cytogenetic studies.

Based on the association of Calp_0165 and Calp_1127 with different diploid subgroups in the SplitsTree network, putative events of autopolyploidization from *C. pyrenaica* cannot be ruled out for these samples. Besides this association, both *C. alpina* and *C. micacea* are connected to coastal species, especially to tetraploid *C. officinalis*, in the chloroplast phylogeny as well as in the SplitsTree output, thus a general association of both *C. micacea* and *C. alpina* with the coastal species complex can be hypothesized. Based on both SplitsTree and STRUCTURE analyses no clear distinctiveness between *C. alpina* and *C. micacea* could be revealed and given the uncertainties in ascribing distinct karyotypes, a separate taxonomic status of the two species might be questioned as also suggested by Gill (2008). Given the assumption of a formerly wide inland distribution of tetraploid *C. officinalis*-type plants in Central Europe (e.g. Koch, 2002; see below for *C. bavarica*, *C. polonica*/*C. tatrae*/*C. borzaeana*), the same can be assumed for Great Britain and regarding the close associations of both taxa with the coastal species complex, especially with *C. officinalis*, both taxa might also be treated as inland populations of *C. officinalis* as done before for *C. alpina* populations by e.g. Clapham et al. (1990) or Koch et al. (1998).

C. bavarica (2n=6x=36)

The evolutionary history of the inland hexaploid *C. bavarica* has been discussed by several studies (Vogt, 1985; Koch et al., 1996; Koch et al., 1998; Abs, 1999; Koch, 2002) and there is a general consensus about its formation via allopolyploidization from diploid *C. pyrenaica* and tetraploid *C. officinalis*. This hypothesis was also supported by the separate STRUCTURE analysis performed in here (Figure II-14). However, different theories exist regarding the timing of this allopolyploidization event. Whereas Vogt (1985) argued for a very recent origin (17^{th}/18^{th} century) from *C. pyrenaica* and putatively cultivated *C. officinalis* from Bavarian monasteries, Koch (2002) demonstrated that an inter- or postglacial allopolyploidization from *C. pyrenaica* and former inland relatives of *C. officinalis* without anthropogenic impact is more likely. This hypothesis is supported by the divergence time estimates based on whole chloroplast genomes (Figure II-9), where the separation of the two *C. bavarica* samples from the respective *C. pyrenaica* samples was dated to ~10,000 and ~11,000 years ago. Both chloroplast and mitochondrial phylogenies are in favor of *C. pyrenaica* serving as the maternal plants in the hybridization process and indicate multiple origins of *C. bavarica* throughout the distribution area that is nowadays split into two isolated metapopulations but was putatively wider and connected by the time of the origination of the species (Koch, 2002). Repeated independent origination has been suggested for several allopolyploid Brassicaceae species before within the genera *Arabidopsis* (Mummenhoff and Hurka, 1995; O'Kane Jr et al., 1996), *Cardamine* (Neuffer and Jahncke, 1997; Franzke and Mummenhoff, 1999), *Draba* (Koch and Al-Shehbaz, 2002) and *Yinshania* (Koch and Al-Shehbaz, 2000). However, the SplitsTree analysis (Figure II-15) shows a slightly different picture with the two *C. bavarica* samples closer together and both associated with *C. pyrenaica* (Cpyr_0506) from Baden-Württemberg (Hohenlohe). Therefore, the scenario of a single ancient origin followed by genetic differentiation over geographic distances as developed in a study by Koch (2002) becomes more likely, that way accounting for the strong geographic partitioning detected therein based on isozyme data.

C. polonica (2n = 6x = 36), C. tatrae (2n = 6x = 42), C. borzaeana (2n=8x=48)

The evolutionary background of the three highly polyploid endemic inland species *C. polonica*, *C. tatrae* and *C. borzaeana* has not been fully unraveled yet, but despite the different karyotypes, similar allopolyploidy events from *C. pyrenaica*-like diploids and *C. officinalis*-like tetraploids have been suggested for all of them (Bajer, 1951; Koch et al., 1996; Koch et al., 1998; Koch et al., 2003; Kochjarová et al., 2006). A close association of the three taxa is indicated in the SplitsTree results (Figure II-15) where they form a separate group, and likewise, STRUCTURE analysis (Figure II-11) revealed similar patterns for the three species with signatures of admixture between the Central European and the arctic gene pool, this way supporting a similar allopolyploid formation. The monophyly of the taxa, as revealed from whole chloroplast genomes, confirms independent polyploidy events and the phylogenetic placement of *C. polonica* closest to *C. pyrenaica* samples from the Slovakian Carpathian Mountains and the Ukraine supports the suggested involvement of *C. pyrenaica* in the formation of this species. The revealed signs of admixture of *C. tatrae* and *C. borzaeana* samples with the *C. danica*-specific cluster in the STRURTURE analysis are also reflected in the SplitsTree network where the whole group is placed next to *C. danica*. Given the basal phylogenetic position of Spanish *C. danica* in the pink lineage of the chloroplast phylogeny together with *C. borzaeana* and arctic 2n=14 diploids, one possible interpretation might be an early southward migration of arctic diploids that are meanwhile extinct in Central Europe but served as parental taxa for the formation of these old polyploid taxa, that way also explaining the arctic footprint in the STRUCURE results. In support of this hypothesis, affinities between *C. tatrae* and arctic *C. groenlandica* have been described before based on cpDNA analyses by Koch et al. (1996). However, no final conclusions on the evolutionary origin of the three Central European polyploid inland taxa can be drawn from the analyses presented herein.

C. danica (2n=6x=42)

As already stated in the general introduction, despite the common base chromosome number of n=7 (Gill, 1976), a closer phylogenetic connection between *C. danica*, the only annual species within the genus, and arctic *C. groenlandica*, was rejected by Koch et al. (1996) based on cpDNA data analyses. Likewise, given the special cytogenetic constitution of *C. danica*, with three chromosome sets of varying size (Gill, 1976), a closer relationship to *C. tatrae*, also didn't seem likely. Thus till now, allopolyploid origination from tetraploid *C. officinalis* relatives and progenitors of *C. aestuaria* seemed the most plausible hypothesis (Koch et al., 1996; Koch et al., 1998). However, here the SplitsTree result (Figure II-15) shows a close relationship between *C. tridactylites* and *C. danica*, that is even more prominent in the 'biallelic sites' analysis (Supplemental Data Set II-15), thus after all suggesting a role of this n=7 taxon in the formation of *C. danica*. Moreover, in the chloroplast phylogeny (Figure II-4), Spanish *C. danica* is grouped in the basal pink lineage together with *C. borzaeana* (2n=6x=42/48) and four arctic diploids with 2n=14, hereby further strengthening the association of *C. danica* with the arctic species complex. However, STRUCTURE results (Figure II-11) do not support this association as no admixture with the arctic cluster was detected and instead, *C. danica* samples showed a pattern clearly distinct from all other samples but without forming a completely separated group.

Close associations between *C. anglica* and *C. danica* have been described based on secondary phenolic compound comparisons (Vogt, 1985) and indeed, in the SplitsTree network, the *C. danica* group is found close to *C. anglica* (Cang_1023) as also supported by STRUCTURE analyses (but see below the discussion of *C. anglica* samples).
In summary, the origin of the coastal hexaploid *C. danica* remains cryptic but an early formation with involvement of n=7 taxa seems likely and subsequent hybridization and introgression among coastal taxa can be assumed, given the frequently described naturally occurring hybrids with *C. danica* (e.g. Pegtel, 1999).

C. anglica (2n=8x=48)

The octoploid *C. anglica*, occurring along Central European coasts from northern France to southern Sweden, was suggested to have originated via auto- or allopolyploidization from a tetraploid taxon similar to recent *C. officinalis* (Koch et al., 1996).
Here, both samples referred to as *C. anglica* in the STRUCTURE analysis (Figure II-11) show signatures of admixture with the *C. danica* specific gene pool, arguing against a polyploidization event that only involved a *C. officinalis*-like taxon since none of the analyzed *C. officinalis* samples shows affinities with the *C. danica* specific genotypes. However, for one of the two *C. anglica* populations analyzed in here, chromosome counts are missing and both measured genome size and chromosome count for the other individual/population (Cang_1157, Scotland, north of Creetown) are deviating from the expected values for *C. anglica* (Table I-3) and rather suggest hybridization between *C. anglica* and *C. officinalis* (see Schulz, 2015), also known as *C. x hollandica* (first described by Henrard, 1915). No cytogenetic data was generated for the other "*C. anglica*" sample in the dataset (Cang_1023, France Dep. Morbihan, La Trinité-sur-Mer). Therefore, possibly none of the analyzed samples was a pure *C. anglica*. SplitsTree results indicate close affinities of Cang_1023 to the *C. danica* complex whereas Cang_1157 is not clearly associated to a certain subgroup. Regarding the cytogenetic uncertainty, no final conclusions on the origin of octoploid *C. anglica* can be made in here. Likewise, due to the described uncertainties, the lack of monophyly as revealed from the generated chloroplast phylogeny, should be evaluated with caution.

2.5.4 Repeated adaptation to arctic/alpine habitats

As outlined above, one of the pending questions regarding the present-day distribution of *Cochlearia* taxa was if adaptation to the extreme environmental conditions of arctic and alpine habitats, especially regarding the Central European high alpine regions, evolved once or several times in parallel. This question can finally be answered based on the insights gained via phylogenetic analysis of whole chloroplast genomes (Figure II-4) in combination with the respective divergence time estimates (Figure II-9). As stated above, the phylogenetic results indicate an early adaptation to arctic habitats and are clearly in favor of subsequently repeated, independent colonization of and thus adaptation to the different high alpine regions in Central Europe. As revealed from the BEAST analysis, these processes must have taken place temporally separated. Whereas the orange *C. tatrae*-lineage diverged about 229 kya, within the Riss glacial, split times for *C. excelsa* are dated to less than 100 kya and thus fall within the Würm glacial and

the same applies for the taxa from alpine habitat types in Great Britain, namely *C. alpina* and *C. micacea*. As a closer look at the chloroplast ML tree reveals, taxa adapted to extremely cold environments can be found in every phylogenetic lineage. Regarding the hypothesized arctic origin of recent taxa (see 2.5.2), this can be interpreted as a result of pre-adaptation to these environments, putatively carried by standing genetic variation.

2.6 References

Abbott, R.J., and Brochmann, C. (2003). History and evolution of the arctic flora: in the footsteps of Eric Hultén. Molecular ecology **12,** 299-313.

Abbott, R.J., and Comes, H.P. (2004). Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (Purple saxifrage). New Phytologist **161,** 211-224.

Abbott, R.J., Smith, L.C., Milne, R.I., Crawford, R.M., Wolff, K., and Balfour, J. (2000). Molecular analysis of plant migration and refugia in the Arctic. Science **289,** 1343-1346.

Abs, C. (1999). Differences in the life histories of two *Cochlearia* species. Folia Geobotanica **34,** 33-45.

Allen, G.A., Marr, K.L., McCormick, L.J., and Hebda, R.J. (2012). The impact of Pleistocene climate change on an ancient arctic–alpine plant: multiple lineages of disparate history in *Oxyria digyna*. Ecology and evolution **2,** 649-665.

Alsos, I.G., Engelskjøn, T., Gielly, L., Taberlet, P., and Brochmann, C. (2005). Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. Molecular Ecology **14,** 2739-2753.

Alsos, I.G., Eidesen, P.B., Ehrich, D., Skrede, I., Westergaard, K., Jacobsen, G.H., Landvik, J.Y., Taberlet, P., and Brochmann, C. (2007). Frequent long-distance plant colonization in the changing Arctic. Science **316,** 1606-1609.

Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K., and Palmer, J.D. (2010). Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Molecular Biology and Evolution **27,** 1436-1448.

Augustin, L., Barbante, C., Barnes, P.R., Barnola, J.M., Bigler, M., Castellano, E., Cattani, O., Chappellaz, J., Dahl-Jensen, D., and Delmonte, B. (2004). Eight glacial cycles from an Antarctic ice core. Nature **429,** 623-628.

Bajer, A. (1951). Cytological studies on *Cochlearia tatrae* Borb. Bulletin International de l'Academie Polonaise des Sciences et des Lettres Série B **3,** 89-118.

Benedict, J.C., DeVore, M.L., and Pigg, K.B. (2011). *Prunus* and *Oemleria* (Rosaceae) flowers from the late early Eocene Republic flora of northeastern Washington state, USA. International Journal of Plant Sciences **172,** 948-958.

Bergey, C.M. (2012). vcf-tab-to-fasta. http://code.google.com/p/vcf-tab-to-fasta.

Bock, D.G., Kane, N.C., Ebert, D.P., and Rieseberg, L.H. (2014). Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytologist **201,** 1021-1030.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics**,** btu170.

Burger, G., Gray, M.W., and Lang, B.F. (2003). Mitochondrial genomes: anything goes. Trends in genetics **19,** 709-716.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution **17,** 540-552.

Chang, S., Yang, T., Du, T., Huang, Y., Chen, J., Yan, J., He, J., and Guan, R. (2011). Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in *Brassica*. BMC genomics **12,** 497.

Clapham, A.R., Tutin, T.G., and Moore, D.M. (1990). Flora of the British isles. (CUP Archive).

Comes, H.P., and Kadereit, J.W. (1998). The effect of Quaternary climatic changes on plant distribution and evolution. Trends in plant science **3,** 432-438.

Crane, M.B., and Gairdner, A. (1923). Species-crosses in *Cochlearia*, with a preliminary account of their cytology. Journal of Genetics **13,** 187-200.

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., and Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. American Journal of Botany **99,** 291-311.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., and Sherry, S.T. (2011). The variant call format and VCFtools. Bioinformatics **27,** 2156-2158.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature methods **9,** 772-772.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics **12,** 499-510.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin **19,** 11-15.

Drummond, A.J., Ho, S.Y., Phillips, M.J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS Biology **4,** e88.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution **29,** 1969-1973.

Earl, D.A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources **4,** 359-361.

Egan, A.N., Schlueter, J., and Spooner, D.M. (2012). Applications of next-generation sequencing in plant biology. American journal of botany **99,** 175-185.

Ehrich, D. (2006). AFLPdat: a collection of R functions for convenient handling of AFLP data. Molecular Ecology Notes **6,** 603-604.

Elkington, T. (1984). Cytogenetic variation in the British flora: Origins and significance. New Phytologist **98,** 101-118.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology **14,** 2611-2620.

Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164,** 1567-1587.

Franzke, A., and Mummenhoff, K. (1999). Recent hybrid speciation in *Cardamine* (Brassicaceae)–conversion of nuclear ribosomal ITS sequences in statu nascendi. TAG Theoretical and Applied Genetics **98,** 831-834.

Franzke, A., Koch, M.A., and Mummenhoff, K. (2016). Turnip Time Travels: Age Estimates in Brassicaceae. Trends in plant science **21,** 554-561.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in plant science **16,** 108-116.

Franzke, A., Hurka, H., Janssen, D., Neuffer, B., Friesen, N., Markov, M., and Mummenhoff, K. (2004). Molecular signals for Late Tertiary/Early Quaternary range splits of an Eurasian steppe plant: *Clausia aprica* (Brassicaceae). Molecular Ecology **13,** 2789-2795.

Gao, L., Su, Y.J., and Wang, T. (2010). Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. Journal of Systematics and Evolution **48,** 77-93.

Gill, B., McAllister, H., and Fearn, G. (1978). Cytotaxonomic studies on the *Cochlearia officinalis* L. group from inland stations in Britain. Watsonia **8,** 395-396.

Gill, E. (2008). Conservation genetics of the species complex *Cochlearia officinalis* L. s.l. in Britain. PhD thesis. University of Edinburgh.

Gill, J. (1965). Diploids in the genus *Cochlearia*. Watsonia **6,** 188-189.

Gill, J. (1971). Cytogenetic studies in *Cochlearia* L. The chromosomal homogeneity within both the 2n = 12 diploids and the 2n = 14 diploids and the cytogenetic relationship between the two chromosome levels. Annals of Botany **35,** 947-956.

Gill, J. (1973). Cytogenetic studies in *Cochlearia* L. (Cruciferae). The origins of *C. officinalis* L. and *C. micacea* Marshall. Genetica **44,** 217-234.

Gill, J. (1976). Cytogenetic studies in *Cochlearia* L. (Cruciferae). The chromosomal constitution of *C. danica* L. Genetica **46,** 115-127.

Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic acids research **31,** 5907-5916.

Henrard, J.T. (1915). *Cochlearia hollandica* mihi. Nederlands Kruidkundig Archief **1915,** 200-203.

Heubl, G.R., and Vogt, R. (1985). Chemosystematische Studien in der Gattung *Cochlearia* L. (Cruciferae). (Chemotaxonomical studies in the genus *Cochlearia* L. (Cruciferae).). Botanische Jahrbücher fur Systematik, Pflanzengeschichte und Pflanzengeographie **107,** 177-194.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature **405,** 907-913.

Hewitt, G. (2004). Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London B: Biological Sciences **359,** 183-195.

Hewitt, G.M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. Biological journal of the Linnean Society **58,** 247-276.

Hohmann, N., Wolf, E.M., Lysak, M.A., and Koch, M.A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. The Plant Cell **27,** 2770-2784.

Holland, P. (1981). Pleistocene refuge areas, and the revegetation of Nova Scotia, Canada. Progress in Physical Geography **5,** 535-562.

Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., and Edger, P.P. (2015). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Molecular biology and evolution **33,** 394-412.

Hultén, E. (1937). Outline of the History of Arctic and Boreal Biota During the Quaternary period. Lehre J Cramer, New York.

Hultén, E. (1971). The circumpolar plants. II. Dicotyledons. Kungliga Svenska Vetenskapsakademiens Handlingar, Fjärde Serien **13,** 1–463.

Huson, D.H. (1998). SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14,** 68-73.

Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution **23,** 254-267.

Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23,** 1801-1806.

Jansen, R.K., Raubeson, L.A., Boore, J.L., Chumley, T.W., Haberle, R.C., Wyman, S.K., Alverson, A.J., Peery, R., Herman, S.J., and Fourcade, H.M. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. Methods in enzymology **395,** 348-384.

Jordon-Thaden, I., and Koch, M. (2008). Species richness and polyploid patterns in the genus *Draba* (Brassicaceae): a first global perspective. Plant Ecology & Diversity **1,** 255-263.

Kadereit, J.W., Arafeh, R., Somogyi, G., and Westberg, E. (2005). Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. Taxon**,** 861-876.

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., and Links, M.G. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. The Plant Cell **26,** 2777-2791.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution **30,** 772-780.

Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research **30,** 3059-3066.

Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic acids research **33,** 511-518.

Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., and Koch, M.A. (2013). BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Plant and Cell Physiology**,** pct158.

Knobloch, E., and Mai, D. (1986). Monograph of the fruits and seeds in the Cretaceous of Central Europe. Rozpravy Ústreõdního Ústavu Geologickéh **47,** 1-219.

Knoop, V. (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Current genetics **46,** 123-139.

Knoop, V., Volkmar, U., Hecht, J., and Grewe, F. (2011). Mitochondrial genome evolution in the plant lineage. In Plant mitochondria (Springer), pp. 3-29.

Koch, M. (2002). Genetic differentiation and speciation in prealpine *Cochlearia*: Allohexaploid *Cochlearia bavarica* Vogt (Brassicaceae) compared to its diploid ancestor *Cochlearia pyrenaica* DC. in Germany and Austria. Plant Systematics and Evolution **232,** 35-49.

Koch, M., and Al-Shehbaz, A. (2000). Molecular Systematics of Chinese Genera *Yinshania*, *Hilliella*, and *Cochleariella* (Brassicaceae): Evidence from plastid trnL intron and nuclear ITS DNA sequence data. Annals of the Missouri Botanical Garden **87,** 246-272.

Koch, M., and Al-Shehbaz, I.A. (2002). Molecular data indicate complex intra-and intercontinental differentiation of American *Draba* (Brassicaceae). Annals of the Missouri Botanical Garden**,** 88-109.

Koch, M., Hurka, H., and Mummenhoff, K. (1996). Chloroplast DNA restriction site variation and RAPD-analyses in *Cochlearia* (Brassicaceae): Biosystematics and speciation. Nordic Journal of Botany **16,** 585-603.

Koch, M., Huthmann, M., and Hurka, H. (1998). Isozymes, speciation and evolution in the polyploid complex *Cochlearia* L. (Brassicaceae). Botanica acta **111,** 411-425.

Koch, M., Mummenhoff, K., and Hurka, H. (1999). Molecular phylogenetics of *Cochlearia* (Brassicaceae) and allied genera based on nuclear ribosomal ITS DNA sequence analysis contradict traditional concepts of their evolutionary relationship. Plant systematics and evolution **216,** 207-230.

Koch, M., Dobeš, C., Bernhardt, K., and Kochjarová, J. (2003). *Cochlearia macrorrhiza* (Brassicaceae): A bridging species between Cochlearia taxa from the Eastern Alps and the Carpathians? Plant Systematics and Evolution **242,** 137-147.

Koch, M.A. (2012). Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). Taxon**,** 76-92.

Koch, M.A., Kiefer, C., Ehrich, D., Vogel, J., Brochmann, C., and Mummenhoff, K. (2006). Three times out of Asia Minor: the phylogeography of *Arabis alpina* L.(Brassicaceae). Molecular Ecology **15,** 825-839.

Kochjarová, J., Valachovič, M., Bureš, P., and MrÁZ, P. (2006). The genus *Cochlearia* L. (Brassicaceae) in the Eastern Carpathians and adjacent area. Botanical Journal of the Linnean Society **151,** 355-364.

Lanfear, R., Calcott, B., Ho, S.Y., and Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution **29,** 1695-1701.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. BMC evolutionary biology **14,** 1.

Lang, G. (1994). Quartäre Vegetationsgeschichte Europas: Methoden und Ergebnisse.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27,** 2987-2993.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25,** 2078-2079.

Lopez, L., Wolf, E., Pires, C.J., Edger, P.P., and Koch, M.A. (submitted). Molecular resources for evolutionary studies in the Brassicaceae family using transcriptomes. Frontiers in Plant Science.

Lynch, M. (2009). Estimation of allele frequencies from high-coverage genome-sequencing projects. Genetics **182,** 295-301.

Maddison, W.P., and Knowles, L.L. (2006). Inferring phylogeny despite incomplete lineage sorting. Systematic biology **55,** 21-30.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist **207,** 437-453.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research **20,** 1297-1303.

Moore, M.J., Bell, C.D., Soltis, P.S., and Soltis, D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences **104,** 19363-19368.

Morris, G.P., Grabowski, P.P., and Borevitz, J.O. (2011). Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. Molecular Ecology **20,** 4938-4952.

Muller, J., Muller, K., and Quandt, D. (2010). PhyDE—Phylogenetic Data Editor, version 0.997.

Mummenhoff, K., and Hurka, H. (1995). Allopolyploid origin of *Arabidopsis suecica* (Fries) Norrlin: evidence from chloroplast and nuclear genome markers. Plant Biology **108,** 449-456.

Neuffer, B., and Jahncke, P. (1997). RAPD analyses of hybridization events in *Cardamine* (Brassicaceae). Folia Geobotanica **32,** 57-67.

Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. Molecular Phylogenetics and Evolution **66,** 17-29.

Nordal, I., and Stabbetorp, O. (1990). Morphology and taxonomy of the genus *Cochlearia* (Brassicaceae) in Northern Scandinavia. Nordic journal of botany **10,** 249-263.

Nordal, I., and Laane, M. (1996). Taxonomic delimitation within *Cochlearia officinalis* s. lat. with particular discussion on the rank of *C. anglica* (Brassicaceae). Symbolae Botanicae Upsalienses **31,** 47-57.

Nordal, I., Eriksen, A., Laane, M., and Solberg, Y. (1986). Biogeographic and biosystematic studies in the genus *Cochlearia* in Northern Scandinavia. Symbolae Botanicae Upsalienses **27,** 83-93.

O'Kane Jr, S.L., Schaal, B.A., and Al-Shehbaz, I.A. (1996). The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. Systematic Botany**,** 559-566.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., and Chang, Z. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature **322,** 572-574.

Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biology **7,** 84.

Pegtel, D.M. (1999). Effect of ploidy level on fruit morphology, seed germination and juvenile growth in scurvy grass (*Cochlearia officinalis* L. s.l., Brassicaceae). Plant Species Biology **14,** 201-215.

Petit, J.-R., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J.-M., Basile, I., Bender, M., Chappellaz, J., Davis, M., and Delaygue, G. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature **399,** 429-436.

Petit, R.J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U.M., van Dam, B., and Deans, J.D. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest ecology and management **156,** 49-74.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., and Lareu, M. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. Frontiers in genetics **4,** 98.

Posada, D. (2008). jModelTest: phylogenetic model averaging. Molecular biology and evolution **25,** 1253-1256.

Pritchard, J., Wen, X., and Falush, D. (2010). Documentation for structure software: Version 2.3. University of Chicago, Chicago, IL.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics **155,** 945-959.

Provan, J., and Bennett, K. (2008). Phylogeographic insights into cryptic glacial refugia. Trends in ecology & evolution **23,** 564-571.

Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. Current protocols in bioinformatics**,** 11.12. 11-11.12. 34.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26,** 841-842.

Rambaut, A., Suchard, M., Xie, D., and Drummond, A. (2014). Tracer v1. 6, Available from http://beast.bio.ed.ac.uk/Tracer.

Reboud, X., and Zeyl, C. (1994). Organelle inheritance in plants. Heredity **72,** 132-140.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. science **298,** 2381-2385.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., and Burleigh, J.G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. BMC Evolutionary Biology **14,** 23.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Research **6,** 283-290.

Saunte, L.H. (1955). Cytogenetical studies in the *Cochlearia officinalis* complex. Hereditas **41,** 499-515.

Schatz, M.C., Delcher, A.L., and Salzberg, S.L. (2010). Assembly of large genomes using second-generation sequencing. Genome research **20,** 1165-1173.

Schönswetter, P., Popp, M., and Brochmann, C. (2006). Rare arctic-alpine plants of the European Alps have different immigration histories: the snow bed species *Minuartia biflora* and *Ranunculus pygmaeus*. Molecular Ecology **15,** 709-720.

Schönswetter, P., Paun, O., Tribsch, A., and Niklfeld, H. (2003). Out of the Alps: colonization of Northern Europe by East Alpine populations of the glacier buttercup *Ranunculus glacialis* L. (Ranunculaceae). Molecular ecology **12,** 3373-3381.

Schranz, M.E., and Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. Plant Cell **18,** 1152-1165.

Schranz, M.E., Lysak, M.A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends in Plant Science **11,** 535-542.

Schulz, R. (2015). Evolutionary insights from cytogenetic analyses: the case of *Cochlearia* L. Master thesis. Universität Heidelberg.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nature biotechnology **26,** 1135-1145.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., and Yamaguchi-Shinozaki, K. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. The EMBO journal **5,** 2043.

Sims, H.J., Herendeen, P.S., Lupia, R., Christopher, R.A., and Crane, P.R. (1999). Fossil flowers with Normapolles pollen from the Upper Cretaceous of southeastern North America. Review of Palaeobotany and Palynology **106,** 131-151.

Skrede, I., Eidesen, P.B., Portela, R.P., and Brochmann, C. (2006). Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). Molecular Ecology **15,** 1827-1840.

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. Journal of theoretical biology **261,** 58-66.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics**,** btu033.

Stebbins, G. (1984). Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. Botanica Helvetica **94,** 1-13.

Steele, P.R., and Pires, J.C. (2011). Biodiversity assessment: State-of-the-art techniques in phylogenomics and species identification. American Journal of Botany **98,** 415-425.

Stewart, J.R., and Lister, A.M. (2001). Cryptic northern refugia and the origins of the modern biota. Trends in Ecology & Evolution **16,** 608-613.

Straub, S.C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany **99,** 349-364.

Straub, S.C., Fishbein, M., Livshultz, T., Foster, Z., Parks, M., Weitemier, K., Cronn, R.C., and Liston, A. (2011). Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC genomics **12,** 1.

Stull, G.W., de Stefano, R.D., Soltis, D.E., and Soltis, P.S. (2015). Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. American journal of botany **102,** 1794-1813.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., and Cosson, J.-F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology **7,** 453-464.

Takahashi, M., Crane, P.R., and Ando, H. (1999). *Esgueiria futabensis* sp. nov., a new angiosperm flower from the Upper Cretaceous (lower Coniacian) of northeastern Honshu, Japan. Paleontological Research **3,** 81-87.

Templeton, A.R., Crandall, K.A., and Sing, C.F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132,** 619-633.

Tribsch, A., and Schönswetter, P. (2003). Patterns of endemism and comparative phylogeography confirm palaeo-environmental evidence for Pleistocene refugia in the Eastern Alps. Taxon **52,** 477-497.

Tribsch, A., Schönswetter, P., and Stuessy, T.F. (2002). *Saponaria pumila* (Caryophyllaceae) and the ice age in the European Alps. American journal of botany **89,** 2024-2033.

Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nature genetics **15,** 57-61.

Vogt, R. (1985). Die *Cochlearia pyrenaica*-Gruppe in Zentraleuropa. Berichte der Bayerischen Botanischen Gesellschaft **56,** 5-52.

Webb III, T., and Bartlein, P. (1992). Global changes during the last 3 million years: climatic controls and biotic responses. Annual Review of Ecology and Systematics **23,** 141-173.

Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., and Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Sciences **2,** 1400042.

Whittall, J.B., Syring, J., Parks, M., Buenrostro, J., Dick, C., Liston, A., and Cronn, R. (2010). Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. Molecular Ecology **19,** 100-114.

Wicke, S., Schneeweiss, G.M., Müller, K.F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant molecular biology **76,** 273-297.

Wolfe, K.H., Li, W.-H., and Sharp, P.M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences **84,** 9054-9058.

Zmudczyńska-Skarbek, K., Barcikowski, M., Zwolicki, A., Iliszko, L., and Stempniewicz, L. (2013). Variability of polar scurvygrass *Cochlearia groenlandica* individual traits along a seabird influenced gradient across Spitsbergen tundra. Polar Biology **36,** 1659-1669.

Chapter 3 – *Cochlearia* Metabolomics

3.1 Abstract

Chapter 3 describes the analysis of the *Cochlearia* cold metabolome via metabolite profiling using gas chromatography-mass spectrometry (GC-MS) in order to detect putative intrageneric variation among metabolomic phenotypes as a possible result of adaptation to local environments.
Therefore, plants taken from 27 populations, covering large parts of the genus' distribution range, were subjected to either a 20 days cold treatment at 5°C or kept under control conditions at 20°C. Four climatic ecotypes were defined based on a hierarchical clustering of 5 temperature-related bioclimatic variables retrieved from the WorldClim database and used as group priors in the metabolite analyses. Plants from all bioclimatic clusters showed a strong metabolomic reaction to the cold treatment and as expected, especially concentrations of carbohydrates and amino acids were significantly increased with slight differences between the four bioclimatic clusters.
As revealed via discriminant analysis of principal components (DAPC), there is considerable overlap between the four clusters especially in the cold response, yet slight variation in the metabolomic responses to the different conditions and the metabolites contributing to this variation could be detected. According to this, the metabolomic phenotype of the central European alpine samples from the High Tatra and Northern Scandinavia is more similar to the arctic cluster under cold conditions than under warm conditions while at the same time showing strong overlap with the other two other central European clusters. This is interpreted either as a putative result of introgression from the physiological phenotype of diploid arctic samples into parts of the Central European phenotypes or as a putatively repeated evolution of cold adaptation facilitated by standing genetic variation.
However, further analyses are needed in order to assess the statistical significance of the metabolomic insights presented herein. If the revealed bioclimatic/physiological characteristics can be confirmed, then subsequent population-genomic studies might likely identify traces of adaptive molecular variation together with genes and pathways involved in the cold adaptation of the genus *Cochlearia*. Moreover, including the metabolomic phenotypes of the Mediterranean sister-genus *Ionopsidium* might reveal additional insights into these fascinating evolutionary processes.

3.2 Introduction

3.2.1 New approaches to study plant responses to abiotic stresses and adaptation

The existing variation between different natural environments both on a spatial as well as on a temporal scale is subjecting plants to a variety of abiotic stresses which have a strong impact on their development, growth and productivity (Urano et al., 2010). Thus, there is great interest in how plants respond to these stresses from a scientific point of view but also for economic reasons for example in crop development and production (e.g. Berry and Bjorkman, 1980; Thomashow, 1999; Jump and Penuelas, 2005; Oh et al., 2007; Fernandes et al., 2008; Li et al., 2008). In light of the current global change, the plastic and evolutionary responses of both plants and animals to novel or changing environments have gained new relevance (e.g. Thomas et al., 2004; Visser, 2008; Franks and Hoffmann, 2012; Franks et al., 2014).

Especially in immobile terrestrial plants, climate has been shown to be one of the central factors shaping the geographical distribution of taxa (e.g. Walther et al., 2002; Walther et al., 2005) and forcing adaptive evolution (e.g. Fournier-Level et al., 2011; Hancock et al., 2011; Zhou et al., 2014). Likewise, edaphic adaptation to special or contrasting soil conditions and its role on plant distribution and diversification has long been studied intensively, e.g. regarding the adaptation to serpentine and non-serpentine soils (e.g. Bratteler et al., 2006; Turner et al., 2010) or the influence of calcicole/calcifuge bedrock preferences on biogeography and diversity in alpine systems (e.g. Conti et al., 1999; Alvarez et al., 2009). Despite the strong interest in plant adaptation to local climatic conditions and contrasting environments (reviewed by Savolainen et al., 2013), the genetic basis of these processes is not yet fully understood. Detecting and understanding past events of local adaptation to extreme ecological conditions can improve the predictions of how far and how fast adaptation to the present environmental change might be possible in the future (Hancock et al., 2011; Franks et al., 2014) and thus, there is a strong interest in unravelling the molecular mechanisms underlying adaptation, the genes involved in these processes and the evolutionary forces that shape the genetic variation required for adaptation (Stapley et al., 2010).

With the rise of the new 'omics' approaches, and especially with the technological progress in high-throughput sequencing, biologists have gained new tools to study the adaptation of living organisms to different environmental stresses (reviewed by Urano et al., 2010), thereby also addressing some of the very basic questions in evolutionary biology that could not have been answered so far, such as the genetic basis of convergent evolution (reviewed by Nadeau and Jiggins, 2010). Over the last years, genome-wide high-throughput sequencing data analyses provided a wealth of valuable insights into the mechanisms of plant abiotic stress responses and adaptation (e.g. Brachi et al., 2010; Turner et al., 2010; Hancock et al., 2011) via both reverse genetics approaches, focusing on candidate genes known to be involved in the respective pathways or stress responses, as well as via forward genetics approaches, performing genome scans in search of outlier loci of interest (reviewed e.g. by Nadeau and Jiggins, 2010; Oleksyk et al., 2010; Stapley et al., 2010). At the same time, this genomic information needed to be linked with gene functions and phenotypes, a gap that was filled by the so called functional genomics approaches including

transcriptomics, proteomics and metabolomics (for reviews see e.g. Somerville and Somerville, 1999; Fiehn et al., 2000; Hall et al., 2002; Holtorf et al., 2002; Chen and Harmon, 2006; Bräutigam and Gowik, 2010). Based on the different 'omics' methods, several regulatory networks responsible for abiotic stress responses have meanwhile been analyzed in more detail (reviewed by Urano et al., 2010), yet many open questions remain to be answered.

Metabolomics studies, the high-throughput analyses of an organism's metabolome via mass spectrometry, are a comparatively young but very promising way to connect the genotype with the phenotype of a study system (Fiehn et al., 2000; Fiehn, 2002) and thus they can shed new light on the impact of both environmental and genetic changes on biological mechanisms (for reviews see Hall et al., 2002; Dunn, 2008; Dunn et al., 2013). Similar to the rapid progress in next-generation sequencing, high-throughput methods for analyzing the metabolism of e.g. plants (for reviews see Schauer and Fernie, 2006; Last et al., 2007) and animals (e.g. Dunn et al., 2011; Mushtaq et al., 2014) have been developed rapidly over the last two decades. The metabolic profiles gained in these studies represent a metabolic phenotype linked to a range of morphological or physiological characteristics, for example to plant biomass production as shown by Meyer et al. (2007) on *Arabidopsis thaliana*, and there is great interest in analyzing the impact of abiotic stresses on a plant's metabolism (e.g. Kaplan et al., 2004; for reviews see Shulaev et al., 2008; Urano et al., 2010).

3.2.2 Edaphic and cold adaptation in arctic/alpine systems and the potential of the *Cochlearia* study system

As illustrated by Figure 1 in the general introduction, the genus *Cochlearia* displays striking ecological associations of lowland taxa with calcareous bedrocks and arctic/alpine taxa with siliceous bedrocks, and as revealed from phylogenomic analyses presented in chapter 2, (high-)alpine taxa, found on siliceous bedrocks in different Central European alpine environments, putatively evolved several times in parallel, likely from neighboring lowland/mountainous taxa on limestone bedrocks and thus assuming repeated niche shifts at the interspecific level as described for several other plant genera before (e.g. in *Saxifraga* (Conti et al., 1999) or *Gentiana* sect. *Ciminalis* (Hungerer and Kadereit, 1998)). Yet at the intraspecific level, the clear distinction between calcicole and calcifuge plant species was suggested to be a result of historical biogeography, thereby assuming niche stability during postglacial recolonization (Alvarez et al., 2009). However, neither the genomic basis nor the ecophysiological background of these phenomena are fully understood by now. Aside from edaphic adaptation, given the suggested "arctic background" of the Central European *Cochlearia* taxa, as indicated from phylogenomic analyses in chapter 2, one hypothesis would be that the repeated adaptation to alpine habitats was facilitated by a strong pre-adaptation to cold present as standing genetic variation. Therefore, by making use of the newly available genomic tools, such as genome-scans in population genomic approaches, the *Cochlearia* study system offers excellent opportunities to study general evolutionary mechanisms behind cold and edaphic adaptation, parallel adaptation and likely adaptation from standing genetic variation, thereby putatively also shedding new light on the formation of spatial biodiversity patterns in the European alpine systems. Yet, as stated above,

phenotypic information is needed in order to tap the full potential of these genomic approaches, and thus, as a starting point for any further in-depth studies within the cold-adapted genus *Cochlearia*, the objective of chapter 3 was to investigate its low-temperature metabolome, in order to define temperature-related metabolomic phenotypes via metabolite profiling and to detect putative variation in the metabolic response to temperature stress.

Especially based on the insights gained from novel omics approaches, it has become clear that cold tolerance is a very complex trait both from a morphological and a physiological point of view (e.g. Fowler and Thomashow, 2002; reviewed by Preston and Sandve, 2013). Although the molecular framework of the adaptation to sub-zero temperatures in plants is not yet completely unraveled, several effects of cold induction are already known and could be part of the plant's cold perception as a first step for initiating a cold response. Among them are (adapted from Smallwood and Bowles, 2002): 1) a change in membrane fluidity 2) effects on secondary structures of RNA/DNA molecules 3) altered enzyme activity 4) early changes in photosynthetic activity and the plant metabolism. During cold acclimation, the process of gradually adapting to low nonfreezing temperatures leading to acquired freezing tolerance, drastic metabolomic reconfiguration has been revealed for *A. thaliana* (Cook et al., 2004; Kaplan et al., 2004) with a central role of the C-repeat binding factors (CBF; Gilmour et al., 1998) and increased concentrations in the majority of all analyzed metabolites (75% according to Cook et al. (2004)), especially in carbohydrates and amino acids. It has been suggested that many of the increased compounds could act as signaling molecules and/or as compatible solutes, especially amino acids, such as alanine, glycine, proline, serine, and soluble carbohydrates like fructose, glucose, maltose and sucrose among others (Kaplan et al., 2004). Analyzing the metabolic responses to cold temperatures on a spatial scale, as also done by Davey et al. (2009) for three geographically and climatically distinct populations of *Arabidopsis lyrata* subsp. *petraea*, might reveal spatial variation of the *Cochlearia* cold metabolome as a possible result of local adaptation and could provide new insights into cold acclimation and cold tolerance.

3.3 Material and methods

3.3.1 Plant material and taxon sampling

A total of 27 populations were selected for the metabolomic analyses (Table III-1). These include 13 *Cochlearia* species/subspecies and cover large parts of the distribution range including arctic regions (Figure III-1). If possible, four plants per population were analyzed, but due to limited amounts of seed material/living plants, fewer samples were considered for some of the populations. The 107 plants selected for metabolomic analyses where either collected in the wild or grown from seed material and all plants were cultivated in a substrate consisting of seedling potting soil, quartz sand, and either composted earth or a peat based substrate (Ökohum GmbH, Herbertingen, Germany) under greenhouse conditions in the Botanical Garden Heidelberg until the beginning of the experiment (see 3.3.3 Cold treatment).

Table III-1. Samples included in the metabolomic analyses with information on origin and analyzed sample number (No. Ind.).

Species	Pop.-no.	Locality		Lat	Lon	alt	No. ind.
C. sessilifolia	221	Alas.	Kodiak Island, Anton Larson Bay	57.9	-152.6	1 m	2
C. islandica	210	IS	Stokkseyri (South Coast)	63.8	-21.0	0 m	4
C. tatrae	145	SK	W Carpathians, Tatra Mts., Velicka Dolina	49.2	20.2	1765 m	6
C. alpina	148	UK	Scotland: Teesdale	54.7	-2.4	621 m	3
C. alpina	153	UK	Scotland: Cairngorm Mountains; Corrie an't snechda	57.1	-3.7	1134 m	4
C. alpina	203	UK	Scotland, Moffat, Hartfell Rig	55.4	-3.4	714 m	4
C. micacea	162	UK	Scotland: Beinn an Dothaidh	56.5	-4.7	743 m	4
C. micacea	163	UK	Scotland: Meall nan Tarmachan	56.5	-4.3	590 m	4
C. officinalis	204	UK	Scotland: Coast south of Ayr; Dunure	55.4	-4.8	4 m	4
C. officinalis	169	NO	Trs, Storfjord, Skibotn	69.4	20.3	289 m	4
C. scotica	156	UK	Scotland: Burghead, coast	57.7	-3.5	5 m	4
C. x hollandica	207	UK	Scotland: Sandyhills, south of Dumfries	54.9	-3.7	6 m	4
C. x hollandica	206	UK	Scotland: north of Creetown	54.9	-4.4	11 m	4
C. danica	152	UK	Scotland: Anstruther harbor	56.2	-2.7	1 m	4
C. danica	269	DE	BW, Rheinstetten, along highway B36	49.0	8.3	118 m	4
C. scotica	159	UK	Scotland: Achmelvich, salt marsh	58.2	-5.3	8 m	4
C. scotica	161	UK	Scotland: Ballachulish, salt marsh	56.7	-5.2	2 m	4
C. pyrenaica	199	UK	England: Malham Tarn Fen	54.1	-2.2	380 m	4
C. pyrenaica	200	UK	England: Ribblesdale, Ingleborough, Rakespring	54.2	-2.4	676 m	3
C. pyrenaica	201	UK	England, Alston, Nenthead	54.8	-2.3	449 m	4

C. pyrenaica	33	SK	Carpathians, Veľká Fatra National Park, Jazierce	49.0	19.3	555 m	3
C. pyrenaica	212	ES	Asturias, between Villar de Vildas and La Pornacal	43.1	-6.3	1000 m	4
C. pyrenaica	32	AT	Lower Austria, Scheibbs, Lunz, Lehen	47.8	15.0	576 m	4
C. macrorrhiza	21	AT	Lower Austria, Moosbrunn	48.0	16.4	185 m	4
C. bavarica	193	DE	Bavaria, Glonn, river Glonn	48.0	11.9	556 m	3
C. bavarica	267	DE	Bavaria, Katzbrui near Unteregg	48.0	10.5	683 m	4
C. borzaeana	140	RO	Carpathians, Distr. Maramures, Muntii	47.7	25.0	1226 m	4

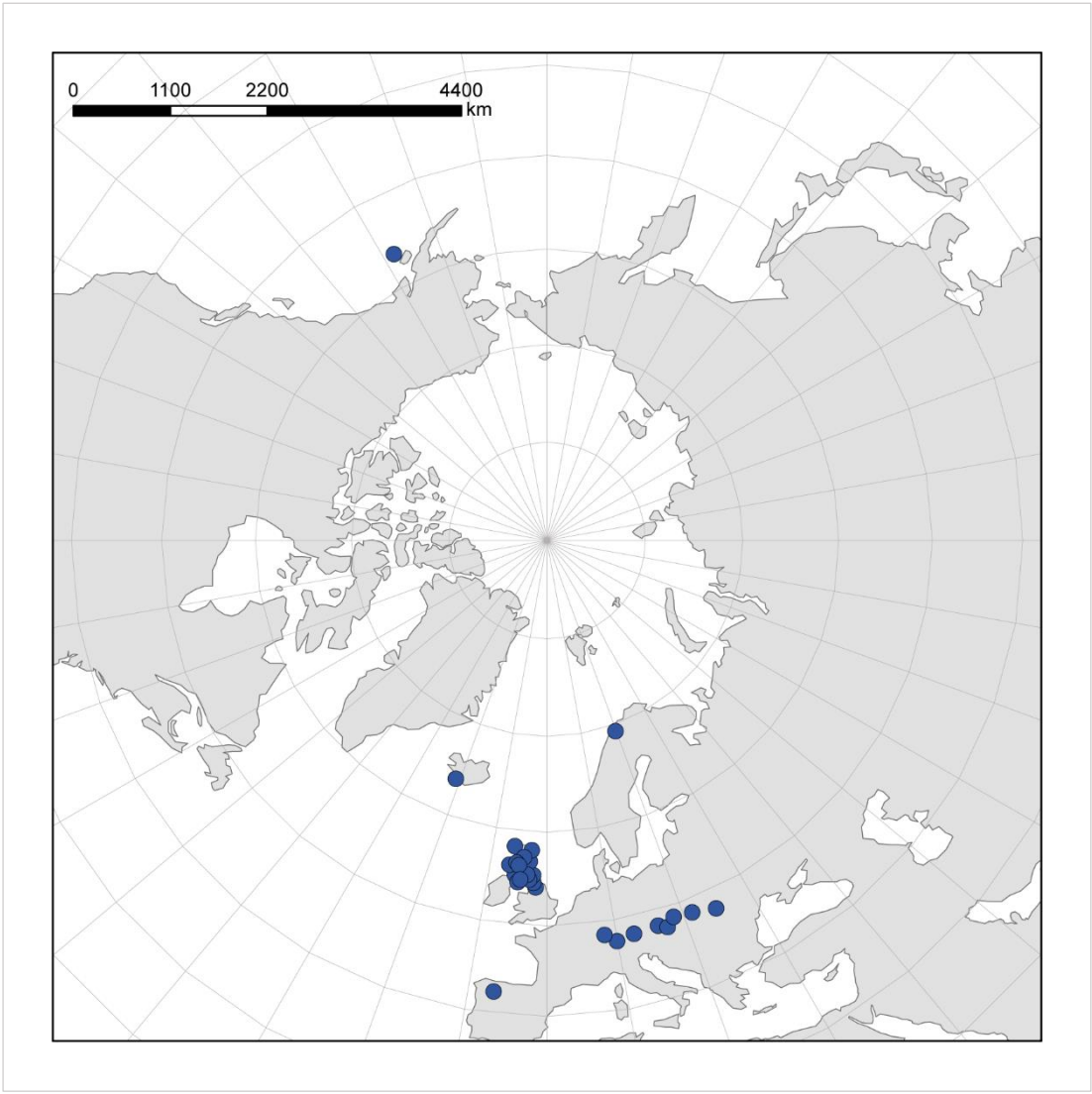

Figure III-1. Geographical distribution of 27 populations, comprising 13 different *Cochlearia* taxa, considered for metabolomic analyses.

3.3.2 Habitat characterization and ecotype definition

In order to analyze the phenotypic variation within the cold metabolomes as a possible result of adaptation to local environments, bioclimatic variables were downloaded from the high-resolution climate data WorldClim grids (http://www.worldclim.org, Hijmans et al., 2005) at a resolution of 30 arcseconds (~1 km² / pixel). For all 27 populations included in the metabolomic analyses (with given or approximated coordinates), the standard 19 topo-climatic factors describing precipitation (bioclimatic variables 1-11) and temperature (bioclimatic variables 12-19) were obtained using the R packages 'raster' (Hijmans, 2016b) and 'geosphere' (Hijmans, 2016a) and following an R script written by F. Michling (unpublished; Supplemental Data Set III-1). The full list of all 19 bioclimatic variables for each analyzed population is given with Supplemental Data Set III-2. Climatic ecotypes of each population were finally defined based on the temperature-related bioclimatic variables *Annual Mean Temperature* (BIO1), *Max Temperature of Warmest Month* (BIO5), *Min Temperature of Coldest Month* (BIO6), *Mean Temperature of Warmest Quarter* (BIO10) and *Mean Temperature of Coldest Quarter* (BIO11) that provide a good estimation of climatic differences across the distribution range of the genus. The selected variables are likely to be of relevance in local cold adaptation and might thus have an association with variation in the cold metabolism of *Cochlearia*.

Based on the five chosen temperature-related bioclimatic variables of each population (given with Supplementary Data Set III-2, sheet 3 'all_pops_subset1'), a hierarchical cluster analysis was performed in R version 3.3.1. After adding a constant value (+150) to the original temperature data points, this way avoiding negative values, the data was log transformed using the log() function and scaled using the scale() function. Hereafter, Euclidian distances between observations were computed using the dist() function (method = "euclidian"). Hierarchical clustering was then applied using the function hclust() from the 'stats' package according to the Ward method (method = "ward.D2") (Ward Jr, 1963), which has been shown to be an appropriate method for cluster analysis on climate data (Unal et al., 2003). The R package 'NbClust' (Charrad et al., 2014) with the argument 'index' set to "all" (26 computed clustering indices), was used to evaluate the optimal number of clusters in the dataset. Finally, the cluster dendrogram was visualized using the R package 'factoextra' (Kassambara, 2015).

3.3.3 Temperature treatment

All 107 plants selected for metabolomic analysis were initially kept in a plant room under a 16/8 h day/night, 20°C/18°C day/night cycle for two weeks in order to acclimatize them to the same conditions. After this time, leaf samples of every plant were collected, frozen in liquid nitrogen and stored at -80°C until metabolomic analysis. Now, half of the plants from each population were transferred to a cold chamber with a 16/8 h day/night, 5°C day/night cycle, while the remaining plants stayed in control conditions of 20°C/18°C. After 20 days, leaf samples were harvested again from all plants under cold and control conditions, frozen in liquid nitrogen and transferred to -80°C. The experimental setup is illustrated in Figure III-2.

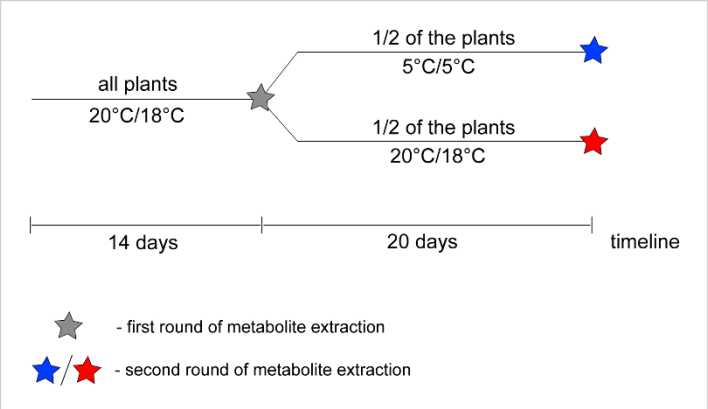

Figure III-2. Experimental setup of the temperature treatment with a timeline for metabolite extractions.

3.3.4 Metabolite extraction and analyses

All metabolomic measurements and annotations were carried out by Dr. Emmanuel Gaquerel at the Metabolomics Core Technology platform Heidelberg. Metabolite profiling was performed using gas chromatography-mass spectrometry (GC-MS) and largely following the primary metabolites extraction steps as described by Roessner et al. (2001). Therefore, 15 – 40 mg of the previously collected and frozen leaf material were homogenized by grinding in liquid nitrogen and hereafter mixed with 360 µL cold methanol. 20 µg of ribitol were added as an internal normalizing standard. After extracting the sample for 15 min at 70°C, it was mixed thoroughly with 200 µL chloroform and 400 µL water and centrifuged subsequently. 200 µL of the methanol-water upper phase containing the primary metabolites were taken off and concentrated to dryness in a vacuum concentrator. A two-step derivatization procedure including methoximation of the dried residue followed by silylation was performed as described by Lisec et al. (2006). To this end, the residue was first re-suspended in a methoxyamine-hydrochloride/pyridine solution for a methoxymization of the carbonyl groups. The sample was then heated for 90 min at 37°C and further silylated with N-methyl-N-trimethylsilyltrifloracetamide at 37°C for 30 min.

GC-MS analysis was performed on a gas chromatograph system equipped with quadrupole mass spectrometer (GC-MS-QP2010, Shimadzu, Duisburg, Germany). For this, 1 µL of each sample was injected in split mode with a split ratio of 1:20 and the separation of derivatized primary metabolites was carried out on a RTX-5MS column (Restek Corporation, Bellefonte, Pennsylvania, USA). Two injection replicates were performed, resulting in a total of 214 samples and 20 blanks in order to assess the instrumental parameter stability. Metabolites were detected using optimized instrumental settings as described by Lisec et al. (2006) and annotated by using a two-pronged approach. Obtained raw data files were converted into an ANDI-MS universal file format for spectrum deconvolution and compound identification via the reference collection of the Golm Metabolome Database (GMD, http://gmd.mpimp-golm.mpg.de/) using the AMDIS program (Automated Mass Spectral Deconvolution and Identification System; www.amdis.net/) for GC-MS data interpretation. Kovats retention indices were calculated for deconvoluted mass spectra from measurements of an alkane mixture and hereafter compared with best hits obtained via the

GMD database. The Shimadzu GCMS solutions software (v2.72) interface was used for a targeted metabolite annotation versus an in-house library of authentic standards that were analyzed under the above analytical conditions.
CSV files were produced for peak areas obtained for quantifier ions selected for target compounds following the annotation procedure as described above. In total, 40 compounds (Table III-2) were selected that were directly annotatable as known compounds or as unknown compounds that were reproducibly detected across all analyzed samples. Several carbohydrates could not be identified with certainty, due to general difficulties in carbohydrate analysis by GC-MS, and are thus denoted as 'unidentified hexose peak 1', 'unidentified hexose peak 2' and 'disaccharide' respectively. Scaling of peak areas was done according to the extracted amount of leaf tissue in each measurement and normalization was based on weight and the area of ribitol which served as a normalizing internal standard, this way accounting for putative technical deviations across the different measurement batches. In general, this approach does not access absolute levels of quantification. However, it is possible to compare variations in normalized levels of primary metabolites across samples. Moreover, for these normalized levels, z-score normalized data was generated in order to bring all detected variables on the same scale, this way avoiding disturbance caused by high intensity variables.

Table III-2. List of 40 selected metabolic compounds sorted by general compound classes. Compounds were selected for integration based on annotatability and reproducibility of peak detection throughout the dataset.

Compound class	**Selected target compounds**
Alkaloids	Pseudo-tropine, Tropine
Amino acids	Alanine, Asparagine, Aspartic acid, Glutamic acid, Glycine, Isoleucine, Proline, Serine, Threonine, Valine
Carbohydrates	Fructose, Unidentified hexose peak 1, Glucose, Glucose,1,6-anhydro-beta, Mannose, *myo*-Inositol, Disaccharide, Unidentified hexose peak 2, Sucrose
Fatty acids	Hexanoic acid, Octadecanoic acid
Organic acids	2-Ketoglutaric acid, Ascorbic acid, 4-amino butyrate, Caffeic acid, Citric acid, Dehydroascorbic acid, Ferulic acid, Fumaric acid, Glyceric acid, Malic acid, Oxalic acid, Phosphoric acid, Threonic acid
Miscellaneous	1-pyrroline-3-hydroxy-5-carboxylic acid, Ethanolamine
Unknown	RT:19.905 unknown, RT:20.075 unknown

3.3.5 Multivariate statistical analysis of metabolomics and integration of bioclimatic population clusters

Using the aov() function embedded in the R package 'stats', a one-way Analysis of Variance (ANOVA) was performed on normalized metabolite data grouped by bioclimatic clusters as revealed via hierarchical cluster analysis (see Figure III-3) in order to test for statistically significant differences in the metabolite concentrations between control and cold treatment in the different bioclimatic clusters.
In addition, a discriminant analysis of principal components (DAPC; Jombart et al., 2010) was carried out taking the previously determined bioclimatic population clusters as group priors. DAPC

was chosen as a consequence of missing discriminating structure and explanatory power in initially performed PCA analyses (data not shown). This method, though primarily designed for analyzing large sequencing datasets of genetically structured populations based on multivariate statistical methods (Jombart et al., 2010), promised to be a good option for assessing – if present – the variation in the cold response between populations/clusters and to identify metabolites contributing most significantly to this variation. The strength of the DAPC lies in the combination of a principal components analysis (PCA) as an initial step of data transformation followed by a discriminant analysis (DA) based on the determined principal components (PCs). In contrast to a stand-alone PCA, DA aims at maximizing the variation between clusters and minimizing the variation within (Jombart et al., 2010). A DAPC circumvents methodical limitations of the DA via the initial PCA and retains the analytical strength of a DA in discriminating pre-defined clusters (Jombart et al., 2010).

The DAPC was performed in R using the function dapc() from the package 'adegenet' (Jombart, 2008; Jombart and Ahmed, 2011). Normalized input data (based on weight and ribitol as in internal standard) was used for all DAPC runs (given with Supplemental Data Set III-7) and variables were centered and scaled and as suggested by Jombart and Collins (2015). In order to find the optimal number of PCs to retain in each respective DAPC analysis, a cross-validation for discriminant analysis of principal components was performed using the xvalDapc() function embedded in the 'adegenet' package prior to each DAPC analysis. XvalDapc() performs replicated DAPCs based on varying numbers of retained PCs and was carried out on a training set consisting of 90% of the data. The remaining 10% of the data served as a validation set to determine the outcome success of the predicted individual posterior group membership probabilities under the respective level of retained PCs. XvalDapc() was performed with 1000 replicates for each number of PC retention and the optimal number of PCs was determined based on the lowest root mean squared error associated with the predictive success. DAPC was performed based on respective group priors (see below) and with normalized input data centered and scaled. The determined number of PCs and all discriminant functions were retained. The R function scatter() as embedded in the 'ade4' package (Dray and Dufour, 2007) was used to produce scatterplots of the DAPC results and the function loadingplot() from the 'adegenet' package was used to plot the contribution of the different primary metabolites to the respective discriminant functions. Contributions above a threshold of 0.07 were highlighted for future analyses and interpretation. Following Jombart and Collins (2015), additional scatterplots were generated with minimum spanning trees based on the (squared) distances between the four clusters within the entire space, thus illustrating actual proximities between the clusters inside the entire space.

A first DAPC was performed with groups defined by treatment ("control 20°C samples" – first measurement of the samples chosen for the 20 days control treatment; "control 5°C samples" – first measurement of the samples chosen for cold treatment; "control treatment" – (second) measurement of the samples kept under 20°C; "cold treatment" – measurement of the samples after 20 days of cold treatment) in order to detect the general metabolomic response to the different conditions. This analysis was then further specified by running the DAPC under a grouping of 'treatment x bioclimatic cluster' – with the bioclimatic clusters as retrieved via hierarchical clustering. Hereafter, group priors were changed to represent the bioclimatic clusters and DAPC

was performed for all data points (before and after treatment) as well as separately for control and cold treatment only.

3.4 Results

3.4.1 Climatic data and habitat characterization

As expected, the analysis of the five selected temperature-related bioclimatic variables revealed differences in the climatic conditions for the 27 population localities included in the metabolomic study with a range of the mean annual temperature from 0.8°C in northern Norwegian *C. officinalis* (Pop.-no. 169) to 10.4°C in a German inland population of *C. danica* (Pop.-no. 269). Average temperatures of the coldest month were not found to be lowest in the two arctic populations but in an alpine population of *C. tatrae* (Pop.-no. 145) with -10.7°C and the northern Norwegian population of *C. officinalis* (Pop.-no. 169) with -12.2°C. The Icelandic population of *C. islandica* (Pop.-no. 210) experiences the smallest annual temperature range, and the highest average temperature of the warmest month was detected for the Austrian lowland population of *C. macrorrhiza* (26°C; Pop.-no. 21), a species that is meanwhile extinct in the wild. A detailed documentation of the extracted topo-climatic factors is given with Supplemental Data Set III-2.

Hierarchical clustering of climate data

According to the majority rule, the optimal number of clusters was determined to be 3 using the 'NbClust' package in R (Charrad et al., 2014). However, based on the interpretation of the dendrogram produced in the hierarchical clustering (Figure III-1), an optimum of 4 clusters was chosen for further statistical analyses, as suggested by "kl" criterion (Krzanowski and Lai, 1988), "Hartigan" criterion (Hartigan, 1975) and the graphical method "dindex" (Lebart et al., 2000), where the optimal number of clusters was identified by peak evaluation of the D index second differences plot (given with Supplemental Data Set III-3). Results of the NbClust analysis are given with Supplemental Data Set III-4.

The clustering dendrogram with four selected clusters is shown in Figure III-1. Cluster 1 (hereafter also referred to as "coastal cluster") is composed of polyploid coastal populations with one exception, namely an English population of diploid *C. pyrenaica* (Pop.-no. 199). Cluster 2 (hereafter also referred to as "inland cluster") includes Central European inland locations of both diploid and polyploid taxa. Cluster 3 (hereafter also referred to as "alpine cluster") is a cluster composed of three locations from continental Europe with an arctic/alpine habitat type. One of them is a population of the high alpine *C. tatrae* from the High Tatra Mountains (Pop.-no. 145), the others represent *C. borzaeana* from the Carpathians (Pop.-no. 140) and the North-Norwegian population of *C. officinalis* (Pop.-no. 169), located north of the Arctic Circle. Finally, cluster 4 (hereafter also referred to as "arctic/alpine cluster") contains the two arctic locations in Iceland (*C. islandica*, 2n=12; Pop.-no. 210) and Alaska (*C. sessilifolia*, 2n=14; Pop.-no. 221) respectively as well as locations of diploid and tetraploid taxa from alpine habitat types in the UK. The label "warm" in the cluster dendrogram (Figure III-3) chosen for clusters 1 and 2 indicates that both clusters encompass coastal and inland locations with moderately warmer temperatures compared to clusters 3 and 4 composed of arctic and alpine locations, yet clusters 1 and 2 also represent cold-

characterized habitat types of the generally cold-adapted *Cochlearia* taxa. Figure III-4 illustrates the geographical distribution of the four clusters.

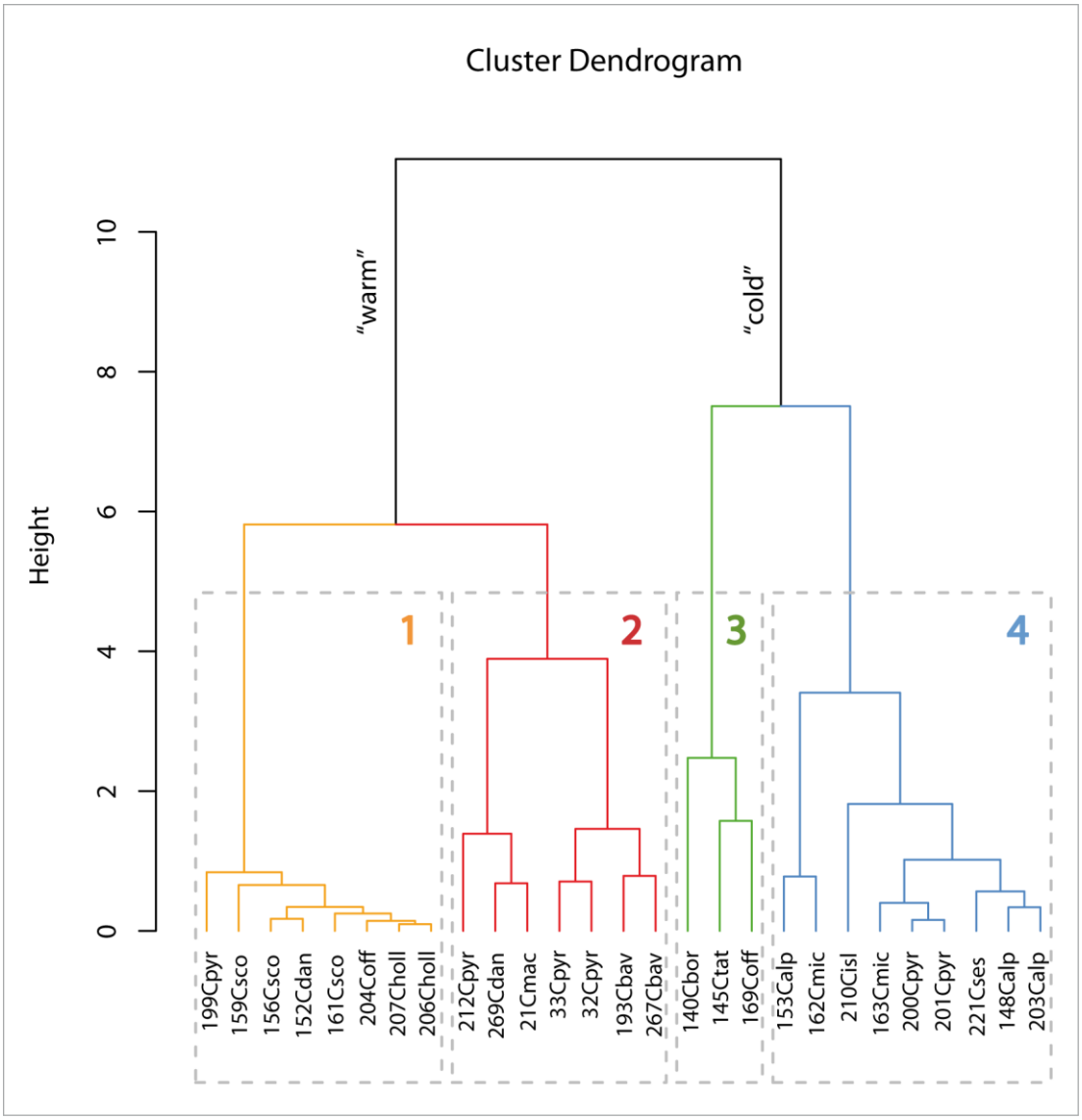

Figure III-3. Hierarchical cluster analysis based on 5 temperature-related bioclimatic variables (*Annual Mean Temp.*, *Max Temp. of Warmest Month*, *Min Temp. of Coldest Month*, *Mean Temp. of Warmest Quarter* and *Mean Temp. of Coldest Quarter*) and 27 populations included in the metabolomic analyses. Tip labels in the bioclimatic cluster dendrogram code for population number and species name. The four retained clusters are represented by different colors.

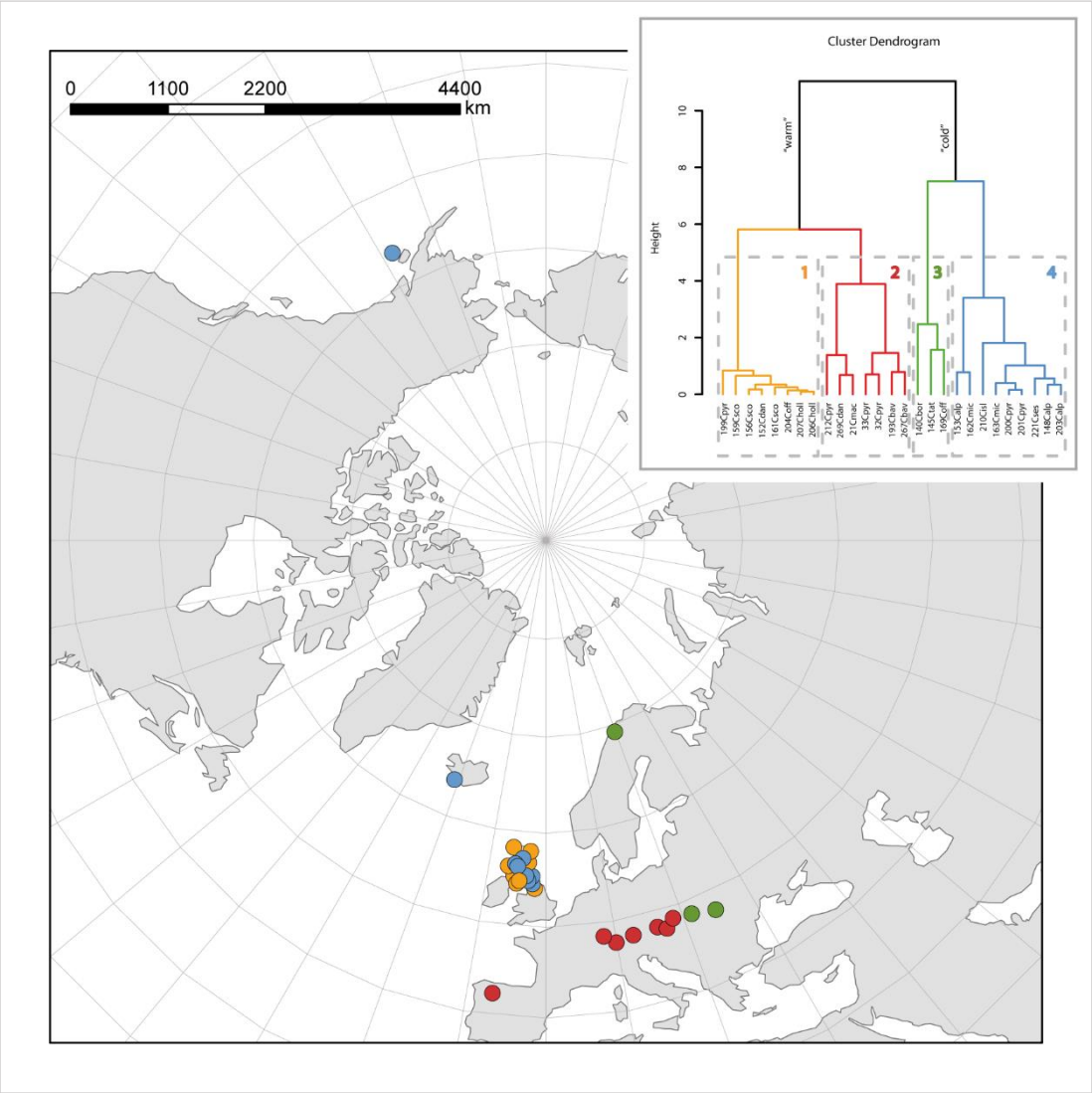

Figure III-4. Geographical distribution of bioclimatic clusters as discovered via hierarchical cluster analysis. Colors of the 27 *Cochlearia* populations considered in the metabolomic analyses are representing clusters 1-4.

3.4.2 Metabolomics – compound analysis (ANOVA)

Out of the 40 target compounds selected for metabolomic analysis, a total of 38 metabolic signatures could be annotated (at least to the level of the general compound class) and two additional compounds were reproducible across all samples but could not be identified. The detected known metabolites fall into general compound classes of alkaloids, amino acids, carbohydrates, fatty acids and organic acids (see above, Table III-2). This set of compounds represents a robust and representative snapshot of the central metabolism in *Cochlearia* (for comparison, Davey et al. (2009) analyzed 32 metabolites (carbohydrates and amino acids) in *Arabidopsis lyrata* subsp. *petraea*).

Measured concentrations of the 40 target metabolites (both raw and normalized results) are given with Supplemental Data Set III-5 and the concentrations normalized by weight and internal standard were used for statistical analyses (see Supplemental Data Set III-5 sheet 7 'normalized grouped by bioclim' for ANOVA analyses input and Supplemental Data Set III-7 for DAPC analyses input). Summaries of the different ANOVA analyses grouped by bioclimatic clusters are given with Supplemental Data Set III-6.

Soluble carbohydrates

The (normalized) concentrations of the analyzed soluble carbohydrates and the polyol *myo*-inositol varied both between treatments and bioclimatic clusters and also within bioclimatic clusters and populations (Figure III-5 and Supplemental Data Set III-5). As shown in Figure III-5, most of the carbohydrates showed statistically significant increased concentrations after 20 days of cold treatment compared to the control treatment. With eight out of nine measured compounds, clusters 2 and 4 had the largest number of significantly increased compounds. Cluster 4 had the highest total concentration of carbohydrates after the cold treatment (Supplemental Data Set III-5, sheet 7 'normalized grouped by bioclim'). Under warm conditions clusters 3 and 4 had higher total concentrations of carbohydrates compared to clusters 1 and 2, mainly resulting from high sucrose concentrations. Overall, clusters 1 and 2 showed a higher total increase of carbohydrates but the largest individual significant increase in response to cold was found for fructose in bioclimatic cluster 3 (~10-fold).

Amino acids

Similar to the results gained for soluble carbohydrates, an increase in concentration after the cold treatment was revealed for most of the detected free amino acids (Figure III-6) but many of these observed increases were not significant. Both bioclimatic clusters 2 and 4 had the highest number of significantly increased amino acids, namely aspartic acid, glutamic acid, proline and serine. Cluster 3 had no amino acids that significantly increased with cold temperatures but the smaller sample size of this cluster (6 samples under control conditions, 8 samples under cold conditions) has to be considered. The amino acids with the highest significant increases were glutamic acid in cluster 4 (~5.4 fold) and proline in cluster 2 (~5.0 fold). Out of all analyzed amino acids, proline was present at highest concentrations in all bioclimatic clusters under both temperature conditions. The highest average concentrations of proline at 20°C were found in bioclimatic cluster 1, while under cold conditions cluster 4 had the highest average proline accumulation. Cluster 4 plants also had the highest accumulation of glutamic acid, a precursor for proline biosynthesis, under cold conditions. The highest total concentration of amino acids under cold conditions was detected for bioclim cluster 4, whereas the highest total increase in amino acid concentration was found for bioclim cluster 2 (~3.6 fold; Supplemental Data Set III-5, sheet 7 'normalized grouped by bioclim').

Alkaloids

As visualized in Figure III-7, the concentrations of tropine and pseudotropine did not differ significantly between cold (5°C) and control treatment (20°C) but showed some variation between the different bioclimatic clusters with the arctic/alpine cluster 4 having the highest concentrations of alkaloids under both temperature conditions.

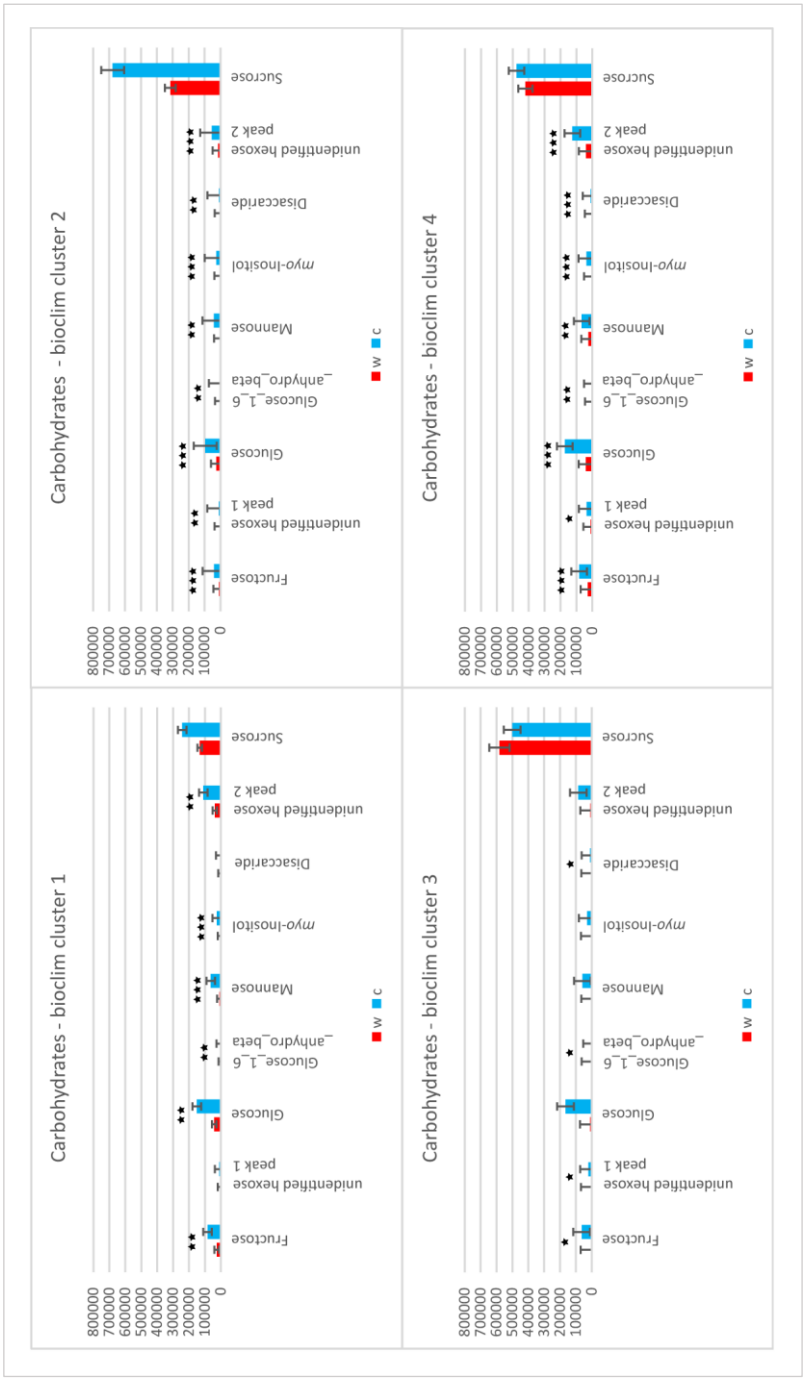

Figure III-5. Concentrations of free carbohydrates in leaves of *Cochlearia* plants from four bioclimatic clusters grown under control (20°C, red) or cold (5°C, blue) conditions for 20 days. Significant differences between control/warm and cold treatments within each cluster as revealed via ANOVA, are indicated by asterisks (* $P \leq 0.05$; ** $P \leq 0.02$; *** $P \leq 0.001$). Means of normalized concentrations per cluster are given with standard errors.

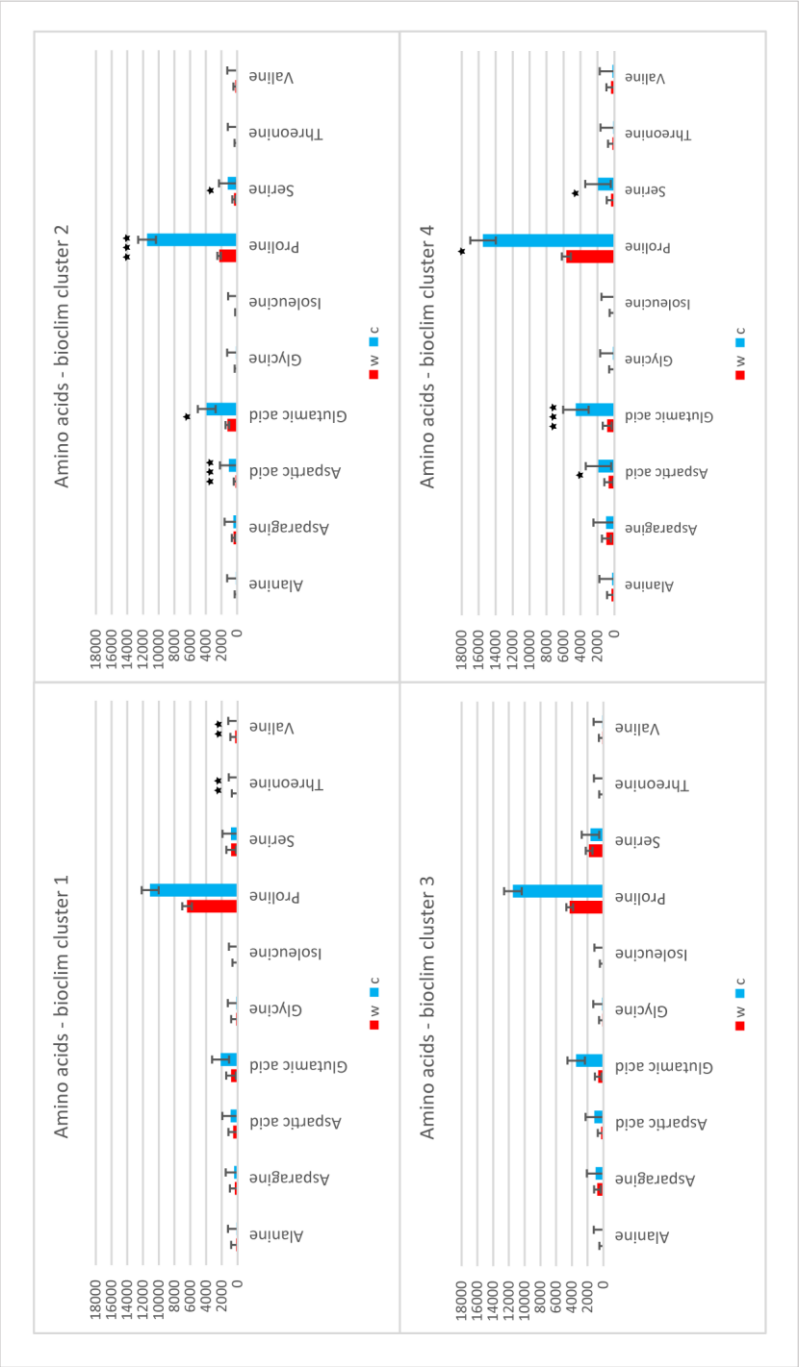

Figure III-6. Concentrations of free amino acids in leaves of *Cochlearia* plants from four bioclimatic clusters grown under control (20°C, red) or cold (5°C, blue) conditions for 20 days Significant differences between control/warm and cold treatments within each cluster as revealed via ANOVA, are indicated by asterisks (* $P \leq 0.05$; ** $P \leq 0.02$; *** $P \leq 0.001$). Means of normalized concentrations per cluster are given with standard errors.

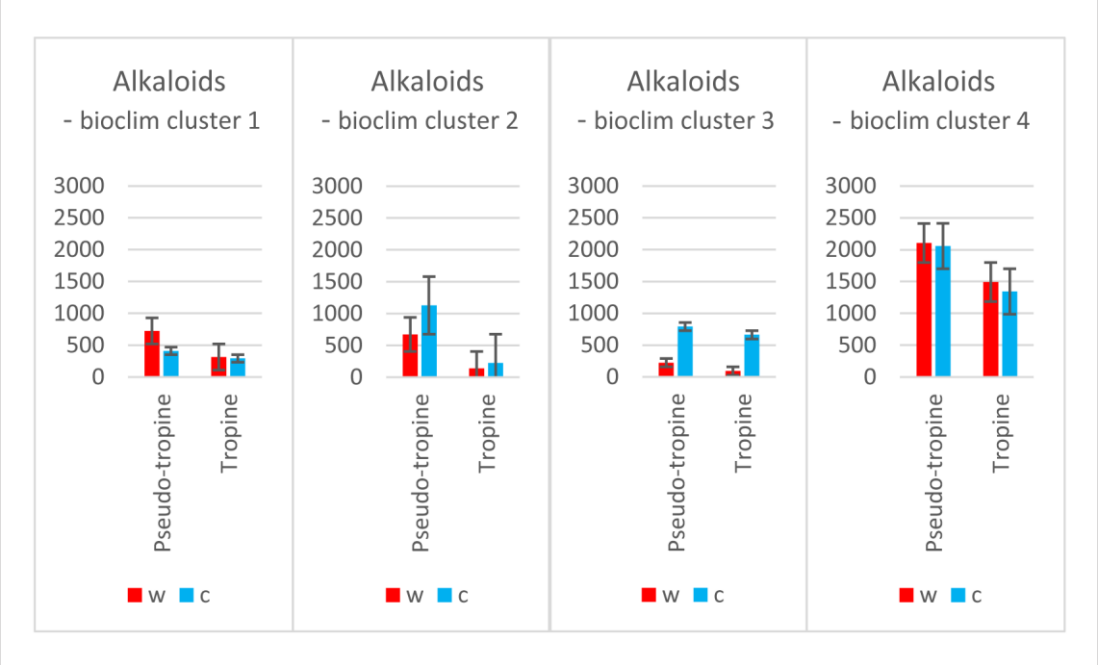

Figure III-7. Concentrations of alkaloids in leaves of *Cochlearia* plants from four bioclimatic clusters grown under control (20°C, red) or cold (5°C, blue) conditions for 20 days. No significant differences between the two treatments within each cluster were revealed via ANOVA. Means of normalized concentrations per cluster are given with standard errors.

Organic acids

Overall, few organic acids showed significantly higher concentrations after cold treatment (Figure III-8), namely ferulic acid (cluster 2), fumaric acid (cluster 3), glyceric acid (clusters 1 and 4) and malic acid (all clusters). Malic acid had the strongest significant increase compared to control conditions in the inland cluster 2 (~4.8 fold) followed by alpine cluster 3 (~4.1 fold). The strongest increase in the total concentration of organic acids was found for bioclimatic cluster 2 (~3.5 fold), while the arctic/alpine cluster 4 had the highest total concentration of organic acids after both control and cold treatment (Supplemental Data Set III-5, sheet 7 'normalized grouped by bioclim'). The coastal cluster 1 had the lowest overall increase in organic acids in response to cold (~1.7 fold).

Fatty acids

No significant differences between warm and cold conditions were detected for the fatty acids hexanoic acid and octadecanoic acid (Figure III-9) but variation between the four bioclimatic clusters can be observed.

Miscellaneous and unknown metabolites

Regarding the four remaining metabolites (two unknown metabolites, RT:19.905 unknown and RT:20.075 unknown; 1-pyrroline-3-hydroxy-5-carboxylic acid and Ethanolamine) no significant increases in compound concentrations were detected between control and cold treatment except

for a significant increase of the 1-pyrroline-3-hydroxy-5-carboxylic acid concentration in response to cold in cluster 2 (Figure III-10).

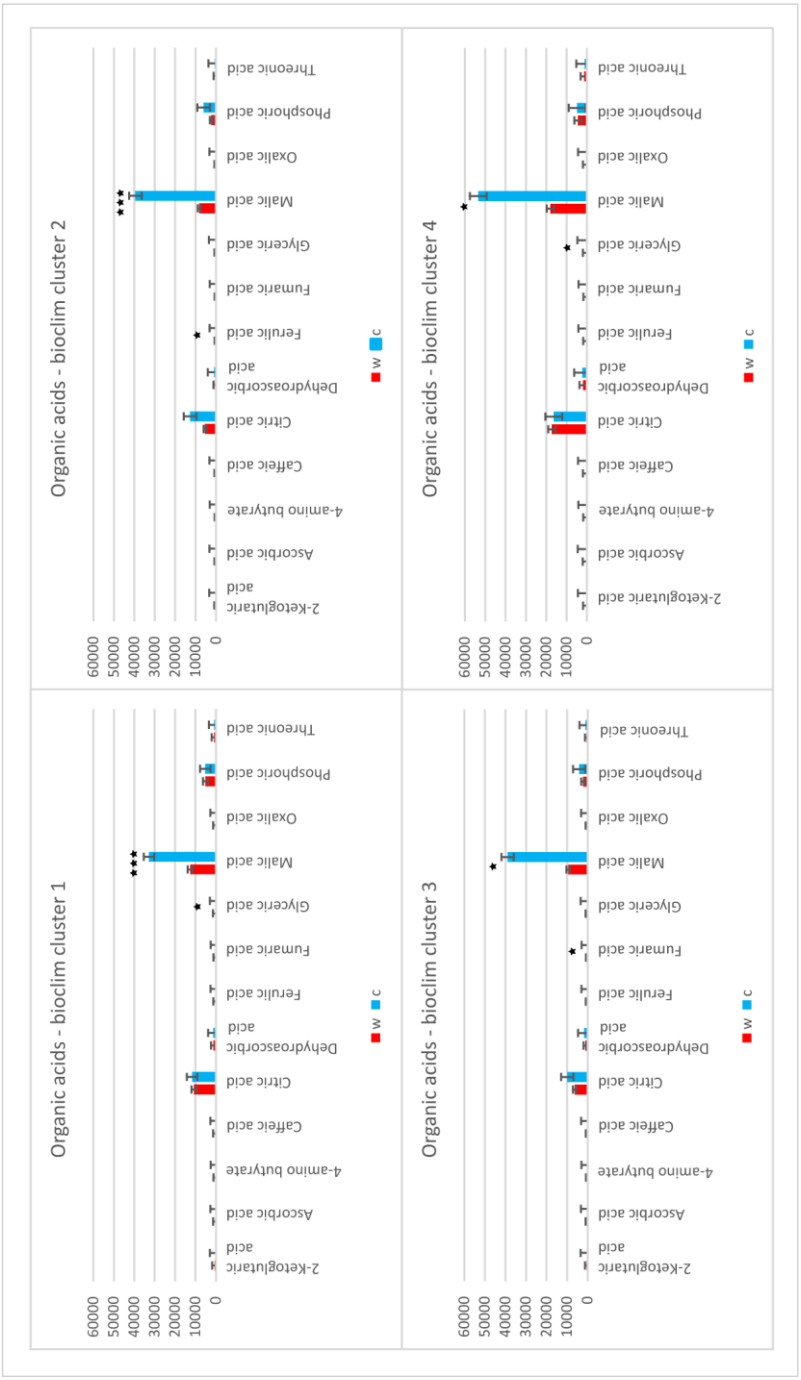

Figure III-8. Concentrations of organic acids in leaves of *Cochlearia* plants from four bioclimatic clusters grown under control (20°C, red) or cold (5°C, blue) conditions for 20 days. Significant differences between control/warm and cold treatments within each cluster as revealed via ANOVA, are indicated by asterisks (* $P \leq 0.05$; ** $P \leq 0.02$; *** $P \leq 0.001$). Means of normalized concentrations per cluster are given with standard errors.

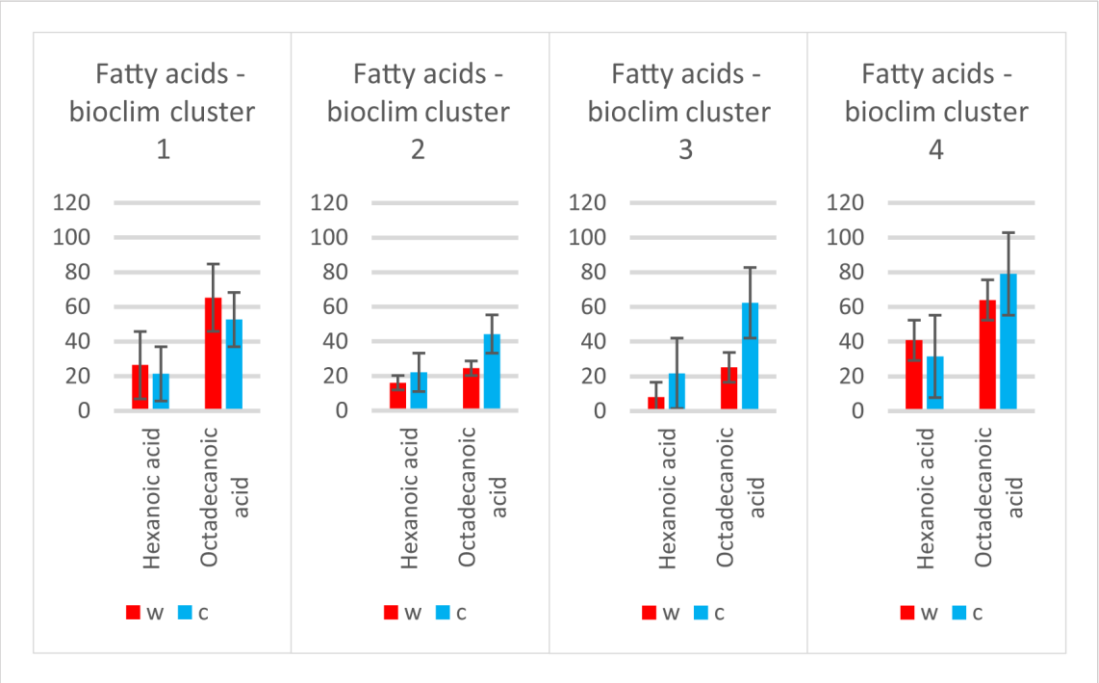

Figure III-9. Normalized concentrations of fatty acids in leaves of *Cochlearia* plants from four bioclimatic clusters grown under control (20°C, red) or cold (5°C, blue) conditions for 20 days. No significant differences between the two treatments were detected for any of the four clusters. Means of normalized concentrations per cluster are given with standard errors.

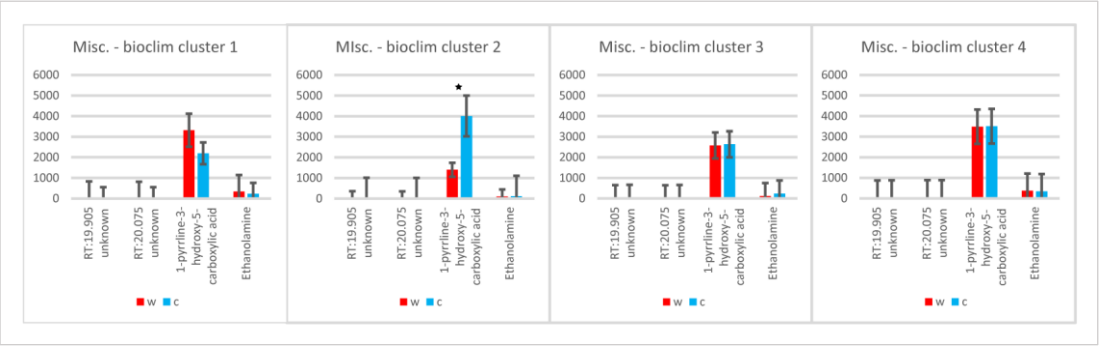

Figure III-10. Normalized concentrations of miscellaneous compounds in leaves of *Cochlearia* plants from four bioclimatic clusters grown under control (20°C, red) or cold (5°C, blue) conditions for 20 days. Significant differences between control/warm and cold treatments within each cluster as revealed via ANOVA, are indicated by asterisks (* $P \leq 0.05$; ** $P \leq 0.02$; *** $P \leq 0.001$). Means of normalized concentrations per cluster are given with standard errors.

3.4.3 Metabolomics – DAPC

DAPC 1 – All data points grouped by treatment

The *DAPC 1*, based on all measured metabolomic data points with group priors defined by treatment (see above), revealed a clear separation between 20°C conditions and cold treatment at 5°C (Figure III-11), indicating a strong differential metabolic response to the different temperatures. As expected, the three groups under control conditions show a nearly complete overlap. The first 10 PCs, explaining 83.4% of the total variation within the metabolite data, were retained for DAPC based on the results of the cross validation (Supplemental Data Set III-8). Here, group memberships were successfully assigned in 49.3% of all replications compared to 25% by random assignment. Thus, the assignment success of the analysis was comparatively low, which is easily explained by the fact that three out of the four analyzed groups represent the same temperature condition of 20°C and thus show a strong overlap (Figure III-11).

Figure III-11 illustrates that discriminant function 1, separating control and cold conditions, explains most of the observed variance. The contributions of the different metabolites to the distinction between samples under control and cold conditions are listed in Supplemental Data Set III-9 and illustrated in Figure III-12. Glutamic acid, glucose, aspartic acid and mannose showed the highest contribution and the top ten contributing compounds are composed of six carbohydrates (glucose, mannose, unidentified hexose peak 2, fructose, *myo*-inositol, disaccharide), three amino acids (glutamic acid, aspartic acid, proline) and one organic acid (malic acid).

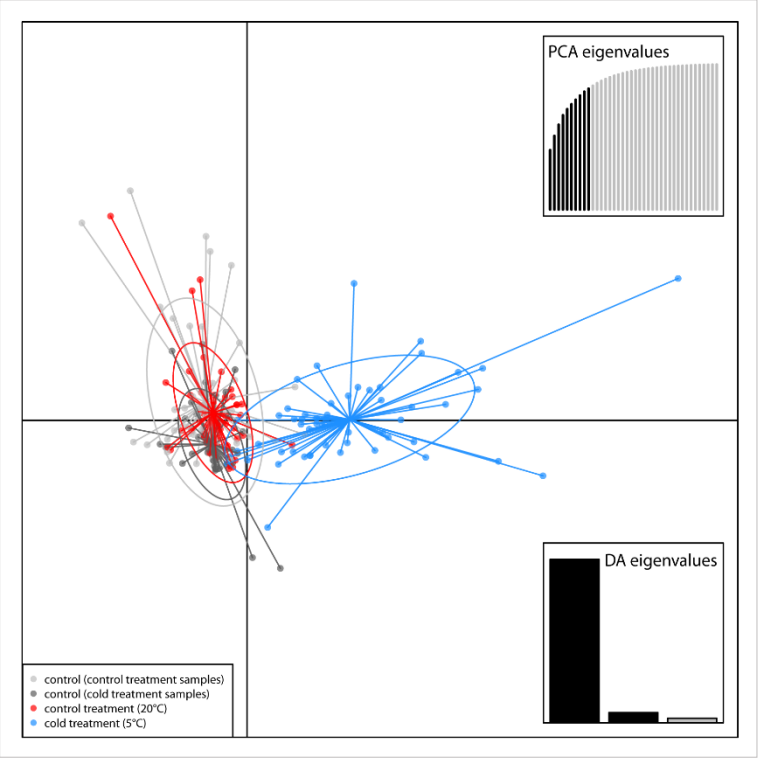

Figure III-11. Result of a DAPC based on all data points, grouped by treatment. Colors indicate the time of metabolite extraction as well as temperature conditions (control conditions (light and dark grey): 20°C (first measurement); cold treatment: 5°C; control conditions: 20 °C (second measurement)). Displayed scatterplot is based on 10 retained PCs and DA eigenvalues 1 and 2.

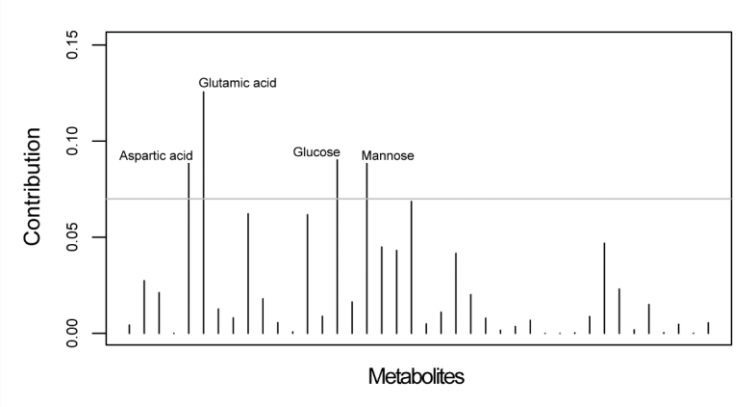

Figure III-12. Loading plot of metabolite contributions to discriminant function number 1 (separating cold and warm conditions), in *DAPC 1* based on all data points grouped by type of treatment. The grey line indicates a threshold of 0.07 and metabolites above this threshold are labeled accordingly.

DAPC 2 – All data points grouped by treatment x bioclimatic cluster

The second DAPC describes the relationship between warm and cold conditions in more detail by combining it with the four bioclimatic clusters. Here, the first 20 PCs were retained according to the cross-validation results (Supplemental Data Set III-8), this way conserving 95.3% of the total variance. Again, discriminant function 1 nicely separates warm and cold treatment (Figure III-14), whereas axes 2 and 3 separate bioclimatic clusters to a certain degree (Figures III-14, III-15). As shown in Figures III-14 and III-15, there is a generally strong overlap between the four bioclimatic clusters under both temperature conditions, yet in the cold treatment the two "warmer" bioclimatic clusters 1 and 2 are slightly separated from the two arctic/alpine clusters 3 and 4 based on discriminant function 3 (Figure III-15). Respective contributions of the different compounds are given with Supplemental Data Set III-9 and loadings of discriminant function 3 are illustrated in Figure III-13. The amino acid proline obviously contributed the most to discriminant function 3, followed by 2-ketoglutaric acid, alanine and ascorbic acid.

The strong overlap between the clusters is also reflected by the low assignment success as revealed via cross-validation, with the highest success rate of 24.6% achieved for 20 PCs compared to 6% of successful assignments via random chance. Thus, although no high success rate was achieved, the chosen model still performed about four times better than random assignment.

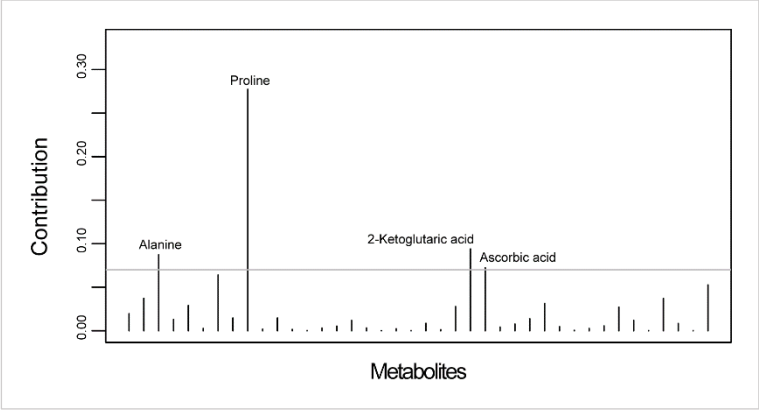

Figure III-13. Loading plot of metabolite contributions to discriminant function 3 in *DAPC 2* based on all data points grouped by type of treatment x bioclimatic cluster. The grey line indicates a threshold of 0.07 and metabolites above this threshold are labeled accordingly.

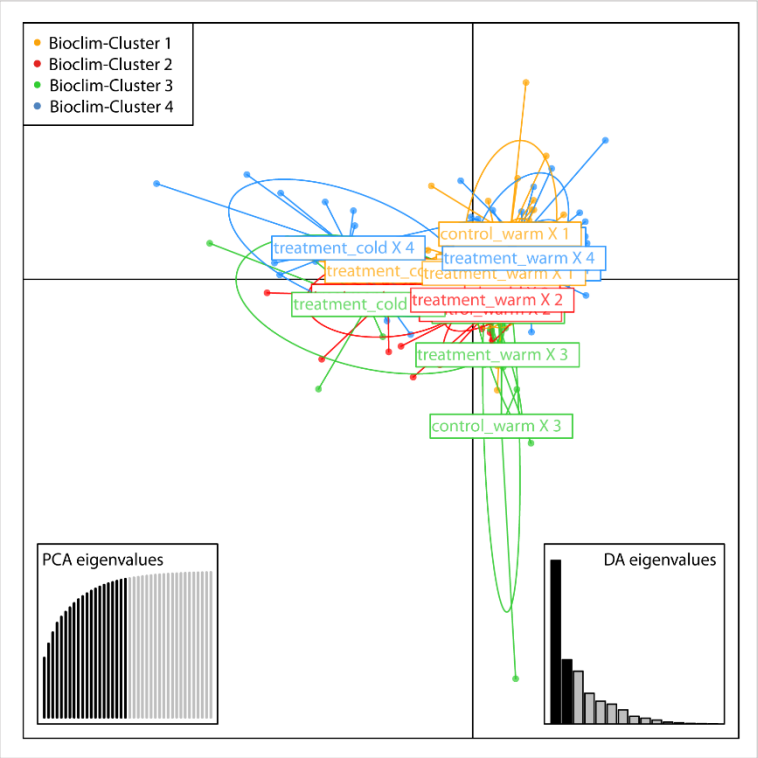

Figure III-14. Result of *DAPC 2* grouped by treatment x bioclimatic clusters and based on all metabolomic data points and 20 retained PCs, representing 95.3% of the total variance. DA axes 1 and 2 are shown.

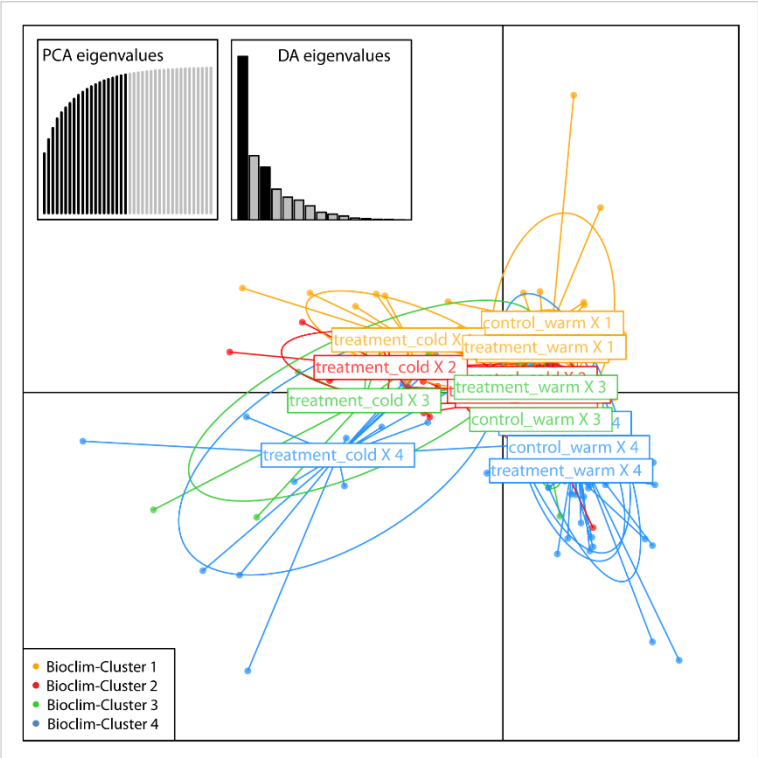

Figure III-15. Result of *DAPC 2* grouped by treatment x bioclimatic clusters and based on all metabolomic data points and 20 retained PCs, representing 95.3% of the total variance. DA axes 1 and 3 are shown.

DAPC 3 – All data points grouped by bioclimatic cluster

For *DAPC 3*, again all data points were analyzed and group priors were set to match the four bioclimatic clusters. About 98.8% of the total variance were retained by using the first 30 PCs with a mean successful assignment rate of 66.1% (25.2% by random chance) as revealed from cross-validation (Supplemental Data Set III-8).

The DAPC result shows that bioclimatic groups are not clearly separated by the metabolomic data (Figures III-16, III-17). However, there is separation to a certain degree. As illustrated in the scatterplots, DA axis 1 separates the arctic/alpine cluster 4, from the remaining clusters whereas axis 2 mainly separates the coastal cluster 2. Figure III-16 nicely shows the strong overlap between the inland and the alpine cluster. However, axis 3 (Figure III-17) separates the alpine cluster 3, while inland and coastal show a stronger overlap here. In Figure III-18, a minimum spanning tree, calculated from squared distances between the four bioclimatic clusters within the entire data space (all discriminant functions), was added to visualize overall proximities of the clusters. It nicely confirms the close association of the red inland and the green alpine cluster and the more separate position of the arctic/alpine cluster 4 which is found closest to the coastal cluster 2.

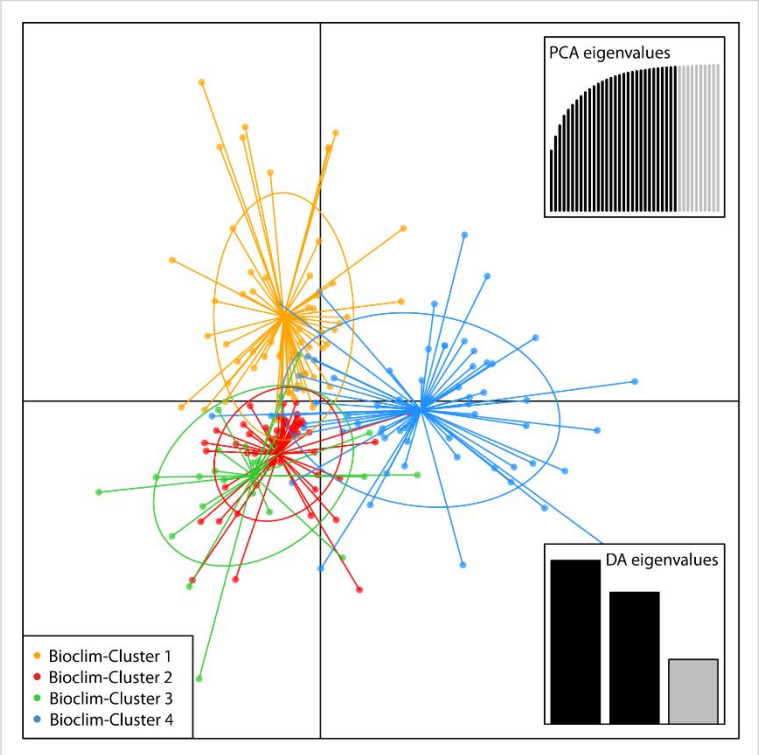

Figure III-16. Result of *DAPC 3* based on all data points, grouped by the four bioclimatic clusters. 30 PCs (98.8% of the total variance) were retained; DA axes 1 and 2 are shown.

The metabolite most significantly contributing to discriminant function 1 and thus to the separation of the arctic/alpine cluster 4 from the remaining clusters was determined to be the alkaloid tropine, followed by citric acid, an unidentified hexose (peak 2) and 2-ketoglutaric acid (Figure III-19 and Supplemental Data Set III-9). As illustrated in Fig. III-20, for discriminant function 2, which mainly separates the coastal cluster, the compounds showing the strongest contribution are ethanolamine with a putative role in salt tolerance of coastal taxa (Parvaiz and Satyawati, 2008),

followed by ascorbic acid and the carbohydrates *myo*-inositol and fructose. Finally, the compounds contributing the most to discriminant function 3 were determined to be ascorbic acid, serine and glutamic acid (Figure III-21).

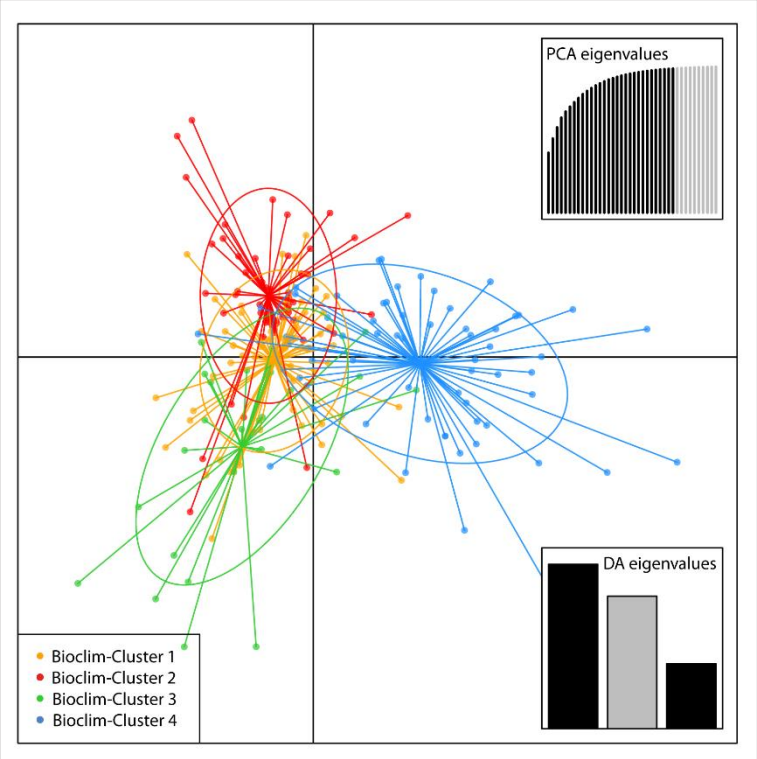

Figure III-17. Result of *DAPC 3* based on all data points, grouped by the four bioclimatic clusters. 30 PCs (98.8% of the total variance) were retained; DA axes 1 and 3 are shown.

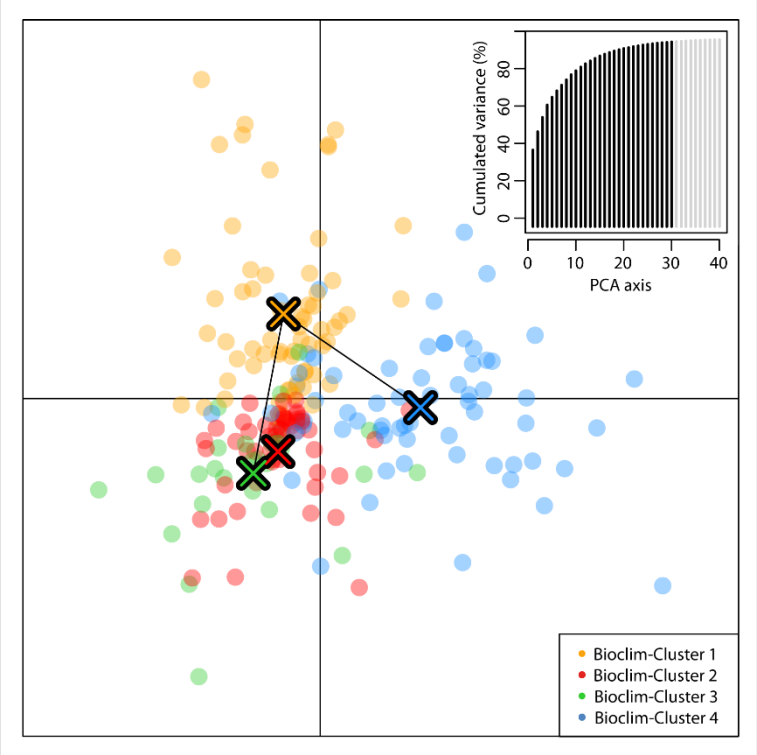

Figure III-18. Result of *DAPC 3* based on all data points (control conditions and 5°C), grouped by four bioclimatic clusters - DA eigenvalues 1 and 2 displayed. A minimum spanning tree (group centers marked with crosses) illustrates proximities between bioclimatic clusters in the whole data space based on 30 PCs, and 98.8% of the total variance.

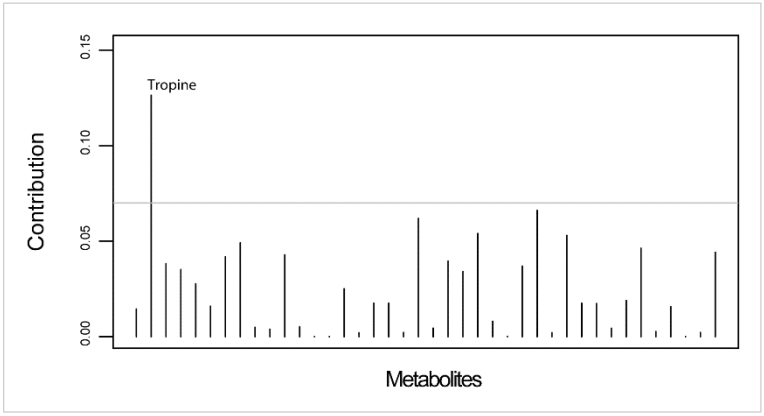

Figure III-19. Loading plot of metabolite contributions to *DAPC 3* (all data points grouped by bioclimatic cluster), discriminant function number 1.

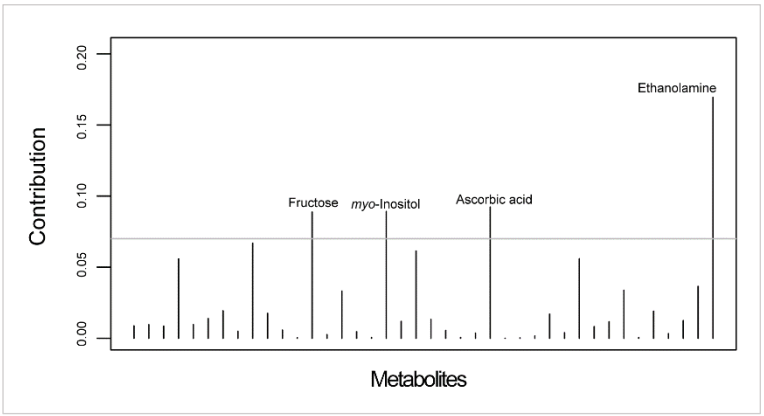

Figure III-20. Loading plot of metabolite contributions to *DAPC 3* (all data points grouped by bioclimatic cluster), discriminant function number 2.

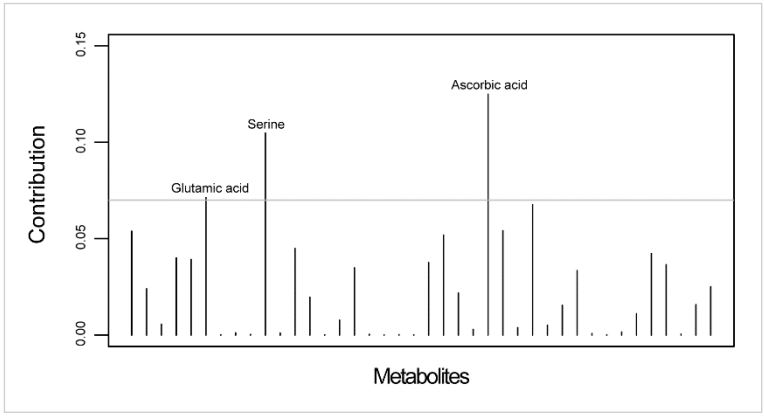

Figure III-21. Loading plot of metabolite contributions to *DAPC 3* (all data points grouped by bioclimatic cluster), discriminant function number 3.

DAPC 4 – Samples after 20 days under control conditions (20°C), grouped by bioclimatic clusters

The optimal number of PCs to retain for *DAPC 4*, as revealed via cross-validation, was 25 with a successful assignment rate of 61.5% (Supplemental Data Set III-8). For this dataset (consisting of 49 plant samples), this number exceeded the maximum recommended number of retained PCs (number of samples divided by three) which may lead to over-fitting and unstable results (Jombart, 2008; Jombart and Collins, 2015). Therefore, DAPC was performed by retaining only 16 PCs and thus 96.2% of the total variance.

Figure III-22 illustrates the separation of bioclimatic clusters, with discriminant function 1 mainly separating the orange coastal cluster and the blue arctic/alpine cluster from the Central European inland samples in the overlapping red and green clusters. However, the separation is not very strong and there is considerable overlap between the clusters. Axis 2 separates coastal and arctic/alpine from Central European inland samples while axis 3 (Figure III-23) did not reveal any clear separation. Again, the scatterplot based on DA axes 1 and 2 is also shown with a minimum spanning tree based on all discriminant functions (Figure III-24). The overall grouping illustrates the close relationship between the two Central European inland clusters (red and green) and connects the arctic cluster with the coastal cluster as already seen in the results of *DAPC 3*.

On the metabolite level, glycine, 2-ketoglutaric acid, proline, 4-amino butyrate and tropine showed the strongest contributions to discriminant function 1 (Figure III-25, Supplemental Data Set III-9), whereas glutamic acid, phosphoric acid, sucrose and caffeic acid contributed most significantly to the structure as revealed from discriminant function 2 (illustrated in Figure III-26).

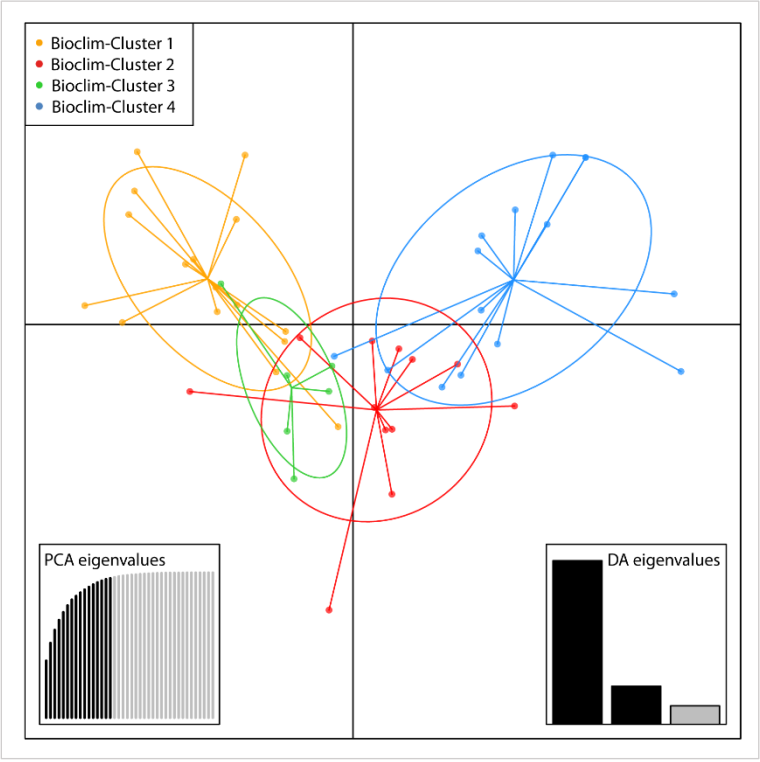

Figure III-22. Scatter plot of *DAPC 4*, based on metabolite data of plant samples after 20 days at 20°C, grouped by four bioclimatic clusters. 16 PCs (96.2% of total variance) were retained and discriminant functions 1 and 2 are shown.

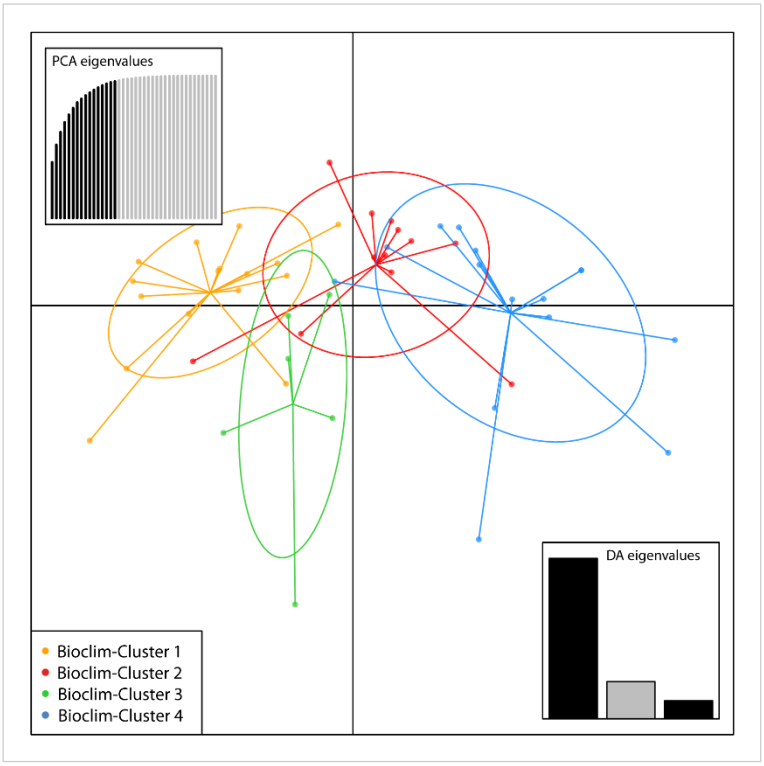

Figure III-23. Scatter plot of *DAPC 4* (samples after 20 days at 20°C, grouped by bioclimatic clusters). 16 PCs (96.2% of total variance) were retained and discriminant functions 1 and 2 are shown.

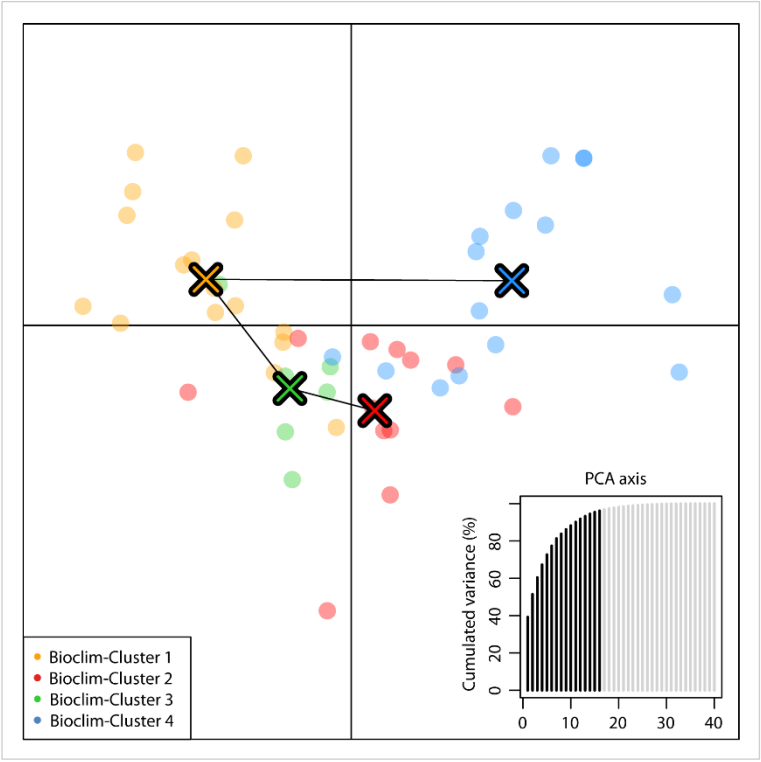

Figure III-24. Scatter plot of *DAPC 4* (samples after 20 days at 20°C, grouped by bioclimatic clusters), based on DA axes 1 and 2 with a minimum spanning tree based on all discriminant functions (16 retained PCs, 96.2% of total variance).

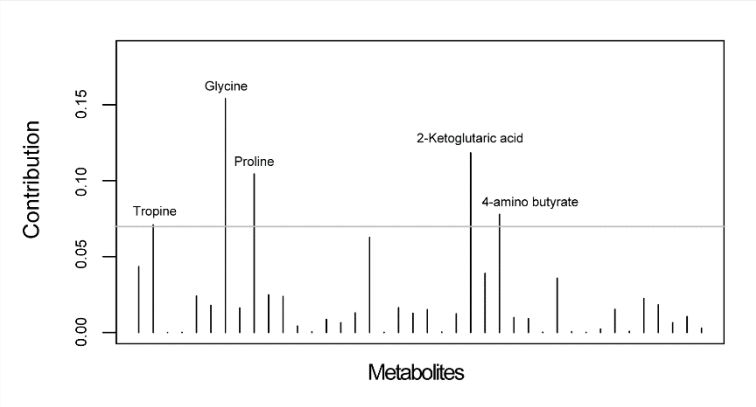

Figure III-25. Loading plot of metabolite contributions to *DAPC 4* (samples after 20 days at 20°C, grouped by bioclimatic clusters), discriminant function number 1.

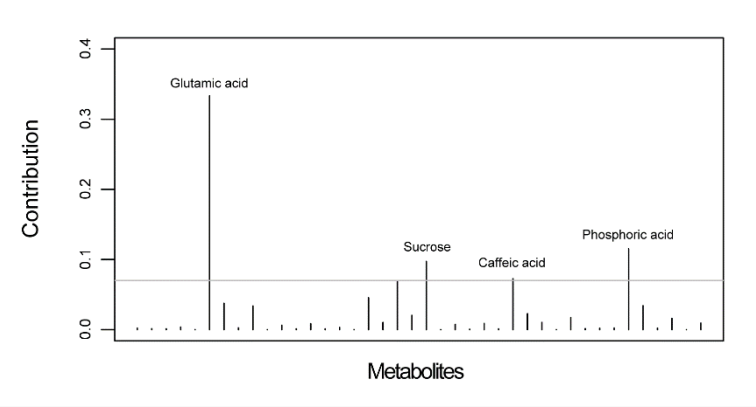

Figure III-26. Loading plot of metabolite contributions to *DAPC 4* (samples after 20 days at 20°C, grouped by bioclimatic clusters), discriminant function number 2.

DAPC 5 – Samples after cold treatment, grouped by bioclimatic clusters

The initial cross-validation for *DAPC 5*, based on metabolite data extracted from samples after cold treatment, revealed the optimal number of PCs to retain to be 13, thereby conserving 94.5% of the total variance. With only 46.9% of successful group assignments, the overall performance was worse than for *DAPC 4* (after 20 days under control conditions, 20°C) and accordingly, the resulting separation of the four clusters was less pronounced, yet general tendencies could be observed (Figures III-27, III-28).

While the coastal, inland and arctic clusters appear to be somewhat separated, the green alpine cluster has a central position and shows full overlap with the other clusters based on discriminant functions 1 and 2 (Figure III-27). The first discriminant function separates 1) the coastal cluster, 2) arctic and arctic/alpine clusters and 3) the inland cluster, whereas discriminant function 2 mainly separates the arctic/alpine cluster 4 from the remaining clusters. As illustrated in Figure III-28, discriminant function 3 slightly separates the alpine cluster 3 from the remaining clusters.

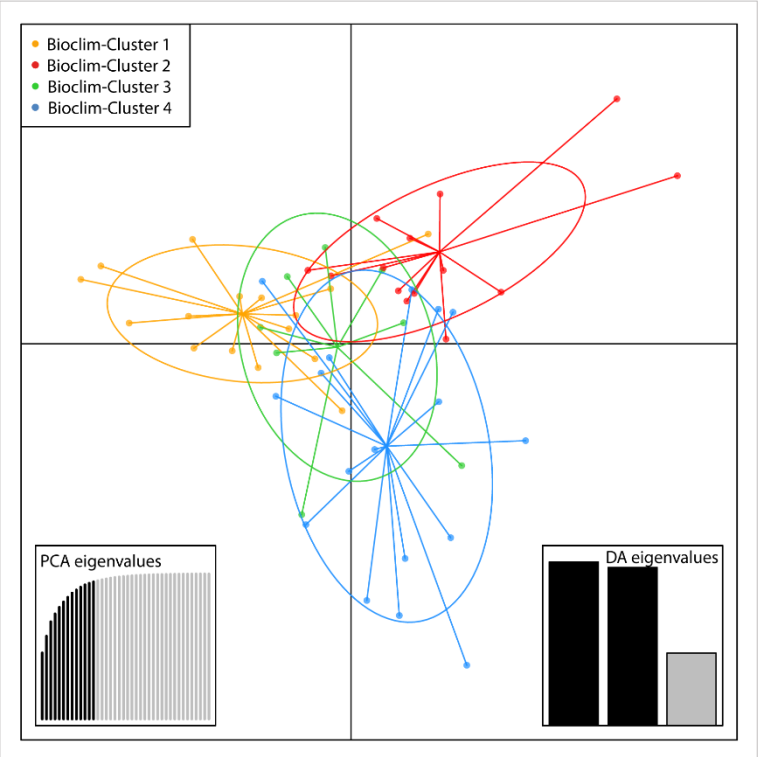

Figure III-27. Scatterplot of DAPC 5, based on plant samples after cold treatment, grouped by four bioclimatic clusters. 13 PCs (94.5% of total variance) were retained and discriminant functions 1 and 2 are shown.

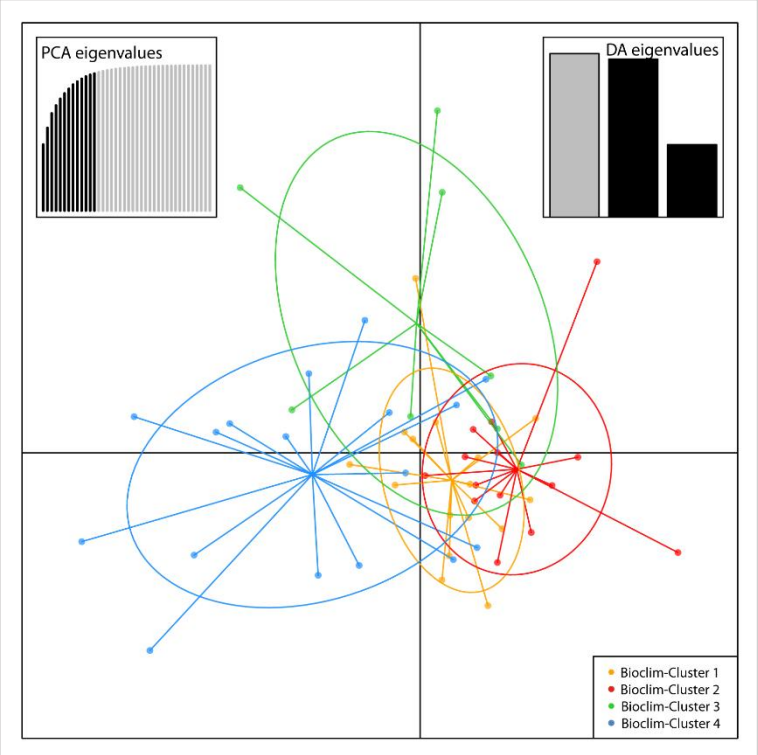

Figure III-28. Scatterplot illustrating discrimi-nant functions 2 and 3 of *DAPC 5*, plants after cold treatment, grouped by four bioclimatic clusters. 13 PCs (94.5% of total variance) were retained.

As revealed from a minimum spanning tree in Figure III-29, the general grouping pattern changed compared to *DAPC 4* under the control treatment, bringing the arctic/alpine cluster closer to the green alpine cluster now placed right at the center of the four bioclimatic clusters.

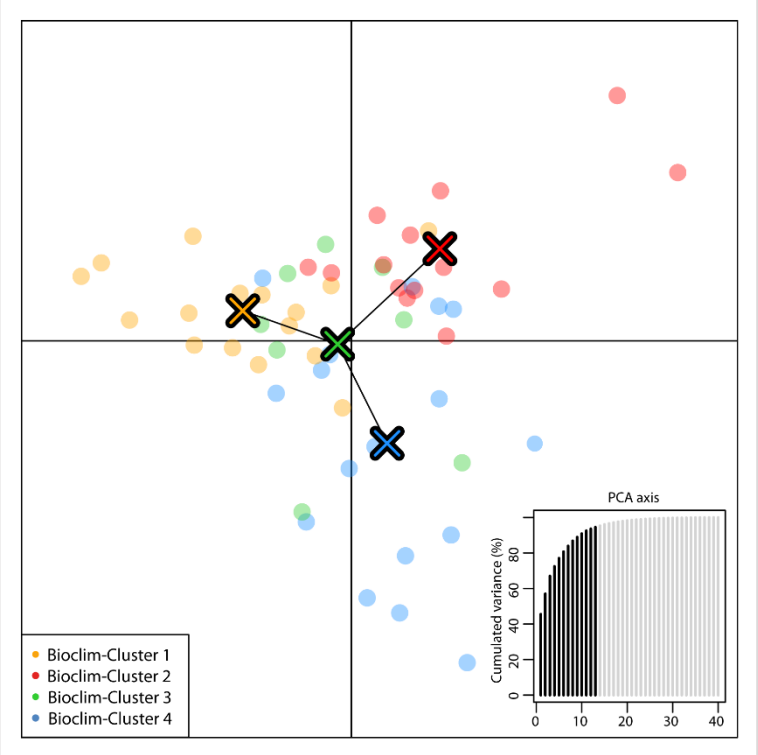

Figure III-29. Scatter plot of *DAPC 5*, plants after cold treatment, based on DA axes 1 and 2 with a minimum spanning tree based on all discriminant functions (13 retained PCs, 95.5% of total variance).

Loading plots for discriminant functions 1 to 3 are given with Figures III-30 to III-32. The loading plot for discriminant function 2 (Figure III-31), separating the arctic/alpine cluster from the remaining clusters, revealed strong contributions of the carbohydrate sucrose, the osmoprotectant proline and citric acid to the observed clustering.

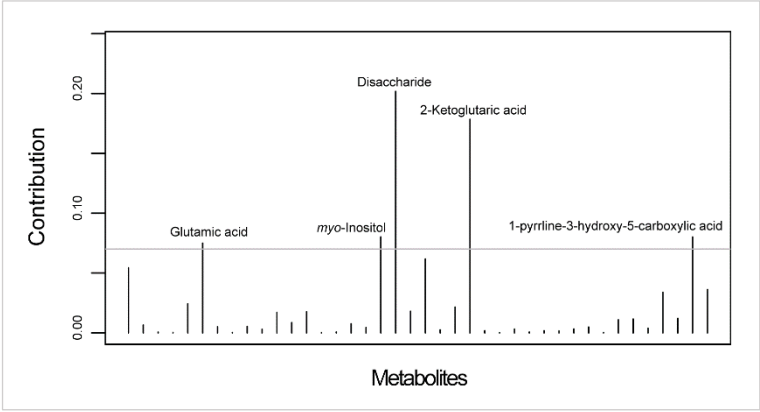

Figure III-30. Loading plot of metabolite contributions to *DAPC 5* (samples after cold treatment grouped by bioclimatic clusters), discriminant function number 1.

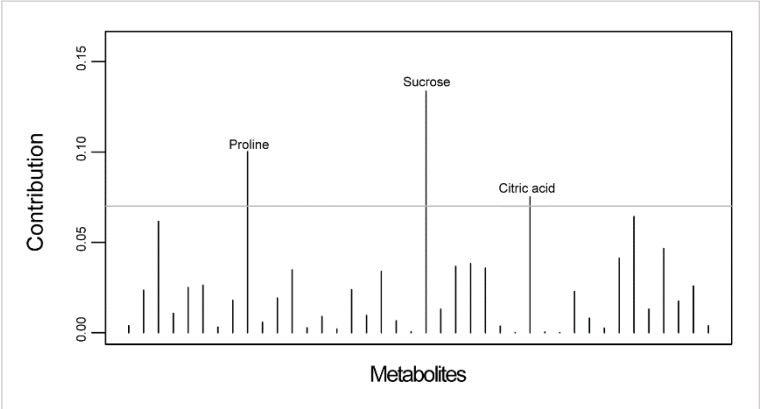

Figure III-31. Loading plot of metabolite contributions to *DAPC 5* (samples after cold treatment grouped by bioclimatic clusters), discriminant function number 2, separating the blue arctic/alpine cluster from the remaining clusters.

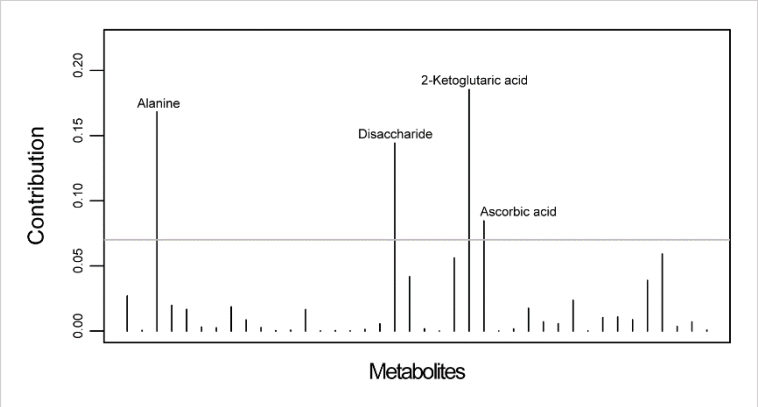

Figure III-32. Loading plot of metabolite contributions to *DAPC 5* (samples after cold treatment grouped by bioclimatic clusters), discriminant function number 3, separating the green alpine cluster from the remaining clusters (Figure III-27).

3.5 Discussion and outlook

3.5.1 Insights from metabolomic compound analysis – investigating the cold metabolome of *Cochlearia*

The strong metabolomic reaction of the genus *Cochlearia* in response to cold, as revealed in here based on targeted metabolite analysis, is well in accordance with similar studies in other plant taxa, for example in the genus *Arabidopsis* (Kaplan et al., 2004; Hannah et al., 2006; Davey et al., 2009), where Cook et al. (2004) described increased concentrations for 75% of all measured metabolites in the response of *A. thaliana* to low temperatures.

Especially an increase in soluble carbohydrates is one of the well-known reactions to cold stress in plants (e.g. Santarius and Milde, 1977; Koster and Lynch, 1992; Klotke et al., 2004; Usadel et al., 2008; Davey et al., 2009) and has also been described for *Cochlearia* before (Kappen and Ullrich, 1970). In the present study, this effect was clearly apparent from the high number of significantly increased carbohydrate concentrations in response to cold (Figure III-5), and the fact that six soluble carbohydrates were found to be among the top ten compounds contributing most significantly to the distinction between control conditions and cold treatment in *DAPC 1* (Supplemental Data Set III-9). Here, aside from its cryoprotective function, the strong contribution of glucose (Figure III-12) might be linked to its role as a signaling molecule in the plant cold response (reviewed by Janská et al., 2010). Other carbohydrates that were significantly increased after 20 days of cold treatment, such as mannose and the polyhydric alcohol *myo*-inositol, have also been shown to increase in response to cold stress in the genus *Arabidopsis* and beneficial effects on freezing tolerance have been suggested (Davey et al., 2009; Maruyama et al., 2009). Whereas an early and rapid increase in sucrose levels in response to plant temperature stress has been described for *A. thaliana* e.g. by Kaplan et al. (2004) and linked to a putative role as a signaling molecule aside from its cryoprotective function, for *Cochlearia*, sucrose concentrations were elevated after cold treatment in bioclimatic clusters 1, 2 and 4, but not significantly increased in any of them. Yet, the generally high sucrose concentrations in all clusters under both control and cold conditions might be advantageous under cold stress and thus reflect the high degree of cold adaptation in *Cochlearia*.

Similar to the accumulation of carbohydrates, increased amounts of amino acids, as revealed for all bioclimatic clusters of *Cochlearia*, have been shown to be a typical reaction of the plant metabolism to abiotic stresses (reviewed by Krasensky and Jonak, 2012) including cold stress e.g. during wintering (Sagisaka and Araki, 1983; Sagisaka, 1987). Proline, the most highly abundant amino acid analyzed in the *Cochlearia* metabolome under both temperature conditions, with a significant increase in response to cold in clusters 2 and 4 (Figure III-6) and a strong contribution to *DAPC 1*, is well known for its role in the response to several abiotic stresses such as drought, salt, heavy metal, high and low temperatures (e.g. Koster and Lynch, 1992; Nanjo et al., 1999; Klotke et al., 2004; Usadel et al., 2008; reviewed by Ashraf and Foolad, 2007; Verbruggen and Hermans, 2008) where it is supposed to act as a protective compound involved in e.g. the maintenance of membrane stability and the buffering of cytosolic pH (Verbruggen and Hermans,

2008; Preston and Sandve, 2013). Moreover, as suggested by Maggio et al. (2002), proline accumulation might have a signaling function contributing to plant adaptation. Another typical stress response was revealed by the increased levels of the amino acids glutamic acid and aspartic acid (significant in clusters 2 and 4), associated with the citric acid cycle. This is consistent with the cold metabolomes of e.g. *Arabidopsis thaliana* (Kaplan et al., 2004) and *A. lyrata* ssp. *petraea* (Davey et al., 2009), where amino acids derived from the citric acid cycle showed strong alterations in response to low temperatures. The strong contributions of glutamic acid and aspartic acid in *DAPC 1* (Figure III-12), further support this finding. Similar to the results gained by Davey et al. (2009), where plants from the coldest studied population showed the strongest increase of citric acid cycle amino acids, the arctic/alpine bioclimatic cluster 4, had the largest increase in glutamic acid under cold conditions.

Another group of compounds that has been described to accumulate in response to cold, e.g. for the *A. thaliana* low-temperature metabolome (Cook et al., 2004), comprises organic acids, among them intermediates of the citric acid cycle such as α-ketoglutarate, fumarate, malate and citrate, which is another indicator of citric acid cycle upregulation in response to cold stress. In *Cochlearia*, the only organic acid that showed significantly increased concentrations in all clusters after cold treatment was malic acid, which was also determined to strongly contribute to the distinction between the two temperature conditions in *DAPC 1*. In the analysis of the *A. thaliana* temperature stress metabolome by Kaplan et al. (2004), malic acid belonged to the group of compounds that showed altered concentrations after both heat and cold shock. About 25% of all analyzed metabolites belonged to this group and about 74% thereof showed a common response, illustrating the existing similarities in different abiotic stress responses and indicating that parts of the molecular basis are shared between them (Smallwood and Bowles, 2002). This also applies for plant responses to cold and drought stress, which have been shown to be similar in many aspects (Thomashow, 1999; Maruyama et al., 2009), given that both cold and drought (as well as salinity) subject plants to osmotic stress with implications on membrane stability, protein function and increased levels of reactive oxygen species (ROS) (Krasensky and Jonak, 2012). Therefore, it has been hypothesized that plant cold tolerance first originated from existing drought tolerance pathways that must have evolved early on in land plant evolution (Preston and Sandve, 2013). As reviewed by Preston and Sandve (2013), repeated/parallel evolution of cold tolerance via individual adjustments of the same or similar molecular mechanisms seems likely based on results from phylogenetics and ancestral trait reconstructions. Given the close phylogenetic relationship between the cold-adapted genus *Cochlearia* and the Mediterranean genus *Ionopsidium* with adaptations to at least seasonally dry habitats, a scenario where the existing adaptation to drought might have been recruited for cold tolerance during times of climate cooling seems possible.

3.5.2 DAPC analyses – in search of intrageneric variation in the *Cochlearia* cold metabolome

The detected differences in metabolite concentrations between the four bioclimatic clusters are suggestive of variation in the cold response and consequently in cold adaptation within the genus *Cochlearia*. Metabolic differentiation in response to cold stress in geographically separated

populations has been shown e.g. for *Arabidopsis lyrata* ssp. *petraea* (Davey et al., 2009), however, the considerable overlap between *Cochlearia* samples from the different bioclimatic clusters in all DAPC analyses and the generally low assignment success of the respective cross-validations indicate a rather low degree of metabolomic differentiation between the four clusters. This implies that the bioclimatic origin of the samples does not have a very strong effect on the clustering of metabolite data in the genus *Cochlearia*, which could be interpreted as a result of the young age of the genus and the generally similar ecological preferences for cold characterized habitats. Still, the slight differences that were detected between the temperature metabolomes of the four clusters indicate that there might have been ecotypic differentiation to some degree. However, there is a risk of overinterpreting these patterns and additional tests, like e.g. a-score optimization (Jombart and Collins, 2015), are needed in order to evaluate the statistical power of the results revealed in here.
Also, the detected variation within some of the analyzed populations illustrates that the response to climatic factors may vary not only over the whole distribution area of a genus but also between individuals of a single plant population (sometimes over very short distances like a few meters) as shown for example on populations of *Triticum dicoccoides* (Li et al., 1999; Li et al., 2001; Li et al., 2000) and *Hordeum spontaneum* (Huang et al., 2002; Owuor et al., 1997) as a result of varying microclimatic/microecological conditions and based on microgeographic genetic divergence within a population (reviewed by Jump and Penuelas, 2005). Thus, a more comprehensive sampling of individuals per population in the presented study would have been beneficial.
Still, based on the data examined in here, some general trends in the responses to different temperature conditions could be observed as will be discussed below.

As already stated above, a metabolomic adjustment to the cold treatment was easily detected from *DAPC 1*, and as confirmed via *DAPC 2*, all of the four bioclimatic clusters showed this response to cold, that way illustrating the fact that cold adaptation within the genus *Cochlearia* is generally high. The strong overlap of the four clusters in *DAPC 2* reflects mentioned difficulties in clearly separating the metabolomic reactions on a bioclimatic scale. Yet, as illustrated by DA axis 3 (Figure III-15), the blue arctic/alpine cluster and the green alpine cluster seem to be slightly closer associated and better separated from the other clusters under cold conditions and the very strong contribution of the stress-responsive amino acid proline (Figure III-3; see above 3.5.1) might indicate its role as one of the key players in the adaptation to extreme cold.
The putative similarity between the cold responses of the two clusters comprising the extreme habitats of arctic and alpine areas is further supported by the result of *DAPC 5* where the minimum-spanning tree revealed closer associations of these two clusters compared to *DAPC 4* under control conditions. In *DAPC 5*, DA axis 2 most clearly separates the blue and the green cluster from the others (Figures III-27, III-28) and again, the two compounds contributing most significantly to this pattern, namely sucrose and proline, can be easily linked to cold adaptation (see above, 3.5.1). At the same time, the green alpine cluster shows strong overlap with the two Central European clusters as illustrated by its central position in the minimum spanning tree. Under control conditions in *DAPC 4* (as well as in *DAPC 3*) the grouping of the four clusters is rather reflecting genomic

clustering as seen in chapter 2 (e.g. Figure II-15) with the arctic phenotype being separated from the others and closest to the coastal phenotype.
The stronger metabolomic similarities of clusters 3 and 4 in response to cold might be interpreted either as a result of introgression of the diploid arctic phenotypes into parts of the Central European phenotypes, or as a result of a repeated adaptation to extreme cold via similar pathways, that might have been facilitated in Central Europe by the genetic legacy of the arctic origin present as standing genetic variation. On a genetic level, putative introgression from the arctic into the Central European genepools or genetic footprints of the arctic past of the genus have already been suggested based on phylogenomic results presented in chapter 2.

3.5.3 Outlook

Additional metabolite analyses should be performed for ecologically relevant taxa that have been missing so far, namely the Central European high-alpine taxon *C. excelsa*, in order to get a more comprehensive picture of the *Cochlearia* cold metabolome. Moreover, including the Mediterranean sister genus *Ionopsidium* into the metabolomic analyses would enable direct comparisons between these two ecologically divergent genera and might provide new insights into the evolution of cold adaptation in *Cochlearia* and the putative recruitment from drought adaptation.
As mentioned above, in order to assess the informative value of the interesting patterns revealed via DAPC analyses, further statistical tests should be performed. If these are in support of the detected intrageneric variation in the cold response of *Cochlearia*, then the described temperature-related metabolomic phenotypes can be used as a starting point for further analyses on the evolution of cold adaptation within the genus *Cochlearia*. Here, the information gained from analyzing the respective compound contributions could then be utilized in order to identify candidate metabolites and pathways playing important roles in the cold metabolism.
Finally, analyzing not only the primary but also the secondary metabolite profiles might provide additional information regarding the cold adaptation within the genus. Flavonoids for example have been shown to play a functional role in the cold acclimation of *A. thaliana* possibly by membrane and/or protein protection or due to their antioxidant activity (Hannah et al., 2006; Schulz et al., 2016).

3.6 References

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schönswetter, P., Taberlet, P., Brodbeck, S., Gaudeul, M., and Gielly, L. (2009). History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. Ecology Letters **12,** 632-640.

Ashraf, M., and Foolad, M. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. Environmental and Experimental Botany **59,** 206-216.

Berry, J., and Bjorkman, O. (1980). Photosynthetic response and adaptation to temperature in higher plants. Annual Review of Plant Physiology **31,** 491-543.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., and Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genetics **6,** e1000940.

Bratteler, M., Lexer, C., and Widmer, A. (2006). Genetic architecture of traits associated with serpentine adaptation of *Silene vulgaris*. Journal of Evolutionary Biology **19,** 1149-1156.

Bräutigam, A., and Gowik, U. (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biology **12,** 831-841.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., and Charrad, M.M. (2014). Package 'NbClust'. J. Stat. Soft **61,** 1-36.

Chen, S., and Harmon, A.C. (2006). Advances in plant proteomics. Proteomics **6,** 5504-5516.

Conti, E., Soltis, D.E., Hardig, T.M., and Schneider, J. (1999). Phylogenetic relationships of the silver saxifrages (*Saxifraga*, sect. *Ligulatae* Haworth): Implications for the evolution of substrate specificity, life histories, and biogeography. Molecular Phylogenetics and Evolution **13,** 536-555.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F. (2004). A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **101,** 15243-15248.

Davey, M.P., Woodward, F.I., and Quick, W.P. (2009). Intraspecfic variation in cold-temperature metabolic phenotypes of *Arabidopsis lyrata* ssp. *petraea*. Metabolomics **5,** 138-149.

Dray, S., and Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. Journal of statistical software **22,** 1-20.

Dunn, W.B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. Physical biology **5,** 011001.

Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R., and Griffin, J.L. (2011). Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chemical Society Reviews **40,** 387-426.

Dunn, W.B., Erban, A., Weber, R.J., Creek, D.J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., and Kopka, J. (2013). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics **9,** 44-66.

Fernandes, J., Morrow, D.J., Casati, P., and Walbot, V. (2008). Distinctive transcriptome responses to adverse environmental conditions in *Zea mays* L. Plant Biotechnology Journal **6,** 782-798.

Fiehn, O. (2002). Metabolomics–the link between genotypes and phenotypes. Plant molecular biology **48,** 155-171.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. Nature biotechnology **18,** 1157-1161.

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J., and Wilczek, A.M. (2011). A map of local adaptation in *Arabidopsis thaliana*. Science **334,** 86-89.

Fowler, S., and Thomashow, M.F. (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. The Plant Cell **14,** 1675-1690.

Franks, S.J., and Hoffmann, A.A. (2012). Genetics of climate change adaptation. Annu Rev Genet **46,** 185-208.

Franks, S.J., Weber, J.J., and Aitken, S.N. (2014). Evolutionary and plastic responses to climate change in terrestrial plant populations. Evolutionary Applications **7,** 123-139.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. (1998). Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. The Plant Journal **16,** 433-442.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., and Bino, R. (2002). Plant metabolomics the missing link in functional genomics strategies (American Society of Plant Biologists).

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., Toomajian, C., Roux, F., and Bergelson, J. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. Science **334,** 83-86.

Hannah, M.A., Wiese, D., Freund, S., Fiehn, O., Heyer, A.G., and Hincha, D.K. (2006). Natural genetic variation of freezing tolerance in *Arabidopsis*. Plant Physiology **142,** 98-112.

Hartigan, J.A. (1975). Clustering algorithms. (New York: Wiley).

Hijmans, R.J. (2016a). Geosphere: Spherical Trigonometry [R package version 1.5-5].

Hijmans, R.J. (2016b). Geographic Data Analysis and Modeling [R package raster version 2.5-8].

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International journal of climatology **25,** 1965-1978.

Holtorf, H., Guitton, M.-C., and Reski, R. (2002). Plant functional genomics. Naturwissenschaften **89,** 235-249.

Huang, Q., Beharav, A., Li, Y., Kirzhner, V., and Nevo, E. (2002). Mosaic microecological differential stress causes adaptive microsatellite divergence in wild barley, *Hordeum spontaneum*, at Neve Yaar, Israel. Genome **45,** 1216-1229.

Hungerer, K.B., and Kadereit, J.W. (1998). The phylogeny and biogeography of Gentiana L. sect. Ciminalis (Adans.) Dumort.: a historical interpretation of distribution ranges in the European high mountains. Perspectives in Plant Ecology, Evolution and Systematics **1,** 121-135.

Janská, A., Maršík, P., Zelenková, S., and Ovesná, J. (2010). Cold stress and acclimation–what is important for metabolic adjustment? Plant Biology **12,** 395-405.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24,** 1403-1405.

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics **27,** 3070-3071.

Jombart, T., and Collins, C. (2015). A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0.

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics **11,** 1.

Jump, A.S., and Penuelas, J. (2005). Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters **8,** 1010-1020.

Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y., and Guy, C.L. (2004). Exploring the temperature-stress metabolome of *Arabidopsis*. Plant physiology **136,** 4159-4168.

Kappen, L., and Ullrich, W.R. (1970). Verteilung von Chlorid und Zuckern in Blattzellen halophiler Pflanzen bei verschieden hoher Frostresistenz. Berichte der Deutschen Botanischen Gesellschaft **83,** 265-275.

Kassambara, A. (2015). Factoextra: extract and visualize the results of multivariate data analyses. R package version **1**.

Klotke, J., Kopka, J., Gatzke, N., and Heyer, A. (2004). Impact of soluble sugar concentrations on the acquisition of freezing tolerance in accessions of *Arabidopsis thaliana* with contrasting cold adaptation – evidence for a role of raffinose in cold acclimation. Plant, Cell & Environment **27,** 1395-1404.

Koster, K.L., and Lynch, D.V. (1992). Solute accumulation and compartmentation during the cold acclimation of *Puma rye*. Plant Physiology **98,** 108-113.

Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. Journal of experimental botany **63,** 1593-1608.

Krzanowski, W.J., and Lai, Y. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. Biometrics**,** 23-34.

Last, R.L., Jones, A.D., and Shachar-Hill, Y. (2007). Towards the plant metabolome and beyond. Nature Reviews Molecular Cell Biology **8,** 167-174.

Lebart, L., Morineau, A., and Piron, M. (2000). Statistique exploratoire multidimensionnelle. (Paris: Dunod Editeur).

Li, B., Wei, A., Song, C., Li, N., and Zhang, J. (2008). Heterologous expression of the TsVP gene improves the drought resistance of maize. Plant Biotechnology Journal **6,** 146-159.

Li, Y.-C., Fahima, T., Krugman, T., Beiles, A., Röder, M.S., Korol, A.B., and Nevo, E. (2000). Parallel microgeographic patterns of genetic diversity and divergence revealed by allozyme, RAPD, and microsatellites in *Triticum dicoccoides* at Ammiad, Israel. Conservation Genetics **1,** 191-207.

Li, Y., Fahima, T., Beiles, A., Korol, A., and Nevo, E. (1999). Microclimatic stress and adaptive DNA differentiation in wild emmer wheat, *Triticum dicoccoides*. Theoretical and Applied Genetics **98,** 873-883.

Li, Y., Krugman, T., Fahima, T., Beiles, A., Korol, A., and Nevo, E. (2001). Spatiotemporal allozyme divergence caused by aridity stress in a natural population of wild wheat, *Triticum dicoccoides*, at the Ammiad microsite, Israel. Theoretical and Applied Genetics **102,** 853-864.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006). Gas chromatography mass spectrometry–based metabolite profiling in plants. Nature protocols **1,** 387-396.

Maggio, A., Miyazaki, S., Veronese, P., Fujita, T., Ibeas, J.I., Damsz, B., Narasimhan, M.L., Hasegawa, P.M., Joly, R.J., and Bressan, R.A. (2002). Does proline accumulation play an active role in stress-induced growth reduction? The plant journal **31,** 699-712.

Maruyama, K., Takeda, M., Kidokoro, S., Yamada, K., Sakuma, Y., Urano, K., Fujita, M., Yoshiwara, K., Matsukura, S., and Morishita, Y. (2009). Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. Plant physiology **150,** 1972-1980.

Meyer, R.C., Steinfath, M., Lisec, J., Becher, M., Witucka-Wall, H., Törjék, O., Fiehn, O., Eckardt, Ä., Willmitzer, L., and Selbig, J. (2007). The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **104,** 4759-4764.

Mushtaq, M.Y., Marçal, R.M., Champagne, D.L., van der Kooy, F., Verpoorte, R., and Choi, Y.H. (2014). Effect of acute stresses on zebra fish (*Danio rerio*) metabolome measured by NMR-based metabolomics. Planta Medica **80,** 1227-1233.

Nadeau, N.J., and Jiggins, C.D. (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends in Genetics **26,** 484-492.

Nanjo, T., Kobayashi, M., Yoshiba, Y., Sanada, Y., Wada, K., Tsukaya, H., Kakubari, Y., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Biological functions of proline in morphogenesis and osmotolerance revealed in antisense transgenic *Arabidopsis thaliana*. The Plant Journal **18,** 185-193.

Oh, S.J., Kwon, C.W., Choi, D.W., Song, S.I., and Kim, J.K. (2007). Expression of barley HvCBF4 enhances tolerance to abiotic stress in transgenic rice. Plant Biotechnology Journal **5,** 646-656.

Oleksyk, T.K., Smith, M.W., and O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society of London B: Biological Sciences **365,** 185-205.

Owuor, E.D., Fahima, T., Beiles, A., Korol, A., and Nevo, E. (1997). Population genetic response to microsite ecological stress in wild barley, *Hordeum spontaneum*. Molecular Ecology **6,** 1177-1187.

Parvaiz, A., and Satyawati, S. (2008). Salt stress and phyto-biochemical responses of plants-a review. Plant Soil and Environment **54,** 89.

Preston, J.C., and Sandve, S.R. (2013). Adaptation to seasonality and the winter freeze. Front Plant Sci **4,** 167.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A.R. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. The Plant Cell **13,** 11-29.

Sagisaka, S. (1987). Amino acid pools in herbaceous plants at the wintering stage and at the beginning of growth. Plant and cell physiology **28,** 171-178.

Sagisaka, S., and Araki, T. (1983). Amino acid pools in perennial plants at the wintering stage and at the beginning of growth. Plant and Cell Physiology **24,** 479-494.

Santarius, K., and Milde, H. (1977). Sugar compartmentation in frost-hardy and partially dehardened cabbage leaf cells. Planta **136,** 163-166.

Savolainen, O., Lascoux, M., and Merila, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics **14,** 807-820.

Schauer, N., and Fernie, A.R. (2006). Plant metabolomics: towards biological function and mechanism. Trends in plant science **11,** 508-516.

Schulz, E., Tohge, T., Zuther, E., Fernie, A.R., and Hincha, D.K. (2016). Flavonoids are determinants of freezing tolerance and cold acclimation in *Arabidopsis thaliana*. Scientific Reports **6**.

Shulaev, V., Cortes, D., Miller, G., and Mittler, R. (2008). Metabolomics for plant stress response. Physiologia Plantarum **132,** 199-208.

Smallwood, M., and Bowles, D.J. (2002). Plants in a cold climate. Philosophical Transactions of the Royal Society of London B: Biological Sciences **357,** 831-847.

Somerville, C., and Somerville, S. (1999). Plant functional genomics. Science **285,** 380-383.

Stapley, J., Reger, J., Feulner, P.G., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. Trends in ecology & evolution **25,** 705-712.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F., De Siqueira, M.F., Grainger, A., and Hannah, L. (2004). Extinction risk from climate change. Nature **427,** 145-148.

Thomashow, M.F. (1999). Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annual review of plant biology **50,** 571-599.

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T., and Nuzhdin, S.V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nature genetics **42,** 260-263.

Unal, Y., Kindap, T., and Karaca, M. (2003). Redefining the climate zones of Turkey using cluster analysis. International journal of climatology **23,** 1045-1055.

Urano, K., Kurihara, Y., Seki, M., and Shinozaki, K. (2010). 'Omics' analyses of regulatory networks in plant abiotic stress responses. Current opinion in plant biology **13,** 132-138.

Usadel, B., Blaesing, O.E., Gibon, Y., Poree, F., Hoehne, M., Guenter, M., Trethewey, R., Kamlage, B., Poorter, H., and Stitt, M. (2008). Multilevel genomic analysis of the response of transcripts, enzyme activities and metabolites in *Arabidopsis* rosettes to a progressive decrease of temperature in the non-freezing range. Plant, Cell & Environment **31,** 518-547.

Verbruggen, N., and Hermans, C. (2008). Proline accumulation in plants: a review. Amino acids **35,** 753-759.

Visser, M.E. (2008). Keeping up with a warming world; assessing the rate of adaptation to climate change. Proceedings of the Royal Society of London B: Biological Sciences **275,** 649-659.

Walther, G.-R., Berger, S., and Sykes, M.T. (2005). An ecological 'footprint'of climate change. Proceedings of the Royal Society of London B: Biological Sciences **272,** 1427-1432.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J.-M., Hoegh-Guldberg, O., and Bairlein, F. (2002). Ecological responses to recent climate change. Nature **416,** 389-395.

Ward Jr, J.H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American statistical association **58,** 236-244.

Zhou, Y., Zhang, L., Liu, J., Wu, G., and Savolainen, O. (2014). Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. Molecular ecology **23,** 3504-3522.

Availability of additional material

Additional data CD contains the following files:

Chapter 1:

Supplemental Data Set I-1:	Detailed results of *Cochlearia* flow cytometry measurements generated by R. Schulz and E. Wolf
Supplemental Data Set I-2:	Results of Cochlearia chromosome counts made by R. Schulz, P. Sack and T. Mandáková
Supplemental Data Set I-3:	Results of the survey of published and own chromosome numbers and genome sizes with geographical distribution
Supplemental Data Set I-4:	Results of simple linear regression analyses as well as rank correlation tests performed in R

Chapter 2:

Supplemental Data Set II-1:	Detailed sequencing results
Supplemental Data Set II-2:	Chloroplast genome alignment for phylogenetic analyses
Supplemental Data Set II-3:	Mitochondrial genome alignment for phylogenetic analyses
Supplemental Data Set II-4:	Annotated mitochondrial *de novo* contigs in Genbank Flat File Format (*.gb) – used as a mitochondrial genome reference
Supplemental Data Set II-5:	*Cochlearia pyrenaica de novo* transcriptome assembly in fasta format – used as a transcriptome reference
Supplemental Data Set II-6:	Sampling details of transcriptome-wide nuclear SNP calling datasets
Supplemental Data Set II-7:	Partitioning schemes for RAxML and BEAST analyses (chloroplast genome)
Supplemental Data Set II-8:	STRUCTURE input files
Supplemental Data Set II-9:	Alignment of transcriptome-wide nuclear SNPs for RAxML analysis of *Cochlearia* diploids
Supplemental Data Set II-10:	Alignment of transcriptome-wide nuclear SNPs for RAxML analysis of *Cochlearia* diploids and *Ionopsidium megalospermum*
Supplemental Data Set II-11:	Alignment of transcriptome-wide nuclear SNPs for SplitsTree analysis of *Cochlearia* samples
Supplemental Data Set II-12:	Full RAxML tree based on (nearly) complete chloroplast genomes

Supplemental Data Set II-13:	Full RAxML tree based on partial mitochondrial genomes
Supplemental Data Set II-14:	STRUCTURE result 'all samples' under *correlated allele frequencies*
Supplemental Data Set II-15:	SplitsTree network generated from biallelic sites only (107,269 variants)
Supplemental Data Set II-16:	Results of deltaK analyses used for detecting the optimal number of clusters in STRUCTURE analyses
Supplemental Data Set II-17:	STRUCTURE result 'all diploids' under un*correlated allele frequencies*
Supplemental Data Set II-18:	STRUCTURE result 'European diploids' under un*correlated allele frequencies*

Chapter 3:

Supplemental Data Set III-1:	R script (written by F. Michling, unpublished) used for the extraction of bioclimatic variables from WorldClim grids
Supplemental Data Set III-2:	Bioclimatic variables extracted for 27 populations included in the metabolomic analyses
Supplemental Data Set III-3:	D index plot, generated by NbClust analysis for determining the optimal number of clusters for bioclimatic data
Supplemental Data Set III-4:	Results of NbClust analysis for determining the optimal number of clusters for bioclimatic data
Supplemental Data Set III-5:	Raw and normalized concentrations extracted for 40 selected compounds
Supplemental Data Set III-6:	ANOVA results for metabolomic compound analyses grouped by bioclimatic clusters
Supplemental Data Set III-7:	DAPC analyses input file
Supplemental Data Set III-8:	Results of cross validations for different DAPC analyses
Supplemental Data Set III-9:	Contributions of individual metabolites to the respective DAPC runs

Acknowledgements

I have been lucky enough to have experienced the help and support of many people without whom this work would not have been possible and who need to be acknowledged and thanked here:

Professor Dr. Marcus A. Koch – for trusting me with this thesis, for his continued optimistic support and encouragement and for all the invaluable discussions and patient explanations on this fascinating genus he knows so well.

Professor Dr. Claudia Erbar – for taking the time to be my second supervisor and for her constant encouragement and support during these last years. Moreover, I'd like to thank both Professor Dr. Claudia Erbar and Professor Dr. Peter Leins for taking me in as a student assistant in their plant biodiversity course. I really enjoyed being a part of it and besides teaching the students, I learned a lot of things in there myself.

Dr. Markus Kiefer – for joining my thesis advisory committee, and for his constant support in all the countless computer problems I experienced or caused during the last years.

Dr. Emmanuel Gaquerel – for the great collaboration and his kind support in any metabolomics related question.

Dr. Lúa López – for being such a wonderful *Cochlearia* team member, for building me a great transcriptome reference, for joining me through mud and snow on our adventurous sampling trips and for being the best driving and singing partner during those, for being my statistics Obi-Wan and most importantly for being a true soulmate and friend.

Rainer Schulz – for completing the 'team *Cochlearia*', for his diligent work and his inspiring conscientiousness during his master thesis, for sharing his wisdom, his wit and his friendship with me.

Dr. Nora Hohmann – for being the best companion I could have asked for during the ups and downs of NGS data analysis and for being such an amazing office roommate, couch partner, cookie provider and friend.

Florian Michling – for his great assistance with any R related problem, for always supporting me whenever support of any kind was needed, and for many good talks about work, life and everything in between.

David Ibberson – for the great and friendly support during the sequencing of my samples.

Dr. Martin Lysak and Dr. Terezie Mandáková – for the nice collaboration on *Cochlearia* cytogenetics.

Torsten Jakob – for devotedly taking care of the *Cochlearia* collection in the Botanical Garden.

Dr. Peter Sack – for being the "good soul" of the group, for always helping out and assisting me with cytogenetics, herbarium vouchers, seed samples and more; and last but not least, for being such a nice lunch and porridge companion.

Chao Yang – for offering to water my plants on several weekends, for always being helpful and for being a really nice colleague.

Lisa Kretz - for her support in any lab related issues and for many nice lab talks.

All other current and former group members for creating a very nice working atmosphere and for all the great times we had.

The people who shared their data or plant material with me: Chris Pires, Pat Edger, Michal Ronikier, Judita Kochjarova, Lovisa Gustafsson, Estelle Gill, Inger Nordal, Peter Harsch, and many more.

My friends outside the lab – because this journey would have not been the same without them.

My whole family and especially my parents – for all their love and support in everything I do.

My husband Moritz – for his encouragement, his patience, his unconditional support and for many other things but above all for making our life as wonderful as it is.

And finally, our son Jakob who joined me on the last steps of this journey – for being both the best motivation and the most lovely distraction during these past months of thesis writing. This thesis is dedicated to him.