

RESEARCH ARTICLE

Open Access

The evolutionary history of the *Arabidopsis lyrata* complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier

Roswitha Schmickl¹, Marte H Jørgensen², Anne K Brysting², Marcus A Koch^{1*}

Abstract

Background: The genomes of higher plants are, on the majority, polyploid, and hybridisation is more frequent in plants than in animals. Both polyploidisation and hybridisation contribute to increased variability within species, and may transfer adaptations between species in a changing environment. Studying these aspects of evolution within a diversified species complex could help to clarify overall spatial and temporal patterns of plant speciation. The *Arabidopsis lyrata* complex, which is closely related to the model plant *Arabidopsis thaliana*, is a perennial, outcrossing, herbaceous species complex with a circumpolar distribution in the Northern Hemisphere as well as a disjunct Central European distribution in relictual habitats. This species complex comprises three species and four subspecies, mainly diploids but also several tetraploids, including one natural hybrid. The complex is ecologically, but not fully geographically, separated from members of the closely related species complex of *Arabidopsis halleri*, and the evolutionary histories of both species complexes have largely been influenced by Pleistocene climate oscillations.

Results: Using DNA sequence data from the nuclear encoded cytosolic phosphoglucosomerase and Internal Transcribed Spacers 1 and 2 of the ribosomal DNA, as well as the *trnL/F* region from the chloroplast genome, we unravelled the phylogeography of the various taxonomic units of the *A. lyrata* complex. We demonstrate the existence of two major gene pools in Central Europe and Northern America. These two major gene pools are constructed from different taxonomic units. We also confirmed that *A. kamchatica* is the allotetraploid hybrid between *A. lyrata* and *A. halleri*, occupying the amphi-Beringian area in Eastern Asia and Northern America. This species closes the large distribution gap of the various other *A. lyrata* segregates. Furthermore, we revealed a threefold independent allopolyploid origin of this hybrid species in Japan, China, and Kamchatka.

Conclusions: Unglaciated parts of the Eastern Austrian Alps and arctic Eurasia, including Beringia, served as major glacial refugia of the Eurasian *A. lyrata* lineage, whereas *A. halleri* and its various subspecies probably survived in refuges in Central Europe and Eastern Asia with a large distribution gap in between. The North American *A. lyrata* lineage probably survived the glaciation in the southeast of North America. The dramatic climatic changes during glaciation and deglaciation cycles promoted not only secondary contact and formation of the allopolyploid hybrid *A. kamchatica*, but also provided the environment that allowed this species to fill a large geographic gap separating the two genetically different *A. lyrata* lineages from Eurasia and North America. With our example focusing on the evolutionary history of the *A. lyrata* species complex, we add substantial information to a broad evolutionary framework for future investigations within this emerging model system in molecular and evolutionary biology.

* Correspondence: mkoch@hip.uni-hd.de

¹Heidelberg University, Heidelberg Institute of Plant Sciences, Biodiversity and Plant Systematics, Im Neuenheimer Feld 345, D-69120 Heidelberg, Germany

Background

Molecular biological research during the last decade has largely focussed on model organisms such as *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Arabidopsis thaliana*. Now that knowledge in molecular genetics, cell and developmental biology of these organisms has greatly increased, closely related organisms emerge as promising for studying characteristics not possible to elucidate with model and/or single organisms [1,2].

Arabidopsis lyrata L. is a close relative of *A. thaliana*, from which it diverged approximately five million years ago [3,4]. *Arabidopsis lyrata* s.l. represents a small species complex of four species and several putative subspecies with a circumpolar arctic-alpine distribution (Additional file 1, Table S1). Populations have been adapted to various ecological conditions, including the harsh environment of the arctic tundra, cryptic warm-stage refugia (exposed rocks, rocky slopes) in Central Europe, and different edaphic conditions with substrates such as dolomite, silicious bedrocks, and even heavy metal rich serpentine soil in Central Europe (Lower Austria, personal observation) and the USA [Maryland; [5]] [6]. Most members of the species complex are perennial diploid outbreeders ($2n = 2x = 16$), but also tetraploid cytotypes occur [7,6,8]. There are numerous aspects of the biology of the *A. lyrata* complex that differ from and cannot be addressed in *A. thaliana*, like self-incompatibility and perennial life cycle.

The *A. lyrata* complex has already proven to be a suitable study system for the analysis of character traits such as flowering time [9,10] or pathogen defense [11]. Additionally, molecular mechanisms for the function of sporophytic self-incompatibility have been investigated [12-20], and comparative approaches to analyse sporophytic self-incompatibility in diploids versus polyploids are underway (Jørgensen, unpublished data). Whole genome sequencing of *A. lyrata* was finished last year, and data have been available for a few months (The *A. lyrata* genome sequence assembly v1.0, <http://genome.jgi-psf.org/Araly1/Araly1.info.html>), enabling direct comparisons with the *A. thaliana* genome. However, in contrast to *A. thaliana*, where the evolutionary history has been analysed in more detail [e.g. [21,22]], evolutionary studies on the *A. lyrata* complex have so far been largely restricted to a small number of populations from Central Europe [23-26] or larger sample sizes with a more general genus-wide perspective [27,28]. A detailed European study has revealed that population structure is dominated by regional genetic bottlenecks, and genetic structure exists within continents [Ansell, personal communication]. This suggests a comprehensive global study is necessary to resolve the evolutionary history of this complex.

In this study we present the first worldwide evolutionary history of the *A. lyrata* complex, covering its whole range of distribution and all taxonomically defined units. We use a widely applied nuclear encoded marker system (ITS, internal transcribed spacer region of nuclear encoded ribosomal DNA) to study gene flow between populations, a maternally inherited chloroplast genome marker (*trnL* intron (*trnL*) and *trnL*/F intergenic spacer (*trnL*/F-IGS) of tRNA^{Ser} and tRNA^{Thr}, respectively) to investigate migrational movements due to seed dispersal, and the nuclear encoded housekeeping gene *PgiC* (cytosolic phosphoglucosomerase), a single copy gene, to discriminate between hybridising taxa. We aim to focus on the following four aspects: (1) Unravelling general phylogeographic patterns of the *A. lyrata* complex by identifying the main genetic lineages, and interpreting genetic variation in space and time [29], in the context of both climatic and geological events throughout Pleistocene glaciation cycles; (2) Evaluating the role of hybridisation and polyploidisation in the origin of *A. kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh, an amphi-Beringian member of the *A. lyrata* complex; (3) Explaining Pleistocene and postglacial migration routes by analysing genetic diversity statistics: The arctic-alpine *A. lyrata* complex is one of the rare examples among higher plants with a distribution in both Central Europe and North America and, additionally, a circumpolar distribution - other examples are *Cassiope tetragona* [30], and *Saxifraga oppositifolia* [31]; and (4) Studying the role of Beringia as a refugia for populations of arctic *A. lyrata*, since Beringia is assumed to be one of the major refugia for arctic plants during Pleistocene glaciations [30-36].

Methods

Plant material

Altogether 467 accessions of the *A. lyrata* complex were analysed: 295 accessions newly analysed within this study, 39 accessions sequenced by Schmickl et al. [28], and 133 accessions analysed by Koch and Matschinger [27]. Plant material was mainly collected from herbarium vouchers from BM (Natural History Museum, London), CAS (California Academy of Sciences, San Francisco), DAO (Vascular Plant Herbarium, Agriculture and Agri-Food Canada, Ottawa), DH (Hobert and William Smith Colleges, New York), LE (The V.L. Komarov Botanical Institute, Russian Academy of Sciences, St. Petersburg), LI (Upper Austrian Provincial Museum, Linz), O (Natural History Museum, University of Oslo, Oslo), W (Natural History Museum, Vienna), and partly collected in the field, documented at HEID (Herbarium University of Heidelberg, Heidelberg). Taxon determination followed the voucher labels, and was verified with

floras and determination keys [e.g. [37]; Flora of North America, Al-Shehbaz, personal communication]. In our study we followed the taxonomy of [37]. Twentyfive accessions of the various subspecies of *A. halleri* (L.) O'Kane & Al-Shehbaz were analysed from throughout the distribution range because of the evidence that this taxon served as a putative parent of *A. kamchatica*, previously also treated as *A. lyrata* ssp. *kamchatica* [[27]; and references therein], an evolutionary scenario which

was recently confirmed by [8]. The distribution of the investigated accessions is shown in Figure 1. The accession list is provided with Additional file 2, Table S2. The following short overview of *A. lyrata* and *A. halleri* taxonomy will introduce the concept of Al-Shehbaz and O'Kane [37]. Elven (ed.) [38] pursues a different taxonomic concept, which is summarised in the supplementary material (Additional file 1, Table S1), but will not be discussed here. Apart from *A. thaliana*, the *A. lyrata*

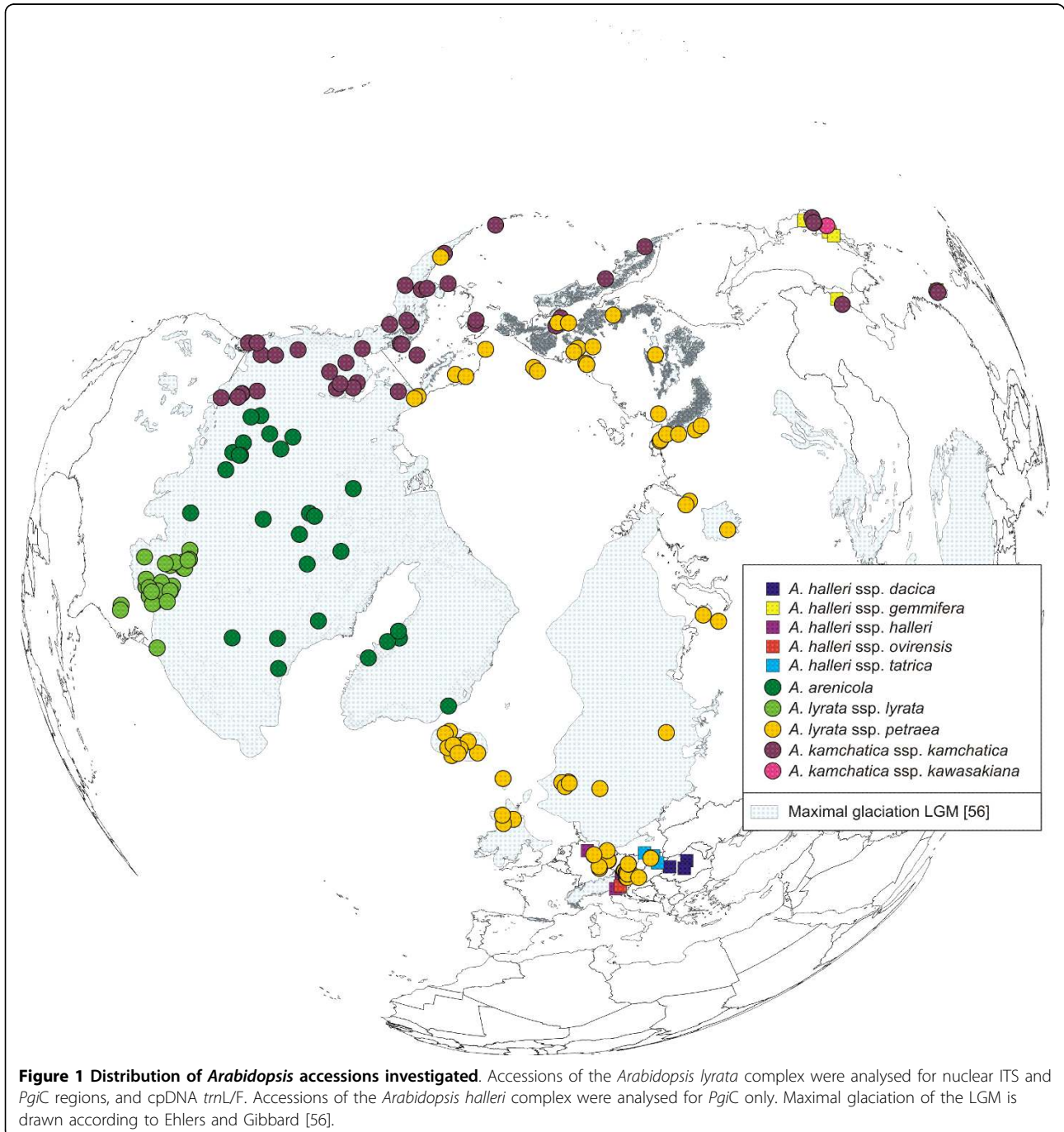


Figure 1 Distribution of *Arabidopsis* accessions investigated. Accessions of the *Arabidopsis lyrata* complex were analysed for nuclear ITS and *PgiC* regions, and cpDNA *trnL/F*. Accessions of the *Arabidopsis halleri* complex were analysed for *PgiC* only. Maximal glaciation of the LGM is drawn according to Ehlers and Gibbard [56].

complex is one of three major species complexes in the genus *Arabidopsis* [27], the other two being the *A. arenosa* (L.) Lawalrée and *A. halleri* complexes [39,40,6,27,41]. The *A. lyrata* complex is considered by different authors to include different numbers of taxa of various ranking, distribution areas, and ploidy levels (Additional file 1, Table S1). Al-Shehbaz and O’Kane [37] treat the complex as one species, *A. lyrata*, with three subspecies: ssp. *lyrata*, ssp. *petraea* (L.) O’Kane & Al-Shehbaz, and ssp. *kamchatica* (Fisch. ex DC.) O’Kane & Al-Shehbaz. Subspecies *lyrata* is considered to be broadly amphi-Pacific and of two ploidy levels ($2n = 16, 32$), ssp. *petraea* to be Northern Eurasian and Central European with the same two ploidy levels, and *kamchatica* to be amphi-Pacific and tetraploid ($2n = 32$). A comprehensive quantitative morphological analysis using multivariate statistics is still not available for the *A. lyrata* complex. Moreover, information on the ploidy level is largely lacking throughout the distribution range, and further cytological investigations might be important for taxon delimitation [7,8]. *Arabidopsis arenicola* (Richardson) Al-Shehbaz, Elven, D.F. Murray & Warwick has only recently been included as part of the *A. lyrata* complex [42]. It has, for a long time, been placed within the genus *Arabis* L. Elven (ed.) [38] considers this taxon to be diploid ($2n = 16$) and distributed in north-eastern North America, and they are supported by Al-Shehbaz in the upcoming Flora of North America [Al-Shehbaz, personal communication]. Many of the mentioned taxa include two ploidy levels, suggesting frequent polyploidisation events within the *A. lyrata* complex. An allopolyploid origin of *kamchatica* has already been confirmed based on nuclear DNA sequences, with *A. lyrata* and *A. halleri* ssp. *gemmifera* (Matsum.) O’Kane & Al-Shehbaz as possible parental taxa [[43,27,28,8]; Jørgensen et al., unpublished data]. Otherwise, little is known with regard to the number of polyploid units and their origins.

Five subspecies have been recognised in *A. halleri*: ssp. *halleri*, ssp. *ovirensis* (Wulfen) O’Kane & Al-Shehbaz, ssp. *dacica* (Heuff.) Kolník, comb. nov., ssp. *tatrica* (Pawl.) Kolník, comb. nov., all distributed in Central Europe, and ssp. *gemmifera* in Eastern Asia, supported by both morphometric analysis (Kólník, unpublished data) and genetic AFLP data (Marhold, unpublished data).

DNA isolation, amplification and sequencing

Total DNA was obtained from dried leaf material and extracted according to the CTAB protocol of Doyle and Doyle [44] with the following modifications: 50-75 mg of dry leaf tissue were ground in 2 ml tubes using a Retsch swing mill (MM 200), 2 units of RNase A per extraction were added to the isolation buffer, and the

DNA pellets were washed twice with 70% ethanol. DNA was dissolved in 50 μ l TE-buffer for storage and diluted 1:3 in TE-buffer before use.

For the cpDNA markers *trnL* intron and *trnL/F* intergenic spacer (*trnL/F*-IGS), primers and PCR cycling scheme followed the protocol of Dobeš et al. [45], using a PTC200 (MJ Research, Waltham, USA) thermal cycler. The PCR reaction volume of 50 μ l contained 1 \times PCR buffer (10 mM TRIS/50 mM KCl buffer, pH 8.0), 3 mM MgCl₂, 0.4 μ M of each primer, 0.2 mM of each dNTP, 1 U Taq DNA polymerase (Amersham Biosciences, Chalfont St Giles, England), and approximately 1 ng of template DNA. Amplified sequences of *trnL/F*-IGS included the complete *trnL/F*-IGS and the first 18 bases of the *trnF* gene. Amplification of the nuclear marker internal transcribed spacer region (ITS) was performed according to Dobeš et al. [46]. PCR reaction conditions were the same as for the two cpDNA markers described above, and PCR cycling scheme was 5 min at 95°C, 35 cycles of 1 min at 95°C, 1 min at 48°C, and 1 min at 72°C, 10 min extension at 72°C, and a final hold at 4°C. PCR products spanned the entire ITS1, 5.8 S rDNA, and ITS2 region.

Before sequencing PCR products were checked for length and concentrations on 1.5% agarose gels and purified with the NucleoFast Kit (Macherey-Nagel, Düren, Germany). Cycle sequencing was performed using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Chalfont St Giles, England) and the original primers. Both strands were amplified in order to gain the complete sequence. PCR products were resolved in 10 μ l loading solution and run on a MegaBace 500 sequencer.

trnL/F and ITS sequence definition and map reconstruction

Plastidic *trnL/F* sequences were defined as haplotypes and suprahaplotypes following our previous studies [27,41,28]: Haplotypes are characterised by multiple *trnF* pseudogenes in the 3’-region of the *trnL/F*-IGS close to the functional *trnF* gene [47,48,27,49,41,50]. When defining *trnL/F* suprahaplotypes, we excluded the pseudogene-rich region. Pseudogenes evolve with a 10 times higher mutation rate than single nucleotide polymorphisms, which makes them non-applicable for phylogenetic reconstructions on the species level [50]. Additionally, their boundaries are under ongoing discussion. An alternative interpretation is provided by Ansell et al. [51]. In summary, haplotypes belonging to one suprahaplotype share the same base order throughout the whole sequence except for the pseudogene-rich region, where they vary in both length and base content. Suprahaplotypes differ from each other only by single point mutations and/or indels. Newly defined *trnL/F* haplotypes were assigned to GenBank numbers [GQ922894-GQ922903] (Additional file 2, Table S2).

ITS sequences were obtained from direct sequencing of PCR products in order to detect hybrid individuals according to ambiguous base positions. Sequences were defined as ITS types and supratypes following our previous studies [27,41,28]: Most ITS types contained ambiguous sites as a result of the direct sequencing approach. They were combined to ITS supratypes by replacing the ambiguous sites by the bases with higher fluorescence intensity in the electropherogram. Sequences with equal fluorescence intensity of the two bases at the ambiguous positions were only found between ITS supratypes b and e and labelled b/e ambiguous. Koch and Matschinger [27] already showed that analyses with either ITS types or ITS supratypes produce the same results. However, working with a limited number of ITS supratypes in contrast to a vast number of ITS types contributes to a clearer data display. Multiple ITS types within a single individual indicate the natural variability among ITS loci within an individual as result of either mutations or gene flow between individuals. In these individuals concerted evolution, which denotes the process of DNA sequence homogenisation among members of multigene families by gene conversion and/or unequal crossing over [52-55], is not yet completed. Newly defined ITS types were assigned to GenBank numbers [GQ922904-GQ922910] (Additional file 2, Table S2).

Maps were constructed using BioOffice version 2.0.6 to create shapefiles and drawn with ArcView version 8.2. Shapefiles for visualising the maximum extent of the ice sheets during the LGM (Last Glacial Maximum) were provided by Ehlers and Gibbard [56].

Network analyses and genetic diversity statistics

Network analyses and genetic diversity statistics were exclusively performed using the *trnL/F* suprahaplotypes, as the pseudogene-rich region is not applicable for phylogenetic reconstructions on the species level [50]. The network was constructed using TCS version 1.21 [57] using the statistical parsimony algorithm [58]. Gaps (except polyT stretches) were coded as single additional binary characters. Newly characterised suprahaplotypes were added to the *trnL/F* network of the *A. lyrata* complex published earlier [28]. Genetic diversity statistics were performed with Arlequin version 3.11 [59]. Pairwise genetic differentiation was calculated among the following nine taxonomic and regional groups: *A. lyrata* ssp. *petraea* from (1) unglaciated Central Europe, (2) glaciated northern Europe, (3) northern Russia and western Beringia, (4) eastern Beringia; *A. lyrata* ssp. *lyrata* from (5) unglaciated North America and the glaciated Great Lakes region; (6) *Arabidopsis arenicola* from glaciated North America and Greenland; and *A. kamchatica* from (7) Japan, (8) north-eastern Russia, (9) Alaska

and western Canada. F_{ST} values, regarding haplotype frequencies only, and Φ_{ST} values, which take into account the genetic relationship between haplotypes as pairwise character differences, were calculated. The significance of differentiation was examined using a permutation test with 1000 permutations. Additionally, genetic diversity was estimated as nucleotide diversity π .

ITS parsimony analysis

ITS data were analysed based on a 652 bp alignment of ITS supratypes [27]. The total number of variable sites was 44, with 16 of them parsimony informative. From altogether 10 supratypes of both the *A. lyrata* complex and *A. halleri* ssp. *gemmifera*, a strict consensus tree was constructed using maximum parsimony with MEGA version 4.1 [60]. Heuristic searches were performed with 10 random addition sequences and Closest Neighbour Interchange (CNI) branch swapping. Bootstrap values were calculated based on 500 replicates. Length of the most parsimonious trees was 48 mutational steps with a consistency index (CI) = 0.89 and a retention index (RI) = 0.96 (autapomorphies excluded). *Arabidopsis thaliana* was used as outgroup.

Primer design for the nuclear marker *PgiC*

Various higher plants are known to have a duplicated locus of the cytosolic enzyme phosphoglucosomerase [61-63], and the loci are normally unlinked. Also within the genus *Arabidopsis*, extensive sequencing of the *PgiC* locus using a cloning strategy revealed a duplication, both in the *A. lyrata* and *A. halleri* complex, and this duplication must have predated the evolutionary split between these two species complexes (Additional file 3, Figure S1; Jørgensen, unpublished data). Both loci were initially simultaneously amplified with the general forward primer 5'-TGCTGTSAGCACTAATCTTGCG-3' and the general reverse primer 5'-TCGAACCCGGGAGAGGTAGACCA-3', following the protocol of Wright et al. [23]. The resulting sequence data showed that a group of alleles at the *PgiC1* locus were exclusively found in *A. halleri* and the allopolyploid *A. kamchatica*. It was thus possible to design *A. halleri*-specific primers that worked as a high-throughput and simple PCR-based screening marker to discriminate between genomes of the *A. lyrata* and *A. halleri* complex. Unfortunately, it was not possible to develop a PCR-based reciprocal marker system characterising alleles from the *A. lyrata* gene pool because of a lack of appropriate DNA sequence variation. In general, the alleles from both duplicated *PgiC* loci were only weakly differentiated among and within species. However, within the *PgiC1* locus we found a deletion of 7 bp length in the *A. lyrata* complex compared to the *A. halleri* complex (Additional file 4, Figure S2). Both groups of alleles

were also substantially differentiated by various SNPs. A primer pair with the forward primer located partly within this 7 bp indel (5'-CATTCAACAGATTGTG-3') and the reverse primer 5'-CCAGTAAACATCATGT-3' was developed to amplify a 92 bp fragment within the *PgiC1* locus (Additional file 5, Figure S3). The PCR reaction volume of 50 µl contained 1× PCR buffer (10 mM TRIS/50 mM KCl buffer, pH 8.0), 2.5 mM MgCl₂, 0.13 µM of each primer, 0.2 mM of each dNTP, 1 U Taq DNA polymerase (Amersham Biosciences, Chalfont St Giles, England), and approximately 1 ng of template DNA. The PCR cycling scheme was 3 min at 94°C, 35 cycles of 20 sec at 94°C, 30 sec at 56°C, and 20 sec at 68°C, 20 sec extension at 68°C, and a final hold at 4°C. By screening the absence/presence of the PCR product of this taxon-specific primer pair, we were able to follow the genetic footprint of the *A. halleri* complex in its allopolyploid hybrids throughout its distribution range.

Results

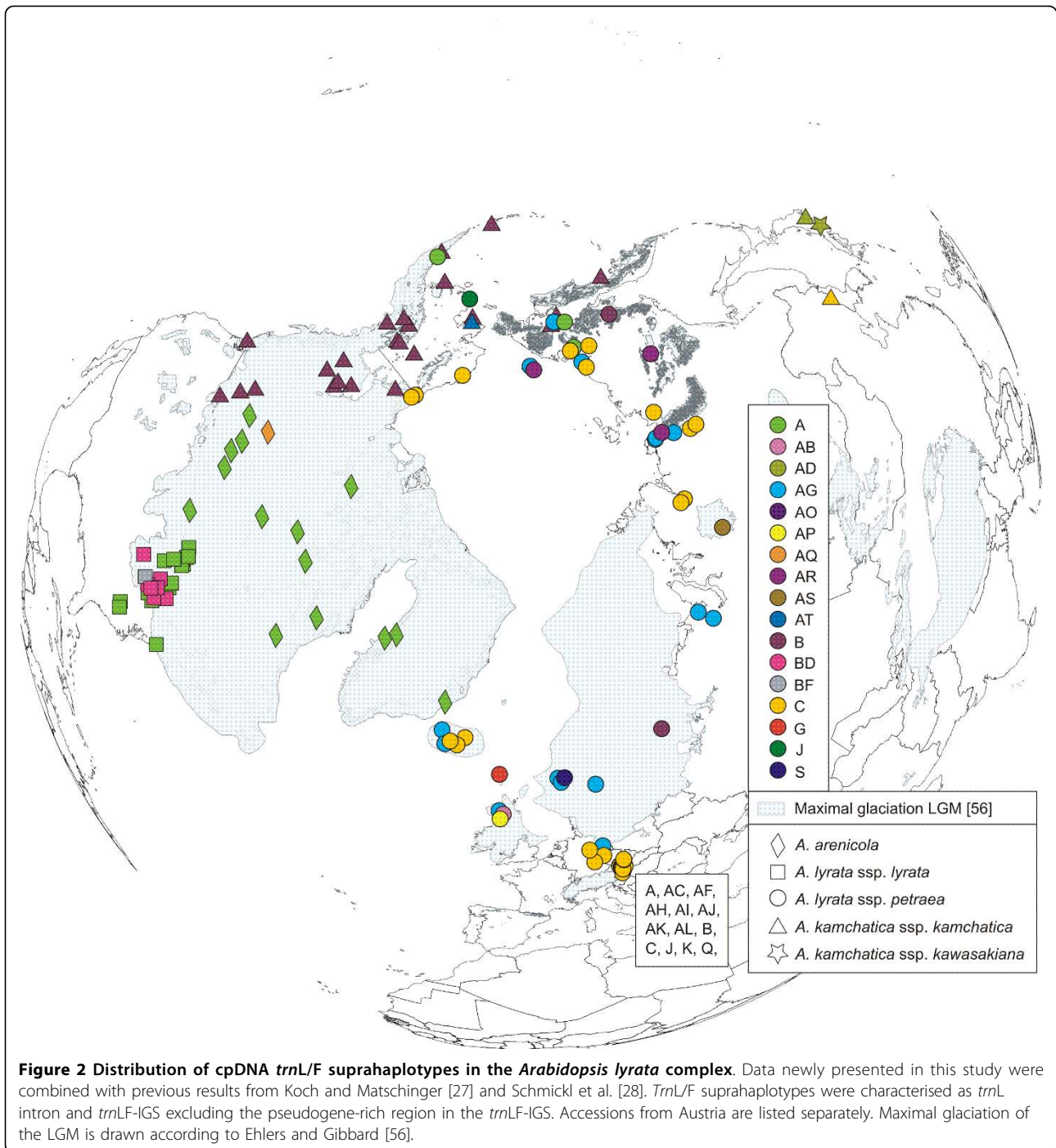
Chloroplast sequence data indicate three main genetic lineages: Eurasia, North America, and the amphi-Pacific region

The investigated accessions (Figure 1), spanning the whole distribution range of the *A. lyrata* complex, were grouped into three genetic and geographically separated lineages: *trnL/F* suprahaplotype C in Eurasia, A in North America, and B in the amphi-Pacific region (Figure 2). The Eurasian lineage with suprahaplotype C had the largest distribution range: unglaciated Central Europe, formerly glaciated northern Europe, arctic Russia, Beringia, and Alaska (north of Brooks Range). The North American lineage, characterised by suprahaplotype A, included the United States (mainly around the Great Lakes), northeastern and central Canada (to the Canadian Rocky Mountains in the west), and Greenland. The third and amphi-Pacific lineage spanned from Kamchatka via Beringia into western Canada (with the Canadian Rocky Mountains as eastern border). These three suprahaplotypes were central in the suprahaplotype network (Figure 3), and all were connected to derived and less widely distributed suprahaplotypes: In the Eurasian lineage AC was widespread and found in Central Europe, mainly Austria, AG predominantly in the north (Iceland, Scandinavia, Russia), and AR in western Beringia (incl. Wrangel Island) (Figure 2). Unique suprahaplotypes, occurring only once in the whole dataset, were observed in Scotland (AB, AS), Austria (AH, AI, AJ, AK, AL, K, R, V), Iceland (AO, AP), Faeroe Islands (G), and Sweden (S). The North American lineage was additionally characterised by the unique suprahaplotypes AQ and BD, and the more widespread BF. Within the amphi-Pacific lineage, AD was detected exclusively in Japan.

Although all three lineages were characterised both by lineage-specific central and “tip” suprahaplotypes, sharing of central suprahaplotypes was observed, e.g. suprahaplotypes A and B were detected in a few accessions of the Eurasian lineage (Figure 2, Additional file 6, Table S3). This finding is congruent with the observation of central suprahaplotype sharing between the three main species complexes of the genus, *A. lyrata*, *A. halleri*, and *A. arenosa* [27]. This observation has been explained by ancestral cpDNA polymorphism predating the radiation of the genus approximately two million years ago [27].

Cytosolic phosphoglucose isomerase identifies *Arabidopsis halleri* ssp. *gemmifera* as one parent of the allopolyploid amphi-Pacific *Arabidopsis kamchatica*

The various *PgiC* alleles detected were defined either as alleles at locus *PgiC1* or *PgiC2*, and, therefore, considered as locus-specific (Additional file 3, Figure S1). However, taxon-specific lineage sorting of the various allele pools after the duplication event was not complete for *PgiC2*. *A. kamchatica* carries *PgiC2* alleles hardly distinguishable from those of *A. septentrionalis* and *A. umbrosa*, which is an additional indicator that an Asian member of the *A. lyrata* complex served as one putative parental taxon. In addition, European *A. lyrata* ssp. *petraea* shares similar alleles with *A. halleri*, also indicating incomplete lineage sorting. The differentiation and lineage sorting of alleles at locus *PgiC1* is more taxon-specific. Here, *A. kamchatica* shares alleles most similar to those of *A. halleri* ssp. *gemmifera* from East Asia, and all of these alleles are significantly distinct from those of Eurasian *A. lyrata*. In summary, it is shown that *PgiC* is not only a suitable marker to screen for hybrid speciation in *A. kamchatica*, but might also be a suitable marker to follow these alleles through space and time. Amplification of *PgiC1* alleles was successful in all *A. kamchatica* accessions (Figure 4, Additional file 5, Figure S3), but, as outlined above, failed in Eurasian and North American members of the *A. lyrata* complex (Figure 4, Additional file 5, Figure S3). *PgiC1* amplification was positive in all *A. halleri* ssp. *gemmifera* accessions and additionally in European subspecies of *A. halleri* (ssp. *dacica*, ssp. *halleri*, ssp. *tatrica*) (Figure 4, Additional file 5, Figure S3). Hence, presence of *PgiC1* alleles without deletion in the forward primer sequence was characteristic for the whole *A. halleri* species complex, except for *A. halleri* ssp. *ovirensis* (data not shown). In this subspecies either a secondary loss of this locus or a complementary mutation in the primer binding site might have occurred. However, *A. halleri* ssp. *ovirensis* is a genetically distinct, local endemite at one single place in the southeastern Austrian Alps with an unclear evolutionary history [27,41].



Gene diversity statistics show highest genetic diversity in the Eurasian lineage, strongly reduced diversity in the North American lineage, and extremely low diversity in the allopolyploid amphipacific lineage

For genetic diversity statistics the distribution of the *A. lyrata* species complex was divided into nine different groups, according to the evidence for genetic lineages and Pleistocene history (Table 1). The Eurasian lineage

(*A. lyrata* ssp. *petraea*) was split into four groups: unglaciated Central Europe, previously glaciated northern Europe or permafrost areas, northern Russia/western Beringia, and eastern Beringia. The North American lineage (*A. lyrata* ssp. *lyrata*) was separated into unglaciated North America/glaciated Great Lakes region and glaciated North America/Greenland. Amphipacific *A. kamchatica* was differentiated into groups from Japan,

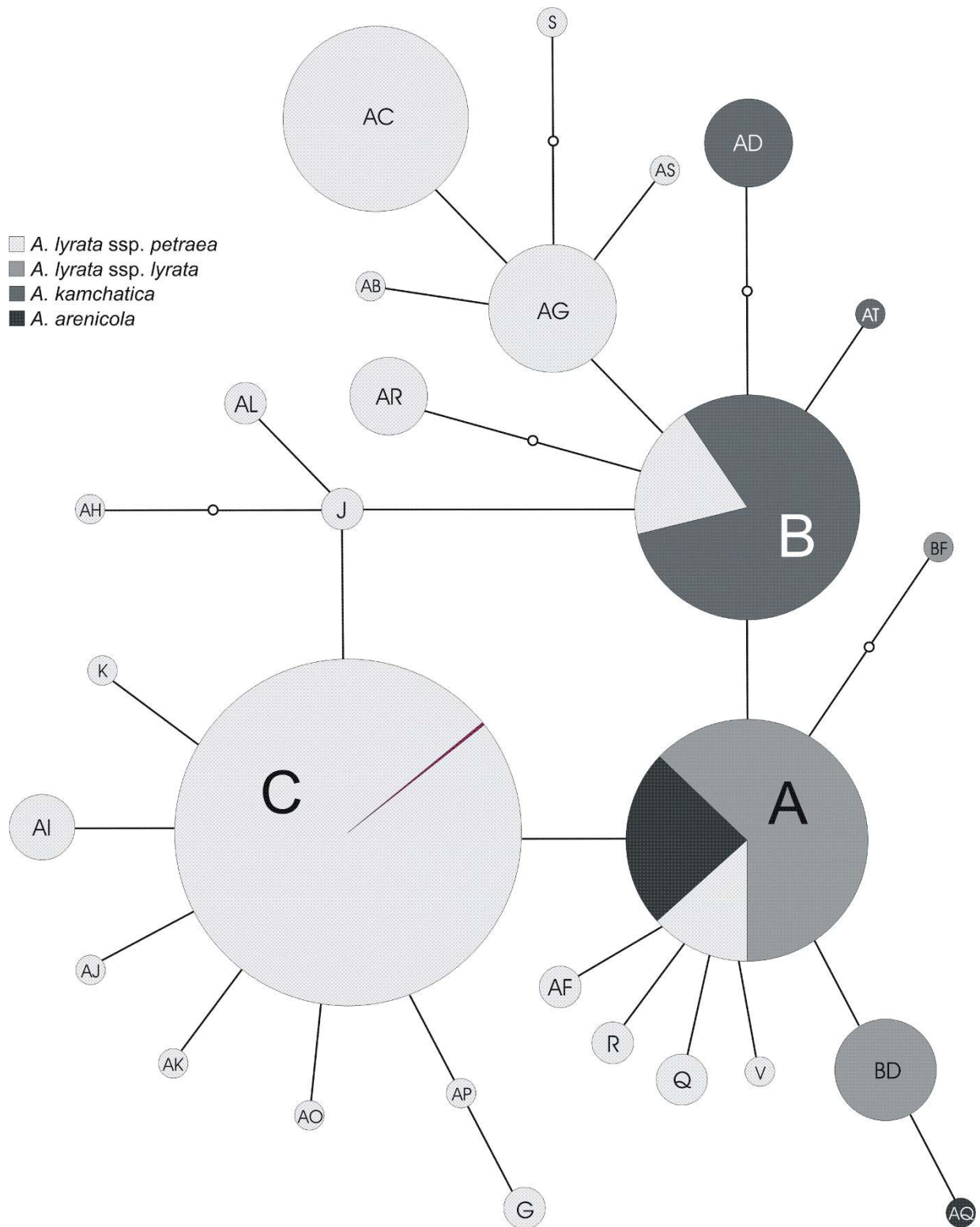
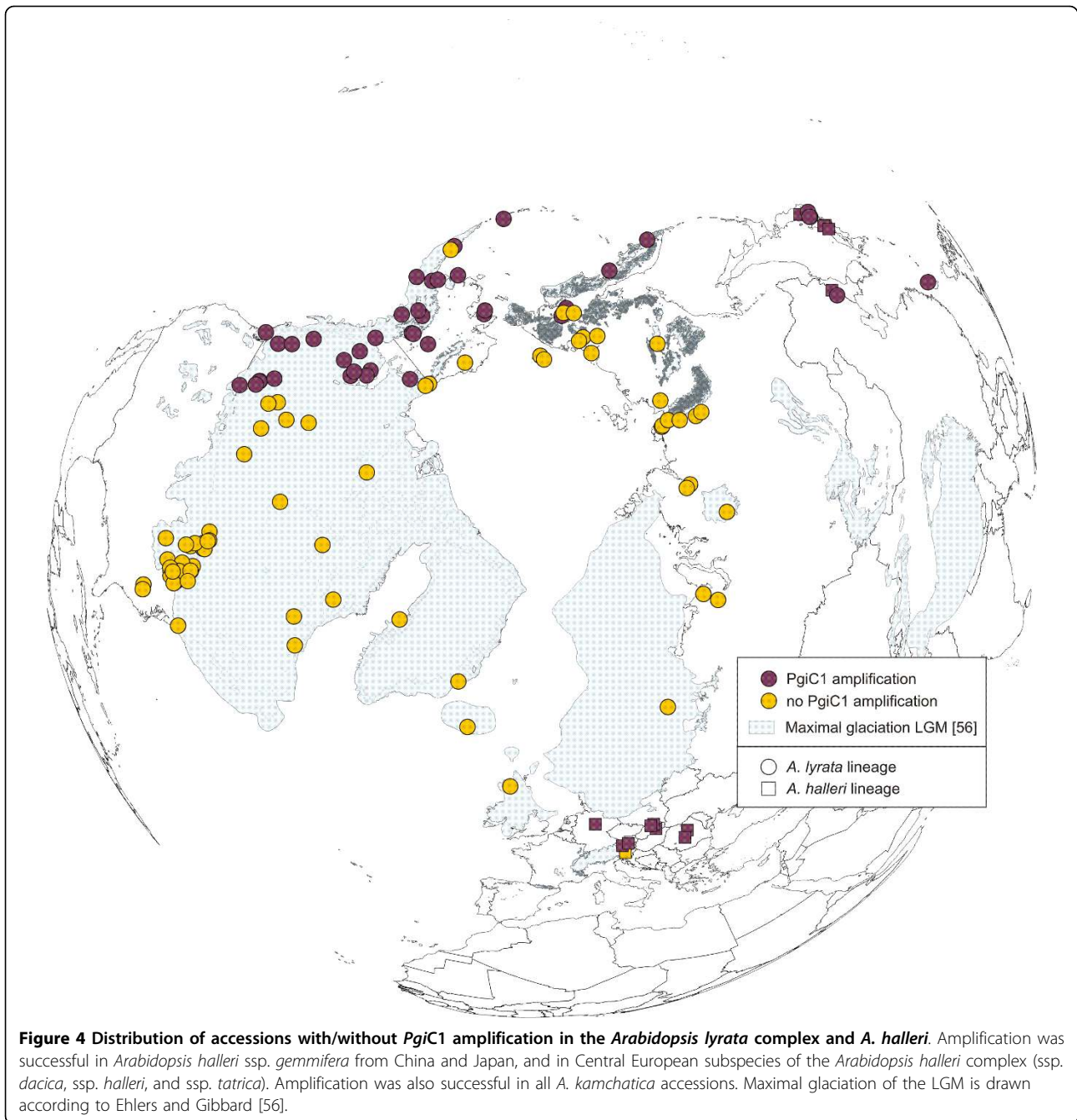


Figure 3 CpDNA *trnL/F* suprahaplotype network of the *Arabidopsis lyrata* complex. Data newly presented in this study were combined with previous results from Koch and Matschinger [27] and Schmickl et al. [28]. *TrnL/F* suprahaplotypes were characterised as *trnL* intron and *trnL-F*-IGS excluding the pseudogene-rich region in the *trnL-F*-IGS. The sizes of the circles indicate the relative frequency of a suprahaplotype in the whole dataset.



Far Eastern Federal District of Russia, and Alaska/western Canada. The Eurasian lineage of *A. lyrata* ssp. *petraea* showed the highest cpDNA-based (*trnL/F*) nucleotide diversity (π), and it was equally high in unglaciated Central Europe ($\pi = 0.0060$), glaciated northern Europe ($\pi = 0.0076$), and northern Russia/western Beringia ($\pi = 0.0077$), but reduced in eastern Beringia ($\pi = 0.0034$) (Table 1). Pairwise Φ_{ST} and F_{ST} calculations showed only slight differentiation among unglaciated Central Europe, glaciated northern Europe,

and northern Russia/western Beringia. For eastern Beringia the data were not statistically significant ($P < 0.05$; Table 2). These observations could indicate long-term survival of the *A. lyrata* complex in unglaciated Central Europe and northern Russia/western Beringia, or post-glacial colonisation of formerly glaciated northern Europe from those two regions. The generally high nucleotide diversity in the Eurasian lineage was caused by a high number of unique and rare suprahaplotypes (Figure 2, Additional file 6, Table S3). In Central Europe

Table 1 Taxonomic and regional genetic differentiation based on cpDNA suprahaplotypes.

| Geographic region | n | $\pi \times 10^2$ |
|---|-----|-------------------|
| <i>A. lyrata</i> ssp. <i>petraea</i> : Unglaciaded Central Europe | 196 | 0.603 +/- 0.329 |
| <i>A. lyrata</i> ssp. <i>petraea</i> : Glaciaded N Europe | 21 | 0.764 +/- 0.425 |
| <i>A. lyrata</i> ssp. <i>petraea</i> : N Russia, W Beringia | 33 | 0.774 +/- 0.422 |
| <i>A. lyrata</i> ssp. <i>petraea</i> : E Beringia | 7 | 0.339 +/- 0.236 |
| <i>A. lyrata</i> ssp. <i>lyrata</i> : Unglaciaded N America, glaciaded Great Lakes region | 55 | 0.080 +/- 0.072 |
| <i>A. arenicola</i> : Glaciaded N America, Greenland | 17 | 0.031 +/- 0.042 |
| <i>A. kamchatica</i> : Japan | 9 | 0.000 +/- 0.000 |
| <i>A. kamchatica</i> : Far Eastern Federal District of Russia | 9 | 0.000 +/- 0.000 |
| <i>A. kamchatica</i> : Alaska, W Canada | 38 | 0.007 +/- 0.019 |

Sample size (n) and nucleotide diversity (π) are provided with the corresponding standard deviation.

the highest number of unique *trnL*/F suprahaplotypes (AC, AF, AH, AI, AJ, AK, AL, J, K, Q, R, V) was found in the foothills of the Eastern Austrian Alps, which remained unglaciaded during Pleistocene climate oscillations. A high number of unique and rare suprahaplotypes was also found in formerly glaciaded northern Europe (AB, AP - Scotland, AO - Iceland, G - Faeroe Islands, S - Norway). Subsequent geographic isolation of these populations during the Holocene warming might have caused restriction of suprahaplotypes to single geographic locations.

In contrast to the Eurasian lineage of *A. lyrata* ssp. *petraea*, the North American lineage showed an approximately tenfold reduction in nucleotide diversity and strong differentiation according to pairwise Φ_{ST} and F_{ST} (Table 2). Nucleotide diversity of accessions from predominantly unglaciaded southeastern North America ($\pi = 0.0008$) was higher than from North America and Greenland ($\pi = 0.0003$) (Table 1), which had been under the Laurentide ice sheet during the LGM, possibly indicating genetic bottlenecks with subsequent rapid postglacial immigration.

Extremely reduced nucleotide diversity was reported from amphi-Pacific *A. kamchatica* ($\pi = 0.0000$) (Table 1). When all three genetic groups of *A. kamchatica* were treated separately, only a single *trnL*/F suprahaplotype was found in each group (Japan: AD; China: C; Far Eastern Federal District of Russia/Alaska/western Canada: B) (Additional file 6, Table S3). Additionally, haplotype diversity was low, especially in *A. kamchatica* with suprahaplotype B. Only one haplotype (no. 84) was detected over a vast amphi-Pacific area from Kamchatka to western Canada (Additional file 7, Table S4).

Refugia as areas of secondary contact of formerly allopatric populations: Beringia as an example

Beringia, an arctic region ranging from Lena River in northeast Russia to Mackenzie River in Alaska and from the Arctic Ocean to mountains in southern Siberia and Alaska, is considered the major refugium for arctic taxa

(reviewed by Abbott and Brochmann [64] and DeChain [65]), as it remained ice-free during Pleistocene climate oscillations. If we consider only the Eurasian and North American lineage of the *A. lyrata* complex, two major ITS groups met in Beringia (Figure 5): (1) the mainly Eurasian group carrying ITS supratype b (Figure 6), comprising Europe (with additional ITS supratypes a, c, d), northern Russia and western Beringia, and (2) the North American group carrying ITS supratypes e and, extremely rarely, γ (Figure 6), including eastern and central North America, Greenland, and eastern Beringia (north of Brooks Range). The main contact zone was located in eastern Beringia, where accessions with *trnL*/F suprahaplotype C, characteristic for the Eurasian lineage, showed ITS supratype e, characteristic for the North American lineage. This is most likely due to ancient and/or recent gene flow from populations of the North American lineage into populations of the Eurasian lineage. Throughout the Beringian area ambiguous sites in ITS DNA sequences, caused by multiple ITS copies within a single genome and incomplete concerted evolution [54,28], were mainly found between ITS supratypes b (Eurasian lineage) and e (North American lineage). These results indicate that gene flow between these genetic groups may be counteracting the effects of concerted evolution. Interestingly, the allopolyploid amphi-Pacific lineage (*A. kamchatica*) also showed ITS supratype b like the Eurasian lineage.

Discussion

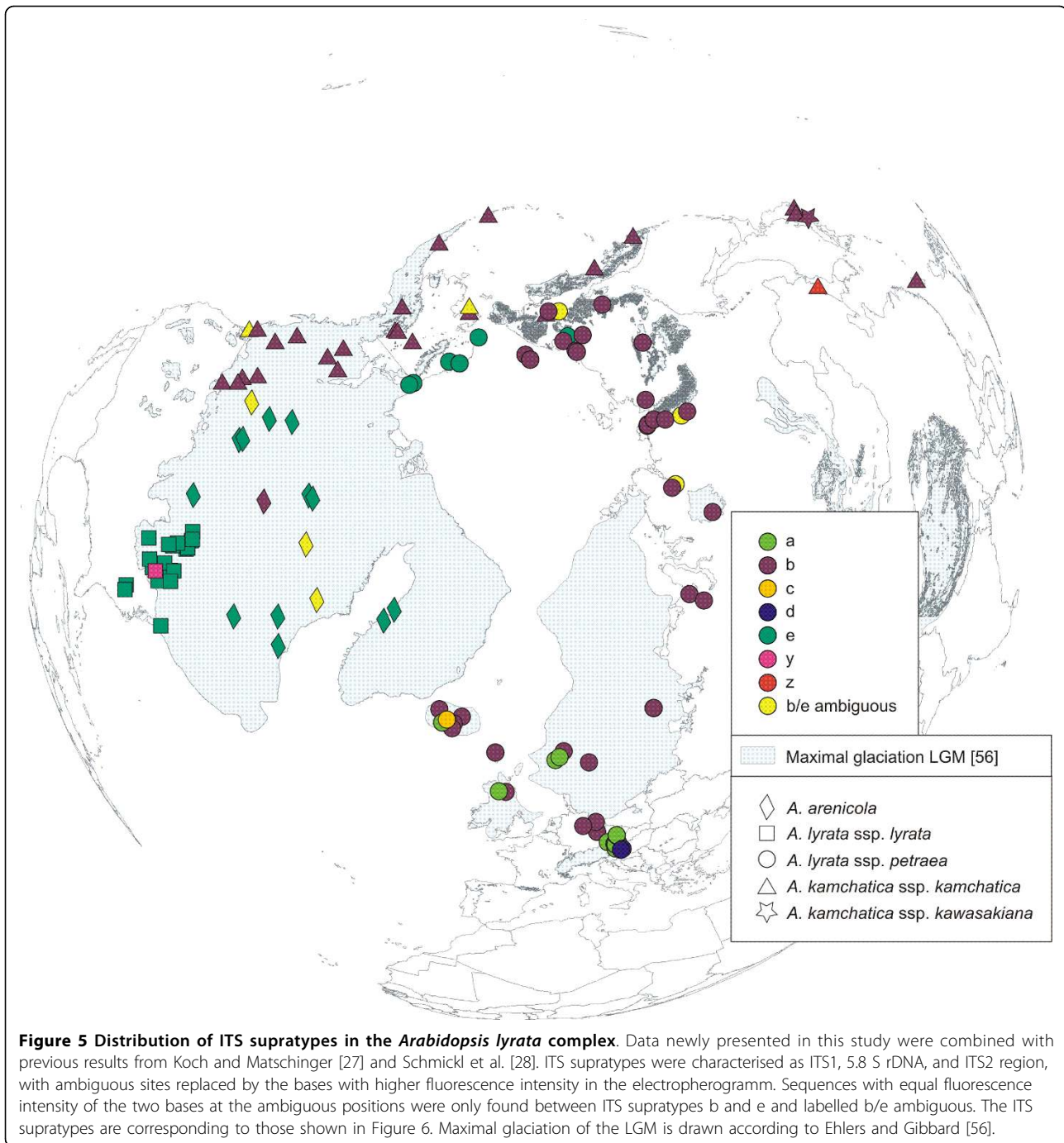
High genetic diversity of the *Arabidopsis lyrata* complex in Eurasia - postglacial migration from Central European and northern Russianrefugia

For arctic-alpine taxa, centres of species and genetic diversity are, in most cases, considered to concur with Pleistocene refugia [64,66,67]. In the genus *Arabidopsis*, both the number of accepted taxa, *trnL*/F suprahaplotype and ITS supratype diversity are highest in Central Europe, indicating this area as a centre of diversity [37,27]. Especially the Eastern Alps could have served as

Table 2 Pairwise genetic differentiation (F_{ST} and Φ_{ST}) among taxonomic and regional groups.

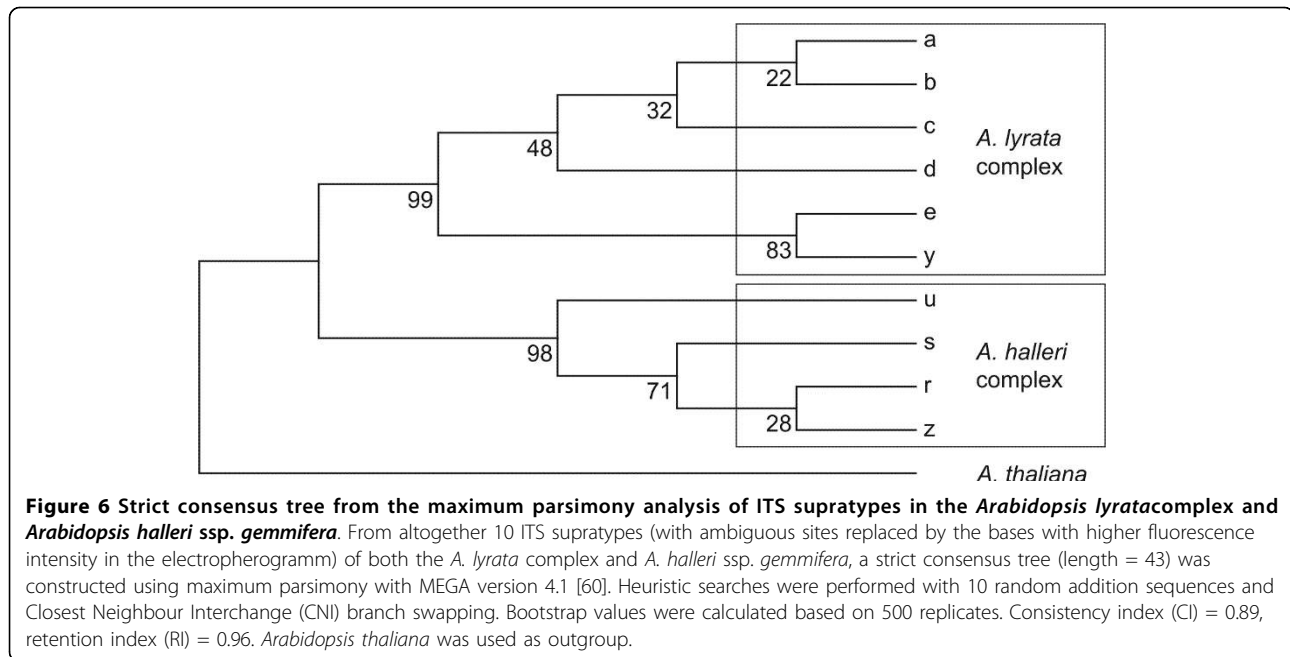
| $\Phi_{ST} \setminus F_{ST}$ | A. lyr. ssp. <i>petraea</i> : Unglac. Central Europe | A. lyr. ssp. <i>petraea</i> : Glac. N Europe | A. lyr. ssp. <i>petraea</i> : N Russia, W Beringia | A. lyr. ssp. <i>petraea</i> : E Beringia | A. lyr. ssp. <i>lyrata</i> : Unglac. N America, glac. Great Lakes region | A. <i>arenicola</i> : Glac. N America, Greenland | A. <i>kamch.</i> : Japan | A. <i>kamch.</i> : Far Eastern Federal District of Russia | A. <i>kamch.</i> : Alaska, W Canada |
|--|--|--|--|--|--|--|--------------------------|---|-------------------------------------|
| A. lyr. ssp. <i>petraea</i> : Unglac. Central Europe | | 0.21832 | 0.17315 | [-0.00029] | 0.47816 | 0.52477 | 0.54797 | 0.53255 | 0.56147 |
| A. lyr. ssp. <i>petraea</i> : Glac. N Europe | 0.14828 | | [0.03253] | 0.17947 | 0.44407 | 0.50004 | 0.48036 | 0.37743 | 0.53503 |
| A. lyr. ssp. <i>petraea</i> : N Russia, W Beringia | 0.19685 | [-0.00620] | | 0.12632 | 0.43290 | 0.48110 | 0.49746 | 0.48437 | 0.60275 |
| A. lyr. ssp. <i>petraea</i> : E Beringia | [-0.00716] | 0.23663 | 0.26978 | | 0.53564 | 0.71126 | 0.76892 | 0.76892 | 0.86207 |
| A. lyr. ssp. <i>lyrata</i> : Unglac. N America, glac. Great Lakes region | 0.37925 | 0.54668 | 0.49762 | 0.79114 | | [0.08456] | 0.70822 | 0.70822 | 0.76248 |
| A. <i>arenicola</i> : Glac. N America, Greenland | 0.33441 | 0.40526 | 0.37345 | 0.77234 | [0.02119] | | 0.92283 | 0.92283 | 0.92688 |
| A. <i>kamch.</i> : Japan | 0.59120 | 0.45048 | 0.40209 | 0.88490 | 0.93046 | 0.97806 | | 1.00000 | 0.95735 |
| A. <i>kamch.</i> : Far Eastern Federal District of Russia | 0.49758 | 0.25161 | 0.19769 | 0.85450 | 0.90536 | 0.96954 | 1.00000 | | [-0.05564] |
| A. <i>kamch.</i> : Alaska, W Canada | 0.54130 | 0.44301 | 0.33922 | 0.94213 | 0.93015 | 0.97891 | 0.97880 | [-0.05564] | |

F_{ST} (above diagonal) and Φ_{ST} (below diagonal) are both estimated from cpDNA suprahaplotypes. Values with $P < 0.05$ are given in brackets (permutation test with 1000 permutations).



a refugium of *A. lyrata* ssp. *petraea*, since both diploid and cytogenetically stabilised tetraploid populations occur there (Schmickl and Koch, unpublished data). For the *A. lyrata* complex, however, Koch and Matschinger [27] concluded that Central Europe is not the only centre of diversity. Our results support this conclusion, showing that the Eurasian lineage of the *A. lyrata* complex is genetically diverse in both Central Europe and arctic Eurasia, including Beringia. To our knowledge

such a pattern has not been observed for any other arctic plant with additional distribution in Central Europe. Either genetic diversity is high in the Arctic compared to the Alps, as in *Saxifraga oppositifolia* L., suggesting long-term evolution in Beringia and more recent colonisation of the Alps [33,31], or genetic diversity is highest in the Alps and decreasing towards the Arctic, as observed in *Arabis alpina* L. [68,69], indicating recent and rapid colonisation from Central Europe. Recent



migration into the Arctic, associated with a loss of genetic diversity due to repeated bottlenecks, was also found in *Ranunculus glacialis* L. [70] and *Rubus chamaemorus* L. [71]. The overall high genetic diversity in the Eurasian *A. lyrata* lineage can probably be explained by long-term glacial survival in multiple refugia (Central Europe, northern Russia, western and eastern Beringia).

So far, periglacial survival of *A. lyrata* was reported only from unglaciated Central Europe, based on microsatellite data [25] and cpDNA sequences [27]. This study showed high genetic diversity also in formerly glaciated northern Europe, caused by numerous unique, locally distributed *trnL/F* suprahaplotypes. A possible explanation might be geographic isolation of populations, which either periglacially survived along the coastline or postglacially migrated into northern Europe, as reported from Ansell [personal communication]. Although several authors suggested periglacial survival on nunataks along the Norwegian coastline, mainly based on geomorphological investigations (reviewed by Brochmann et al. [34]), we assume postglacial colonisation of formerly glaciated northern Europe from both unglaciated Central Europe and northwestern Russia. Subsequent geographic isolation of populations during Holocene warming probably led to the fixation of local sequence types. Although in our study the major *trnL/F* suprahaplotypes C and AG were similarly frequent both in Central and northern Europe, a comparative microsatellite study between populations from Central Europe and Iceland revealed significant differences in marker polymorphism [25]. Riihimäki and Savolainen [9] even

found divergent Central and northern European physiological morphotypes, with earlier and more frequent flowering plants in the south, and later and rarer flowering plants in the north.

We observed high genetic diversity within the Eurasian *A. lyrata* lineage, but also some genetic homogenisation. The *trnL/F* suprahaplotypes C and AG were found all over the distribution range, which could be explained by repeated gene flow between populations during glacial periods, additionally facilitated by rapid long-distance dispersal of the small *Arabidopsis* seeds across the smooth snow surface of the tundra and tundra-steppe by strong winds. This mode of long-distance dispersal could have bridged distances up to 2000 km [72] or even 4000 km (Westergaard et al., in prep.) and seems to be frequent in the Arctic [34,73].

Ancient split of the Eurasian and North American lineage

The strong genetic differentiation we observed between the Eurasian and North American lineages is probably the result of long-term geographic isolation during Pleistocene glaciations. Muller et al. [26] detected genetic divergence between North American and western European populations in a comparative microsatellite study, and found that the former had a lower diversity. However, they focused on a few populations only and included only one North American population. The strong reduction of genetic diversity we observed in the North American lineage compared to the Eurasian lineage is congruent with nuclear and plastidic marker data of Wright et al. [23], but contradicting Balaña-Alcaide et al. [24], who reported

similar genetic diversity based on two nuclear markers. However, both studies included only two populations from each of North America and Europe, and they may not have been representative. Because of the strongly reduced genetic diversity of the North American lineage, we assume this lineage to be derived from the Eurasian lineage. Future studies of more southern populations from the USA will test this assumption. Migration into North America apparently was associated with a strong genetic bottleneck, which restricted populations to ITS supratype e and *trnL/F* suprahaplotype A. Colonisation of North America is possible from different directions, either from Europe, or from East Asia. Both possibilities have been discussed for various circumboreal species (reviewed by Abbott and Brochmann [64]). In the case of *A. lyrata*, colonisation of North America from Russia seems, however, most likely, as the North American genotype was also rarely detected in western Beringia, but not at all in Iceland or Scandinavia.

An amphi-Beringian *Arabidopsis* hybrid zone - due to allopolyploid success?

The majority of arctic polyploids have a history as post-glacial colonisers [74]. It has frequently been assumed that polyploids have a broader adaptive potential for recolonising formerly glaciated areas [75,76]. As one of the major refugia for arctic plants, Beringia has a higher proportion of diploids compared to areas glaciated during Pleistocene [74]. New adaptations in polyploids may evolve either by genome rearrangements [77] and/or epigenetic changes [78-80] within the first generations after polyploid formation, as reported from rapid gene silencing in the allopolyploid *A. suecica* (Fr.) Norrl. ex O.E. Schulz [78]. Hybridisation is frequently involved in polyploidisation, leading to the formation of allopolyploids with one each of the parental genomes. Allopolyploidisations have been reported for several arctic species complexes, such as the high polyploid *Cerastium alpinum* L. complex [81], high polyploid *Primula* sect. *Aleuritia* [82-84], tetra-/hexaploid *Silene* L. [85,86], tetra- to octoploid *Saxifraga* section *Mesogyne* Sternb. [87,88], and tetra- to dodecaploid *Cardamine digitata* Richardson [89]. Introgression, the integration of genetic material from one species into another through repeated backcrossing, was observed between polyploid *Saxifraga cernua* L. and diploid *Saxifraga sibirica* L. [90]. The first known *Arabidopsis* allopolyploid was *A. suecica* with the maternal parent *A. thaliana* and the paternal parent *A. arenosa*, confirmed by artificial crosses [78]. This species probably developed around 20 000 years ago [91] or between 20 000 and 300 000 years ago [92], with a single origin in Fennoscandia [91,92]. The distribution range of this mainly outcrossing [93] allopolyploid species is rather small. The second natural *Arabidopsis*

allopolyploid, *A. kamchatica*, has *A. halleri* ssp. *gemmifera* and a member of the *A. lyrata* complex as parental taxa [[43,27,8]; Jørgensen, unpublished data]. This allopolyploid origin could be confirmed not only for Japanese, but for all *A. kamchatica* accessions. According to chloroplast *trnL/F* data, three different genetic groups were found, which are geographically isolated from each other: (1) accessions of a widespread distribution range from Kamchatka, western and eastern Beringia, to pacific western Canada (*trnL/F* suprahaplotype B), (2) Japanese accessions (*trnL/F* suprahaplotype AD), and (3) accessions from pacific eastern China (*trnL/F* suprahaplotype C). In a more detailed comparison of chloroplast *trnL/F* and nuclear encoded ITS sequence data, two directions of gene flow could be observed: Either the paternal genome originated from a member of the *A. lyrata* species complex (ITS supratype b) and the maternal genome from *A. halleri* ssp. *gemmifera* (*trnL/F* suprahaplotype AD), as already reported for *A. kamchatica* from Japan [27]. Or *A. halleri* ssp. *gemmifera* represented the paternal genome (ITS supratype z, derived from ITS supratype r exclusively found in *A. halleri* ssp. *gemmifera*), and a member of the *A. lyrata* species complex served as donor of the maternal genome (*trnL/F* suprahaplotype C), as in *A. kamchatica* from China. The North American lineage of the *A. lyrata* complex could be excluded as a parent, as neither ITS supratype e nor *trnL/F* suprahaplotype A were detected in *A. kamchatica*. According to these data we suggest at least three independent origins of *A. kamchatica*, first with maternal *A. halleri* ssp. *gemmifera* in Japan, second with paternal *A. halleri* ssp. *gemmifera* in China, and third with an unknown direction of gene flow in Kamchatka, but in all cases with a member of the Eurasian *A. lyrata* lineage as hybridisation partner.

The most profound change in *A. kamchatica* in contrast to its parental species is the switch from outcrossing (with sporophytic self-incompatibility system) to selfing [ssp. *kamchatica*: [15,8]; ssp. *kawasakiana*: [94]]. Selfing increases the possibility of rapid range expansion, as a population can arise from a single individual independent of pollinators and pollen donors. Such a switch is already well known from *A. thaliana*, dated from around 413 000 [95] to one million years ago [96], and is therefore not necessarily correlated with hybridisation and polyploidisation. Change of mating systems is one of the major driving forces for speciation, initiating reproductive isolation of populations [97,98]. However, it is still unclear if there is a correlation between mating systems and hybridisation and/or polyploidisation. The breakdown of self-incompatibility in an artificial cross between *A. thaliana* and *A. lyrata* [99] could indicate a correlation between hybridisation and a switch in mating system, and it could further indicate possible dominance of the selfer *A. thaliana* over *A. lyrata*.

Otherwise, breakdown of the sporophytic self-incompatibility system has been reported mainly from diploid individuals [16,100].

In contrast to *A. suecica*, allopolyploid *A. kamchatica* has a vast distribution range spanning the whole amphiberian region. Our data suggest that this large distribution is partly due to postglacial colonisation of formerly glaciated areas in eastern Beringia and western Canada. The extremely reduced genetic diversity, particularly of the group with suprahaplotype B, suggests that postglacial immigration may have been rapid, possibly facilitated by de novo adaptations as a result of hybridisation and polyploidisation. Moreover, the success of *A. kamchatica* as a rapid coloniser may have been enhanced by the availability of large, open landscapes, where habitats were frequently disturbed by glacial and/or permafrost activity. However, the change in mating systems may have had a strong impact on the success of *A. kamchatica* as a postglacial coloniser, and also on the establishment of the genetic barrier between *A. kamchatica* and the Eurasian and North American lineages.

Beringia as contact zone between the Eurasian and North American lineage of the *Arabidopsis lyrata* complex

Beringia served as a glacial refugium for numerous arctic plant taxa such as *Dryas integrifolia* Vahl. [32], *Saxifraga hirculus* L. [36], *Saxifraga oppositifolia* [33,31], and *Vaccinium uliginosum* L. [35,101]. Periglacial survival in Beringia can also be assumed for the *A. lyrata* complex, in particular, for the Eurasian lineage in arctic western and eastern Beringia north of Brooks Range. The nuclear sequence data indicate that gene exchange with populations of the North American lineage (*A. lyrata* ssp. *lyrata*, *A. arenicola*) occurred inter- and postglacially. However, the data do not support glacial survival of the North American lineage in Beringia, as no plastidic sequence types of this lineage were found in Beringia.

Conclusions

By presenting a worldwide evolutionary history of the *Arabidopsis lyrata* species complex, we provide solid knowledge about centres of genetic diversity, different genetic lineages, their contact zones, and hybrid speciation. We could clearly differentiate three genetic lineages, a Eurasian, a North American, and an amphipacific one. The latter is constituted of the allopolyploid *A. kamchatica*, a hybrid between *A. lyrata* and *A. halleri*. Further investigations of the population dynamics and the role of selfing within this hybrid species should be conducted to gain a deeper understanding of hybrid establishment in the wild.

Additional file 1: Table S1. The two main taxonomic concepts of the *Arabidopsis lyrata* complex. Summarised from Al-Shehbaz and O'Kane [37], including revision from the Flora of North America [Al-Shehbaz, personal communication], and Elven [38]. *Arabidopsis arenicola* was integrated from Warwick et al. [42]. Elven [38] excluded *Arabidopsis lyrata* from taxonomic treatment, as they assumed it to be a non-arctic, boreal taxon.

Additional file 2: Table S2. Within this list all information about taxonomic unit, name on herbarium sheet, herbarium, herbarium number, locality, latitude/longitude, collector, collection date, accession number, ITS type, ITS GenBank number, ITS supertype, *trnL* intron type, *trnL* GenBank number, *trnL/F*-IGS type, *trnL/F*-IGS GenBank number, *trnL* intron + *trnL/F*-IGS type, *trnL/F* suprahaplotype, and *PgiC1* amplification is provided.

Additional file 3: Figure S1. Single most parsimonious tree (length = 553) with bootstrap/jackknife values above 95, based on 37 *Arabidopsis* nuclear DNA *PgiC* sequences. Heuristic searches were performed with 100 random addition sequences and TBR branch swapping, saving three trees per replicate, in TNT [102]. Gaps were treated as fifth state. Consistency index (CI) = 0.69, retention index (RI) = 0.95. Investigated accessions were from the *A. lyrata* complex (ssp. *lyrata*, ssp. *petraea*, *A. septentrionalis*, and *A. umbrosa*) and the *A. halleri* complex (ssp. *gemmaifera*). Taxa with successful amplification of the chosen *PgiC1* fragment, and, consequently, without the deletion in the forward primer site, are marked with the blue box.

Additional file 4: Figure S2. Alignment of the duplicated phosphoglucosomerase loci *PgiC1* and *PgiC2* with the primer binding sites indicated. The forward primer is partly located within the 7 bp deletion between positions 1389 and 1396.

Additional file 5: Figure S3. Selected PCR reactions from the *PgiC* screening: No *PgiC1* amplification in members of the *Arabidopsis lyrata* complex (*A. lyrata* ssp. *lyrata*, *A. arenicola*, *Arabidopsis umbrosa*, and *A. septentrionalis*). Successful *PgiC1* amplification in members of *A. halleri* (ssp. *halleri*, ssp. *dacica*, ssp. *tatica*, and ssp. *gemmaifera*), and *A. kamchatica*.

Additional file 6: Table S3. Taxonomic and regional genetic differentiation based on cpDNA suprahaplotypes. Numbers of cpDNA suprahaplotypes occurring in each region are provided.

Additional file 7: Table S4. List of ITS supertypes, ITS types, *trnL/F* suprahaplotypes, and *trnL/F* haplotypes in the *Arabidopsis lyrata* complex with their corresponding frequencies of occurrence throughout the dataset (*italic*).

Acknowledgements

We thank Ishan Al-Shehbaz (Missouri, USA), Galina Gusarova (Oslo, Norway), Gu Hongya (Beijing, P. R. China), Barbara Mable (Glasgow, Scotland), David L. Remington (Greensboro, USA), Outi Savolainen (Oulu, Finland), and the curators of the Herbariums of the Natural History Museums London and Vienna for providing plant material, Susanne Ball and Stefan Wötzel for laboratory assistance, and Jürgen Ehlers and Phil Gibbard for shapefiles of the LGM. This research was supported by DFG grant KO 2302/5-2 to Marcus A. Koch.

Author details

¹Heidelberg University, Heidelberg Institute of Plant Sciences, Biodiversity and Plant Systematics, Im Neuenheimer Feld 345, D-69120 Heidelberg, Germany. ²Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, PO Box 1066 Blindern, NO-0316 Oslo, Norway.

Authors' contributions

RS carried out the molecular marker studies and statistical analyses, constructed the maps and drafted the manuscript. MHJ constructed the *PgiC* sequence alignment and helped to draft the manuscript. AKB contributed to draft the manuscript. MAK designed and coordinated the

project and drafted the manuscript. All authors read and approved the final manuscript.

Received: 27 November 2009 Accepted: 8 April 2010
Published: 8 April 2010

References

- Mitchell-Olds T: *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol Evol* 2001, **16**:693-700.
- Benfey PN, Mitchell-Olds T: From genotype to phenotype: systems biology meets natural variation. *Science* 2008, **320**:495-497.
- Koch MA, Haubold B, Mitchell-Olds T: Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol* 2000, **17**:1483-1498.
- Koch MA, Haubold B, Mitchell-Olds T: Molecular systematics of the Brassicaceae: evidence from coding plastidic *MATK* and nuclear *CHS* sequences. *Am J Bot* 2001, **88**:534-544.
- Turner TL, Von Wettberg EJ, Nuzhdin SV: Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. *PLoS One* 2008, **3**:e3183.
- Clauss MJ, Koch MA: Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci* 2006, **11**:449-459.
- Zhukova PG, Petrovsky VV, Plieva TV: The chromosome numbers and taxonomy of some plant species from Siberia and Far East. *Botanicheskiy Zhurnal* 1973, **58**:1331-1342.
- Shimizu-Inatsugi R, Lihová J, Iwanaga H, Kudoh H, Marhold K, Savolainen O, Watanabe K, Yakubov VV, Shimizu KK: The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Mol Ecol* 2009, **18**:4024-4048.
- Riihimäki M, Savolainen O: Environmental and genetic effects on flowering differences between northern and southern populations of *Arabidopsis lyrata* (Brassicaceae). *Am J Bot* 2004, **91**:1036-1045.
- Riihimäki M, Podolsky R, Kuitinen H, Koelewijn H, Savolainen O: Studying genetics of adaptive variation in model organisms: flowering time variation in *Arabidopsis lyrata*. *Genetica* 2005, **123**:63-74.
- Clauss MJ, Diemel S, Schubert G, Mitchell-Olds T: Glucosinolate and trichome defenses in a natural *Arabidopsis lyrata* population. *J Chem Ecol* 2006, **32**:2351-2373.
- Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME: Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* 2001, **13**:627-643.
- Schierup MH, Mable BK, Awadalla P, Charlesworth D: Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics* 2001, **158**:387-399.
- Charlesworth D, Mable BK, Schierup MH, Bartolomé C, Awadalla P: Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics* 2003, **164**:1519-1535.
- Mable BK, Beland J, di Berardo C: Inheritance and dominance of self-incompatibility alleles in polyploid *Arabidopsis lyrata*. *Heredity* 2004, **93**:476-486.
- Mable BK, Robertson AV, Dart S, diBerardo C, Witham L: Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* 2005, **59**:1437-1448.
- Charlesworth D, Kamau E, Hagenblad J, Tang C: Trans-specificity at loci near the self-incompatibility loci in *Arabidopsis*. *Genetics* 2006, **172**:2699-2704.
- Hagenblad J, Bechsgaard J, Charlesworth D: Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. *Genetics* 2006, **173**:1057-1073.
- Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB: Selection at work in self-incompatible *Arabidopsis lyrata*: mating patterns in a natural population. *Genetics* 2006, **172**:477-484.
- Schierup MH, Bechsgaard JS, Christiansen FB: Selection at work in self-incompatible *Arabidopsis lyrata*. II. Spatial distribution of S haplotypes in Iceland. *Genetics* 2008, **180**:1051-1059.
- Sharbel TF, Haubold B, Mitchell-Olds T: Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol Ecol* 2000, **9**:2109-2118.
- Beck JB, Schmuths H, Schaal BA: Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol Ecol* 2008, **17**:902-915.
- Wright SI, Lauga B, Charlesworth D: Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Mol Ecol* 2003, **12**:1247-1263.
- Balaña-Alcaide D, Ramos-Onsins SE, Boone Q, Aguadé M: Highly structured nucleotide variation within and among *Arabidopsis lyrata* populations at the FAH1 and DFR gene regions. *Mol Ecol* 2006, **15**:2059-2068.
- Clauss MJ, Mitchell-Olds T: Population genetic structure of *Arabidopsis lyrata* in Europe. *Mol Ecol* 2006, **15**:2753-2766.
- Muller MH, Leppälä J, Savolainen O: Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity* 2008, **100**:47-58.
- Koch MA, Matschinger M: Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 2007, **104**:6272-6277.
- Schmickl R, Jørgensen MH, Brysting AK, Koch MA: Phylogeographic implications for the North American boreal-arctic *Arabidopsis lyrata* complex. *Plant Ecol Div* 2008, **1**:245-254.
- Avice JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC: Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* 1987, **18**:489-522.
- Eidesen PB, Carlsen T, Molau U, Brochmann C: Repeatedly out of Beringia: *Cassiope tetragona* embraces the Arctic. *J Biogeogr* 2007, **34**:1559-1574.
- Abott RJ, Comes HP: Evolution in the Arctic: a phylogeographic analysis of the circumpolar plant, *Saxifraga oppositifolia* (Purple saxifrage). *New Phytol* 2004, **161**:211-224.
- Tremblay NO, Schoen DJ: Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Mol Ecol* 1999, **8**:1187-1198.
- Abbott RJ, Smith LC, Milne RI, Crawford RM, Wolff K, Balfour J: Molecular analysis of plant migration and refugia in the Arctic. *Science* 2000, **289**:1343-1346.
- Brochmann C, Gabrielsen TM, Nordal I, Landvik JY, Elven R: Glacial survival or tabula rasa? The history of North Atlantic biota revisited. *Taxon* 2003, **52**:417-450.
- Alsos IG, Engelskjøn T, Gielly L, Taberlet P, Brochmann C: Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. *Mol Ecol* 2005, **14**:2739-2753.
- Oliver C, Hollingsworth PM, Gornall RJ: Chloroplast DNA phylogeography of the arctic-montane species *Saxifraga hirculus* (Saxifragaceae). *Heredity* 2006, **96**:222-231.
- Al-Shehbaz IA, O'Kane SL Jr: Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). *The Arabidopsis Book* Rockville, MD: American Society of Plant Biologists 2002.
- Elven R, (Ed): Checklist of the Panarctic Flora (PAF) Vascular Plants. Version 2007 [http://www.binran.ru/infsys/paflist/index.htm].
- O'Kane SL Jr, Al-Shehbaz IA: A synopsis of *Arabidopsis* (Brassicaceae). *Novon* 1997, **7**:323-327.
- Al-Shehbaz IA, O'Kane SL Jr, Price RA: Generic placement of species excluded from *Arabidopsis* (Brassicaceae). *Novon* 1999, **9**:296-307.
- Koch MA, Wernisch M, Schmickl R: *Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution. *Taxon* 2008, **57**:933-943.
- Warwick SI, Al-Shehbaz IA, Sauder CA: Phylogenetic position of *Arabis arenicola* and generic limits of *Aphragmus* and *Eutrema* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Can J Bot* 2006, **84**:269-281.
- Shimizu KK, Fujii S, Marhold K, Watanabe K, Kudoh H: *Arabidopsis kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh and *A. kamchatica* subsp. *kawasakiana* (Makino) K. Shimizu & Kudoh, new combinations. *Acta Phytotax Geobot* 2005, **56**:163-172.
- Doyle JJ, Doyle JL: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 1987, **19**:11-15.
- Dobeš CH, Mitchell-Olds T, Koch MA: Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A. × divaricarpa*, and *A. holboellii* (Brassicaceae). *Mol Ecol* 2004, **13**:349-370.
- Dobeš C, Mitchell-Olds T, Koch MA: Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera × divaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and

- chloroplast molecular markers - an integrative approach. *Am J Bot* 2004, **91**:2087-2101.
47. Koch MA, Dobeš C, Matschinger M, Bleeker W, Vogel J, Kiefer M, Mitchell-Olds T: Evolution of the *trn F*(GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. *Mol Biol Evol* 2005, **22**:1032-1043.
48. Dobeš C, Kiefer C, Kiefer M, Koch MA: Plastidic *trnF* UUC pseudogenes in North American genus *Boechera* (Brassicaceae): mechanistic aspects of evolution. *Plant Biol* 2007, **9**:502-515.
49. Koch MA, Dobeš C, Kiefer C, Schmickl R, Klimeš L, Lysak MA: Supernetwork identifies multiple events of plastid *trn F*(GAA) pseudogene evolution in the Brassicaceae. *Mol Biol Evol* 2007, **24**:63-73.
50. Schmickl R, Kiefer C, Dobeš C, Koch MA: Evolution of *trn F*(GAA) pseudogenes in cruciferous plants. *Plant Syst Evol* 2009, **282**:229-240.
51. Ansell SW, Schneider H, Pedersen N, Grundmann M, Russell SJ, Vogel JC: Recombination diversifies chloroplast *trnF* pseudogenes in *Arabidopsis lyrata*. *J Evol Biol* 2007, **20**:2400-2411.
52. Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E: Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proc Natl Acad Sci USA* 1980, **77**:7323-7327.
53. Liao D: Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 1999, **64**:24-30.
54. Koch MA, Dobeš C, Mitchell-Olds T: Multiple hybrid formation in natural populations: Concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). *Mol Biol Evol* 2003, **20**:338-350.
55. Carson AR, Scherer SW: Identifying concerted evolution and gene conversion in mammalian gene pairs lasting over 100 million years. *BMC Evol Biol* 2009, **9**:156.
56. Ehlers J, Gibbard PL, (Eds): *Quaternary glaciations - extent and chronology* Amsterdam: Elsevier 2004.
57. Clement M, Posada D, Crandall KA: TCS: a computer program to estimate gene genealogies. *Mol Ecol* 2000, **9**:1657-1659.
58. Templeton AR, Crandall KA, Sing CF: A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 1992, **132**:619-633.
59. Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005, **1**:47-50.
60. Kumar S, Nei M, Dudley J, Tamura K: MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 2008, **9**:299-306.
61. Gottlieb LD: Evidence for duplication and divergence of the structural gene for phosphoglucosyltransferase in diploid species of *Clarkia*. *Genetics* 1977, **86**:289-308.
62. Ghatnekar L: A polymorphic duplicated locus for cytosolic PGI segregating in sheep's fescue (*Festuca ovina* L.). *Heredity* 1999, **83**:451-459.
63. Ford VS, Gottlieb LD: Single mutations silence PgiC2 genes in two very recent allotetraploid species of *Clarkia*. *Evolution* 2002, **56**:699-707.
64. Abbott RT, Brochmann C: History and evolution of the arctic flora: in the footsteps of Eric Hulten. *Mol Ecol* 2003, **12**:299-313.
65. DeChain EG: A bridge or a barrier? Beringia's influence on the distribution and diversity of tundra plants. *Plant Ecol Div* 2008, **1**:197-207.
66. Tribsch A, Schönswetter P: Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon* 2003, **52**:477-497.
67. Hewitt GM: The structure of biodiversity - insights from molecular phylogeography. *Front Zool* 2004, **1**:4.
68. Koch MA, Kiefer C, Ehrich D, Vogel J, Brochmann C, Mummenhoff K: Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Mol Ecol* 2006, **15**:825-839.
69. Ehrich D, Gaudeul M, Assefa A, Koch MA, Mummenhoff K, Nemomissa S, Brochmann C: Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Mol Ecol* 2007, **16**:2542-2559.
70. Schönswetter P, Paun O, Tribsch A, Niklfeld H: Out of the Alps: colonization of northern Europe by east alpine populations of the glacier buttercup *Ranunculus glacialis* L. (Ranunculaceae). *Mol Ecol* 2003, **12**:3373-3381.
71. Ehrich D, Alsos IG, Brochmann C: Where did the northern peatland species survive the dry glacials: cloudberry (*Rubus chamaemorus*) as an example. *J Biogeogr* 2008, **35**:801-814.
72. Savile DBO: Arctic adaptations in plants. *Canada Department of Agriculture Monograph* 1972, **6**:1-81.
73. Alsos IG, Eidesen PB, Ehrich D, Skrede I, Westergaard K, Jacobsen GH, Landvik JY, Taberlet P, Brochmann C: Frequent long-distance plant colonization in the changing Arctic. *Science* 2007, **316**:1606-1609.
74. Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen AC, Elven R: Polyploidy in arctic plants. *Biol J Linn Soc* 2004, **82**:521-536.
75. Manton I: The problem of *Biscutella laevigata* L. - II. The evidence from meiosis. *Ann Bot* 1937, **1**:439-462.
76. Stebbins GL: Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. *Bot Helv* 1984, **94**:1-13.
77. Soltis PS, Soltis DE: The role of genetic and genomic attributes in the success of polyploids. *Proc Natl Acad Sci USA* 2000, **97**:7051-7057.
78. Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Elvers B: Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 2000, **12**:1551-1568.
79. Wendel JF: Genome evolution in polyploids. *Plant Mol Biol* 2000, **42**:225-249.
80. Liu B, Wendel JF: Epigenetic phenomena and the evolution of plant allopolyploids. *Mol Phylogenet Evol* 2003, **29**:365-379.
81. Brysting AK, Oxelman B, Huber KT, Moulton V, Brochmann C: Untangling complex histories of genome mergings in high polyploids. *Syst Biol* 2007, **56**:467-476.
82. Guggisberg A, Mansion G, Kelso S, Conti E: Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex of *Primula*. *New Phytol* 2006, **171**:617-632.
83. Guggisberg A, Baroux C, Grossniklaus U, Conti E: Genomic origin and organization of the allopolyploid *Primula egalikensis* investigated by in situ hybridization. *Ann Bot* 2008, **101**:919-927.
84. Guggisberg A, Mansion G, Conti E: Disentangling reticulate evolution in an arctic-alpine polyploid complex. *Syst Biol* 2009, **58**:55-73.
85. Popp M, Erixon P, Eggens F, Oxelman B: Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. *Syst Bot* 2005, **30**:302-313.
86. Popp M, Oxelman B: Origin and evolution of North American polyploid *Silene* (Caryophyllaceae). *Am J Bot* 2007, **94**:330-349.
87. Brochmann C, Xiang QY, Brunsfeld SJ, Soltis DE, Soltis PS: Molecular evidence for polyploid origins in *Saxifraga* (Saxifragaceae): the narrow arctic endemic *S. svalbardensis* and its widespread allies. *Am J Bot* 1998, **85**:135-143.
88. Jørgensen MH, Elven R, Tribsch A, Gabrielsen TM, Stedje B, Brochmann C: Taxonomy and evolutionary relationships in the *Saxifraga rivularis* complex. *Syst Bot* 2006, **31**:702-729.
89. Jørgensen MH, Carlsen T, Skrede I, Elven R: Microsatellites resolve the taxonomy of the polyploid *Cardamine digitata* aggregate (Brassicaceae). *Taxon* 2008, **57**:882-892.
90. Kapralov MV, Gabrielsen TM, Sarapultsev IE, Brochmann C: Genetic enrichment of the arctic clonal plant *Saxifraga cernua* at its southern periphery via the alpine sexual *Saxifraga sibirica*. *Mol Ecol* 2006, **15**:3401-3411.
91. Säll T, Jakobsson M, Lind-Halldén C, Halldén C: Chloroplast DNA indicates a single origin of the allotetraploid *Arabidopsis suecica*. *J Evol Biol* 2003, **16**:1019-1029.
92. Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M: A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. *Mol Biol Evol* 2006, **23**:1217-1231.
93. Säll T, Lind-Halldén C, Jakobsson M, Halldén C: Mode of reproduction in *Arabidopsis suecica*. *Hereditas* 2004, **141**:313-317.
94. Sugisaka J, Kudoh H: Breeding system of the annual Cruciferae, *Arabidopsis kamchatica* subsp. *kawasakiana*. *J Plant Res* 2008, **121**:65-68.
95. Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH: The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. *Mol Biol Evol* 2006, **23**:1741-1750.

96. Tang C, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, Hu TT, Clark RM, Nasrallah JB, Weigel D, Nordborg M: **The evolution of selfing in *Arabidopsis thaliana*.** *Science* 2007, **317**:1070-1072.
97. Schoen DJ, Johnston MO, L'Heureux AM, Marsolais JV: **Evolutionary history of the mating system in *Amsinckia* (Boraginaceae).** *Evolution* 1997, **51**:1090-1099.
98. Rieseberg LH, Willis JH: **Plant speciation.** *Science* 2007, **317**:910-914.
99. Nasrallah JB, Liu P, Sherman-Broyles S, Schmidt R, Nasrallah ME: **Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids.** *Genetics* 2007, **175**:1965-1973.
100. Mable BK: **Genetic causes and consequences of the breakdown of self-incompatibility: case studies in the Brassicaceae.** *Genet Res* 2008, **90**:47-60.
101. Eidesen PB, Alsos IG, Popp M, Stensrud Ø, Suda J, Brochmann C: **Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species.** *Mol Ecol* 2007, **16**:3902-3925.
102. Goloboff PA, Farris JS, Nixon KC: **TNT (Tree analysis using New Technology).** Published by the authors, Tucumán 2000.

doi:10.1186/1471-2148-10-98

Cite this article as: Schmickl *et al.*: The evolutionary history of the *Arabidopsis lyrata* complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evolutionary Biology* 2010 **10**:98.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

