Heredity (2008) 100, 47–58

www.nature.com/hdy

ORIGINAL ARTICLE

Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*

M-H Muller[1,2], J Leppälä[1] and O Savolainen[1]
[1]*Department of Biology, University of Oulu, Oulu, Finland and* [2]*UMR DIA-PC 1097, INRA Montpellier, Domaine de Melgueil, Mauguio, France*

The perennial outcrossing *Arabidopsis lyrata* is becoming a plant model species for molecular ecology and evolution. However, its evolutionary history, and especially the impact of the climatic oscillations of the Pleistocene on its genetic diversity and population structure, is not well known. We analyzed the broad-scale population structure of the species based on microsatellite variation at 22 loci. A wide sample in Europe revealed that glaciations and postglacial colonization have caused high divergence and high variation in variability between populations. Colonization from Central Europe to Iceland and Scandinavia was associated with a strong decrease of genetic diversity from South to North. On the other hand, the Russian population included in our data set may originate from a different refugium probably located more to the East. These genome-wide patterns must be taken into account in studies aiming at elucidating the genetic basis of local adaptation. As shown by sequence data, most of the loci used in this study do not evolve like typical microsatellite loci and show variable levels of homoplasy: this mode of evolution makes these markers less suitable to investigate the between-continent divergence and more generally the worldwide evolution of the species. Finally, a strong negative correlation was detected between levels of within-population diversity and indices of differentiation such as F_{ST}. We discuss the causes of this correlation as well as the potential bias it induces on the quantification and interpretation of population structure.
Heredity (2008) **100,** 47–58; doi:10.1038/sj.hdy.6801057; published online 17 October 2007

Keywords: *Arabidopsis lyrata*; microsatellites; postglacial history; population structure; F_{ST}

Introduction

Arabidopsis lyrata is a close relative of *Arabidopsis thaliana*, the model organism for plant molecular biology, and itself is becoming a model for ecology and evolution (Clauss and Koch, 2006). It benefits from a wealth of molecular tools and information accumulated for its congener, and the genomic resources for *A. lyrata* are also rapidly improving (Kuittinen *et al*., 2002, 2004; http://www.jgi.doe.gov/sequencing/why/CSP2006/AlyrataCrubella.html). In contrast to *A. thaliana*, it is a perennial, outcrossing species and forms stable populations (Clauss and Koch, 2006). As such, it is a more tractable organism for population genetics and allows addressing evolutionary questions and testing the validity of theoretical models (Riihimäki *et al*., 2005). For instance, *A. lyrata*, together with other species of the genus *Arabidopsis*, is intensively used for the study of self-incompatibility evolution (Mable *et al*., 2005; Schierup *et al*., 2006) or the genetic basis of adaptation at different geographic scales (for example, Kärkkäinen *et al*., 2004; Riihimäki *et al*., 2005, Kivimäki *et al*. 2007).

Despite this interest, the evolutionary history of the species is not well known. Follwing three subspecies are distinguished (Al-Shehbaz and O'Kane, 2002): *A. lyrata* ssp. *lyrata* (North America), *A. lyrata* ssp. *petraea* (Europe and Siberia) and *A. lyrata* ssp. *kamchatica* (from Eastern Asia to Northwest America). Within Europe, ssp. *petraea* has a patchy distribution extending from Central Europe to Norway (Jalas and Suominen, 1994; Jonsell *et al*., 1995). It occurs under a variety of climatic and ecological conditions, but is most often cold-tolerant and grows in low competition habitats (Jonsell *et al*., 1995; Clauss and Mitchell-Olds, 2006). Studying the large-scale population structure and history of this species is needed as a prerequisite for the study of potentially selected variation (Wright and Gaut, 2005). It is also interesting because *A. lyrata* constitutes a good example of the evolution of a species complex affected by the climatic oscillations of the Pleistocene (Abbott and Brochmann, 2003).

Indeed, the Pleistocene has been a period of intense and recurrent climatic changes that have affected the distribution and genetic diversity of many species, leading to the evolution of differentiated genetic entities and in some cases to speciation (Hewitt, 2004). Cold periods displaced the distribution ranges southwards and European refugia have been described for a variety of organisms in Spain, Italy and the Balkans (Taberlet *et al*., 1998). During interglacial times, species were able to expand northward; in Europe, different routes of colonization have contributed to the present-day distribution of many different species (Taberlet *et al*., 1998). In addition to these North–South movements, the evolution of species having a 'circumboreal' distribution such as *A. lyrata*, has been affected by East–West

Correspondence: Dr M-H Muller, UMR DIA-PC 1097, INRA, Domaine de Melgueil, Mauguio F34130, France.
E-mail: Marie-Helene.Muller@supagro.inra.fr
Received 28 March 2007; revised 12 June 2007; accepted 3 August 2007; published online 17 October 2007

movements, from Europe to America, across the Atlantic, as well as through Russia and Asia (Abbott and Brochmann, 2003). In plants, the vast majority of studies of glacial and postglacial history have concerned tree species (for example, Heuertz *et al*., 2004), and a few other circumboreal species (Alsos *et al*., 2005).

Previous studies on neutral variation, based on isozymes, microsatellites or nucleotide polymorphism, have revealed a strong differentiation between *A.l.* ssp. *petraea* and *A.l.* ssp. *lyrata*, and—generally but not always—a lower level of diversity in ssp. *lyrata*, potentially due to population bottlenecks (Van Treuren *et al*., 1997; Savolainen *et al*., 2000; Ramos-Onsins *et al*., 2004; Wright *et al*., 2006). Within Europe, genetic diversity appears to be strongly influenced by population history and demography (Kärkkäinen *et al*., 2004; Ramos-Onsins *et al*., 2004). A recent study (Clauss and Mitchell-Olds, 2006) analyzed the genetic diversity of ssp. *petraea* in Central Europe and suggested that the species may have persisted in this region during the last glacial maximum. The relationships of Northern European, especially Scandinavian populations, with this potential refugium, the importance of founder effects during recolonization and the level of differentiation within Europe compared to the between continent differentiation (that is, differentiation with the subspecies ssp. *lyrata*) remain to be studied.

With their high level of diversity, ease of use and codominant nature, microsatellite markers have already proven useful in small-scale analyses of *A. lyrata* (Clauss *et al*., 2002; Clauss and Mitchell-Olds, 2006). In the present study, we surveyed microsatellite variation at 22 loci in a sample including seven European populations of ssp. *petraea* and one American population of ssp. *lyrata*. We asked the following main questions: how has postglacial colonization impacted the patterns of within population diversity, and divergence of populations? Have the populations reached a genetic equilibrium after the colonization? What are the implications of the demographic history for studies of molecular ecology and functional genomics?

We show that populations of *A. lyrata* exhibit highly variable levels of diversity and are strongly differentiated, especially within Europe, as a consequence of the different routes of postglacial colonization. Further, we illustrate how the variable within-population diversity can impact the methods of analyzing population differentiation. Finally, our results can serve as baseline information for the study of functionally important variation.

Materials and methods

Sampling of populations and loci

We included seven European populations of *A. lyrata* ssp. *petraea* (Figure 1): Karhumäki (Russia 62°55′N, 34°25′E), Stubbsand (Sweden 63°13′N, 18°90′E), Spiterstulen (Norway, 61°38′N, 8°24′E), Lom (Norway, 61°50′N, 8°30′E, 20 km from Spiterstulen), Reykjavik (Iceland, 64°09′N, 21°58′W), Plech (Germany 49°39′N, 11°29′E) and Bohemia (Czech Republic 50°03′N, 14°06′E), and one American population of *A. lyrata* ssp. *lyrata*—Mayodan (North Carolina, USA, 36°41′N, 79°97′W). Sample sizes ranged from 11 to 30 per population for a total sample size of 186 individuals.

For all populations except Bohemia, we analyzed field-collected seeds sampled from different families. Bohemia seeds were a population sample propagated in the laboratory, kindly provided by Mark McNair. Reykjavik seeds were kindly provided by MH Schierup and Mayodan seeds by CH Langley.

We used 22 microsatellite loci among which 21 were previously described (Table 1). The primers used for SOC1 were 5′ TGTCAAATGTATTCGAGCAAGA and 5′ TGTAAGAGCAAGCACAAGAGGA. Some of these loci are mapped (Kuittinen *et al*., 2004; J Leppäla, unpublished). Except for ELF3 and SOC1, all loci have been derived from *A. thaliana*.

Laboratory methods

DNA was extracted from about 200 mg of plant leaves using the FastDNA kit and the FastPrep instrument (Qbiogene Inc, Carlsbad, CA, USA). The amplification reaction consisted of 10 ng DNA, 0.25 μM each primer, one of which was fluorescently labelled with TET, HEX or FAM, 1 × reaction buffer (Promega, Madison, WI, USA), 2.5 mM $MgCl_2$, 100 μM dNTP, 0.15 U *Taq* DNA polymerase (Promega), in a total volume of 15 μl. The amplification method was as follows: 94 °C for 3 min, 35 cycles of 94 °C for 20 s, 50 °C for 30 s, and 72 °C for 10s, followed by a final extension for 45 min at 72 °C. Electrophoresis was performed on an ABI Prism 377 sequencer. The GENESCAN 3.1 and GENOTYPER 2.0 (Applied Biosystems, Foster City, CA, USA) were used to analyze the DNA fragments and to score the genotypes. We reported all observed size classes. The final extension time was long to ensure that one T was added to all fragments by the *Taq* polymerase.

As most of the microsatellite loci had been derived from *A. thaliana*, their molecular structure and thus their mode of evolution in *A. lyrata* could differ from that of typical microsatellites. For 18 loci, we thus sequenced one allele from each of the populations Karhumäki, Plech, Spiterstulen and Mayodan. For homozygote genotypes, we directly sequenced the polymerase chain reaction product. For heterozygotes, we cloned the polymerase chain reaction products into TOPO TA PCR 2.1 cloning vector (Invitrogen, Carlsbad, CA, USA) and sequenced one single clone per individual. The sequencing reactions were run in the ABI PRISM 377 sequencer and the contigs were read and aligned using the Staden-package (http://staden.sourceforge.net/). For each locus, we scored whether it was a perfect microsatellite (that is, shows one stretch of multiple repeats only), a compound microsatellite (that is, shows more than one stretch of repeats) or not a microsatellite (that is, does not show any clear repeat), and whether we observed point mutations between alleles, or indels outside the repeat area. Sequences are available under accession numbers AM411873–AM411876 and AM413048–AM413110.

Data analysis

The number of alleles and unbiased gene diversity (Nei, 1987) within populations and on the whole data set were calculated using the program GENETIX (Belkhir *et al*., 2001). The number of different alleles per population was standardized for a sample size of 22 gene copies or 11 diploid individuals using the rarefaction method of Petit *et al*. (1998) implemented in the software CONTRIB; the

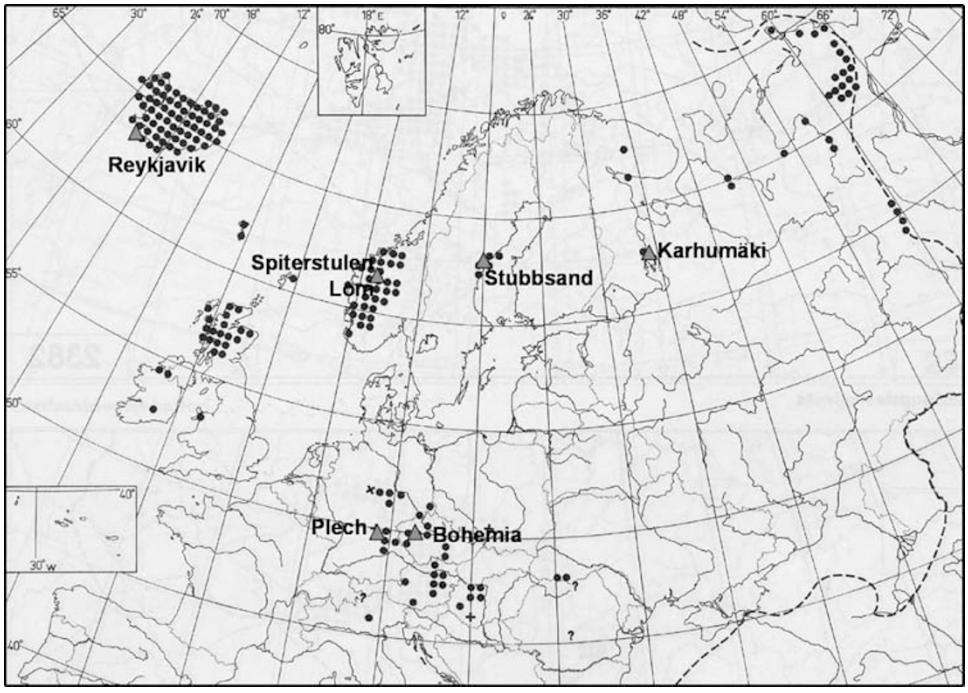

Figure 1 Distribution map of *A. lyrata* ssp. *petraea* in Europe (Jalas and Suominen, 1994). The triangles show the location of the sampled European populations. Lom and Spiterstulen are indicated by the same symbol.

value obtained after rarefaction was denoted allelic richness. The rarefaction method was also applied to the number of private alleles using the software HP-RARE (Kalinowski, 2005) to estimate a private allelic richness. As the different populations showed highly varying levels of allelic richness, and as these differences could by themselves impact the absolute value of the private allelic richness, we also computed the ratio of private allelic richness to allelic richness.

To perform multilocus analyses, we needed to consider subsets of independent loci. Therefore, we analyzed genotypic disequilibrium for all pairs of loci in each population with exact tests using GENEPOP (Raymond and Rousset, 1995), and applied a sequential Bonferroni correction (Rice, 1989). We applied a liberal Bonferroni threshold (equal to 10 times the correct threshold) to be conservative in the conclusion of independence between loci. Then, we used the results of these tests together with the available linkage information to make up all possible subsets of independent loci (20 subsets, see Results). All multilocus analyses described below were made on each subset separately, as well as on the full set of 22 loci.

The significance of differences in genetic diversity parameters between pairs of populations was tested using Wilcoxon signed-rank tests comparing parameters for the same loci in different populations using SAS software (SAS Institute, Cary, NC, USA). Hardy–Weinberg equilibrium was tested for each locus and across loci within each population using exact test implemented in GENEPOP.

Using the software BOTTLENECK (Piry *et al*., 1999), we tested for deviation from mutation-drift equilibrium within each population, using both a sign-test and a Wilcoxon signed-rank test for heterozygosity excess or deficiency (Cornuet and Luikart, 1996). We performed the tests under the IAM (Infinite Allele Model) and SMM (Stepwise Mutation Model) models. Using BOTTLENECK, we also computed the T2 statistic, which is a function of the standardized deviates of the sample heterozygosities relative to the heterozygosity expected given the allelic richness. A positive value of T2 denotes a heterozygosity excess (Cornuet and Luikart, 1996).

Wright's F_{ST} were estimated using the method of Weir and Cockerham (Weir, 1996), using the program GENEPOP. The significance of population differentiation was assessed using exact tests also in GENEPOP. These analyses were performed for individual loci and over loci, in pairwise population comparisons and across all populations.

The value of F_{ST} is strongly dependent on the level of average within population gene diversity (H_S, Hedrick, 1999), and H_S was highly variable among loci. Hedrick (2005) expressed the maximum value that Nei's estimator of F_{ST} (G_{ST}; Nei, 1987) can reach, given the value of H_S, and proposed computing a standardized value of G_{ST} (called G_{ST}') by dividing the estimated value by the maximum value (Hedrick, 2005). As we were using the estimator of F_{ST} of Weir and Cockerham (Weir, 1996), instead of G_{ST}, we manually computed the maximum possible value of this estimator, given the observed levels of intrapopulation diversity. Namely, we modified the numbering of the alleles observed in different populations so that populations were not sharing any alleles. Then we estimated all new F_{ST} values, which now correspond to F_{STmax} values, using GENEPOP on the modified data, and from this, computed standardized, F'_{ST} values.

To depict the relationships between populations, we built neighbor-joining trees (Saitou and Nei, 1987) based on two different matrices of genetic distances, Nei's genetic distance (Nei, 1987) and Cavalli-Sforza's chord distance

Table 1 Description of the loci used in the present study, summary of sequence information and overall level of diversity

Locus	*Ref.*	*LG*	*Position (cM)*	*LG in Arabidopsis thaliana*	*Repeat in A. thaliana*[a]	*Repeat in Arabidopsis lyrata*	*Average number of contiguous repeats*	*Repeat type*	*Other observable sources of polymorphism*	*Number of alleles*	H_T	*Evidence of homoplasy*
AthACS	Clauss *et al*. (2002)	1	0	1	T	A and T	10	Compound	Point mutations	8	0.7909	Yes
AthCDPK9	Clauss *et al*. (2002)	6	38.7	5	TC	TC	6[b]	—	—	5	0.6355	—
AthDET1	Clauss *et al*. (2002)	6	19.6	4	CT	None	1	None	Point mutations, indels	4	0.503	Yes
AthSO191	Clauss *et al*. (2002)	7	62.8	5	ATG	TGA	6	Perfect	None	8	0.5934	No
AthZFPG	Clauss *et al*. (2002)	1	32.3	1	CT	CT	10	Compound	Point mutations	17	0.8219	Yes
ATTSO392	Clauss *et al*. (2002)	1	43.2	1	AAG	AAG	8	Perfect	Indels	15	0.828	No
ELF3	Kuittinen *et al*. (2004)	4	19.7	2	CAA	CAA	16	Compound	Indels	20	0.8674	Yes
F19G10	Clauss *et al*. (2002)	1	30	1	GT	GT	4	Perfect	Point mutations	4	0.4554	Yes
F19K23	Clauss *et al*. (2002)	2	0	1	TTC	TTC	5	Perfect	point mutations, indels	6	0.4596	No
F20D22	Clauss *et al*. (2002)	1	3.3	1	GTTT	CT	7	Perfect	Point mutations, indels	9	0.8069	Yes
ICE12	Clauss *et al*. (2002)	4	56.2	2	CT	TC	7	Compound	Point mutations	6	0.4906	Yes
ICE13	Clauss *et al*. (2002)	1	17.2	1	ATC	CAT	8	Compound	Point mutations, indels	15	0.7787	No
ICE14	Clauss *et al*. (2002)	3*	—	2	GAT	GAT	9	Perfect	Point mutations	6	0.5862	Yes
ICE2	Clauss *et al*. (2002)	NA	—	5	CT	CT	4	Compound	—	5	0.3814	No
ICE3	Clauss *et al*. (2002)	7	77.8	5	CT	—	—	—	—	22	0.8542	—
nga106	Bell and Ecker (1994)	6*	—	5	GA	—	—	—	—	11	0.7915	—
nga112	Clauss *et al*. (2002)	5	54.4	3	CT	CT	6	Compound	Point mutations, indels	14	0.7798	No
nga1139	Ponce *et al*. (1999)	NA	—	4	AG	AG	8	Perfect	Point mutations	11	0.7439	No
nga151	Clauss *et al*. (2002)	6*	—	5	CT	None	1	None	Indels	4	0.5003	Yes
nga248	Bell and Ecker (1994)	NA	—	1	AG	—	—	—	—	10	0.5928	—
nga280	Clauss *et al*. (2002)	1*	—	1	AG	—	—	—	—	5	0.6784	—
SOC1	—	4*	—	2	None	AT	5	Perfect	Point mutations	7	0.7065	Yes

Abbreviations: H_T, total gene diversity; LG, linkage group.
Map position are from Kuittinen *et al*. (2004), except those with * = Leppäla, unpublished.
[a]Clauss *et al*. (2002) or database searches.
[b]Only one sequence obtained, originating from North Carolina.

(Cavalli-Sforza and Edwards, 1967), using Phylip (Felsenstein, 2005). We assessed the confidence of the trees in two ways. First, we built 20 different trees based on the 20 different subsets of loci. Second, we bootstrapped the data (for the full set of loci) a thousand times.

Results

General levels of diversity

Over the whole microsatellite data set, 190 alleles were scored and the average number of alleles detected per locus was 9.63 (range 4–22). The average over loci of total gene diversity (estimated for the whole data set) was 0.67 (range 0.38–0.87). When restricted within Europe, the per-locus statistics were similar (9.23 and 0.65, respectively). The level of polymorphism was highly variable between loci (Table 1).

The observed size classes were rarely concordant with the repeat type described in *A. thaliana* or in *A. lyrata*. Many loci displayed single base-pair differences among alleles, although we would expect 2 or 3 base pair differences. These observed size differences were repeatable when similar samples were amplified or loaded twice, and were confirmed by the sequence information when available (see below).

Sequence variation at microsatellite loci

The general information gathered through the sequence data are summarized in Table 1, and three representative alignments are shown in Figure 2. When it was possible to compare *A. thaliana* and *A. lyrata* sequences, large differences were observed. In addition to base substitutions and indels, the stretch of repeats was often interrupted (compound microsatellite) or absent in *A. lyrata* (Figure 2, locus AthDET1). Moreover, the number of repeats was usually lower in *A. lyrata* than in *thaliana*. The alleles often differed by point mutations or indels outside the main repeat region. Thus, mutation at these loci in *A. lyrata* does not conform to mutation models invoked for microsatellites (for example, Two Phase Model—TPM—or SMM).

From our sequences, we observed that both nucleotide substitutions and indels outside the repeat region created alleles with identical sizes, but differing at the sequence level (for example, homoplasy; Figure 2, AthDET1 and ELF3). Even if not directly observed, this homoplasy could also be deduced from the allele size distribution (F20D22).

Linkage disequilibrium

Among 1208 tests for genotypic disequilibrium between pairs of loci, 80 (6.6%). were significant at the level of $P<0.05$, 45 of which concerning population Bohemia. Overall, only one test was significant when applying Bonferroni correction, and seven tests when applying a more liberal Bonferroni threshold (Table 2). These significant tests involved different pairs of loci and only two involved loci that are genetically linked in *A. lyrata* (Table 2). Among loci for which linkage group was

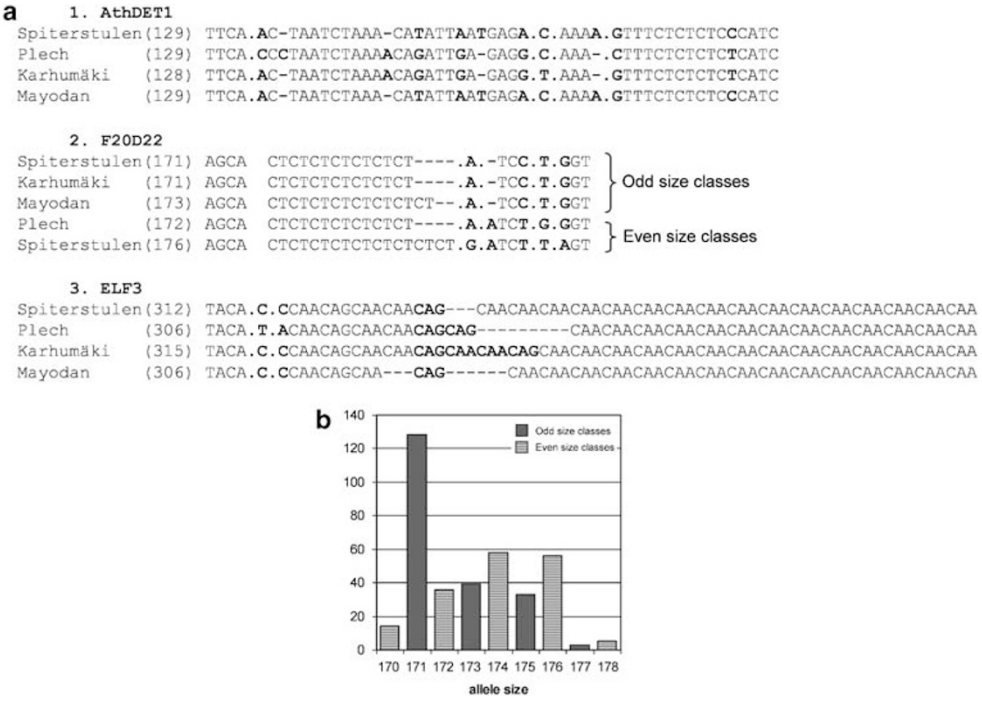

Figure 2 (**a**) Sequence alignment for three representative microsatellite loci. Only the central part of the loci is shown. Nucleotide substitutions and insertions between sequences are highlighted in bold. The size obtained through genotyping is indicated for each allele, as well as the population of origin. (1) AthDET1: compound microsatellite showing evidence of homoplasy. The variation between alleles is due both to indels and point mutation. Homoplasy is observed for the allele size 129. (2) F20D22: perfect microsatellite with other sources of variation than the repeat motif. Homoplasy probably occurs due to point mutations. A single-base insertion gives rise to two kind of size classes (see distribution). Nucleotide substitution within size classes suggests that homoplasy exists even if not directly observed. (3) ELF3: compound microsatellite where the main source of variation is due to variation in repeat number in two distinct stretches of repeats. Homoplasy is observed for the allele size 306. (**b**) Allele size distribution the locus F20D22

Table 2 Summary of genotypic disequilibrium

Locus 1	*Locus 2*	*Population*	*Linkage*	P-*value*
AthZFPG	nga248	Bohemia	NA	<0.00001**
AthDET1	nga1139	Bohemia	NA	0.0004*
ATTSO392	ICE13	Bohemia	26 cM	0.00008*
ICE3	AthSO191	Bohemia	17 cM	0.0004*
AthDET1	AthZFPG	Bohemia	>50 cM	0.0004*
nga1139	ICE2	Lom	NA	0.0002*
F19G10	nga1139	Plech	NA	0.0004*

Locus pairs for which significant genotypic disequilibrium were detected when applying the correct Bonferroni threshold (**$P<0.00005$) and a liberal Bonferroni threshold (*$P<0.0005$).
Linkage: linkage information; NA: not available in *A. lyrata*; >50 cM, different linkage groups; 26 and 17 cM, genetic distance from Kuittinen *et al*. (2004). The first two pairs are located on the same linkage group in *A. thaliana* (Table 1).

known in *A. lyrata*, 867 comparisons were possible: 6.4% were significant among independent loci and 7.7% were significant among linked loci. Linkage only played a small role in the actual disequilibrium and, except in population Bohemia, the loci are largely independent.

Taking into account these results, and the available linkage information, we constructed 20 subsets of 14 or 15 independent loci; these subsets contained loci separated by more than 30 cM, and not showing any significant genotypic disequilibrium with each other (after the liberal Bonferroni correction) in any population.

Variation in the levels of diversity across Europe

The levels of microsatellite diversity were highly variable across populations (Tables 3 and 4). Within Europe, we observed lower diversity statistics in the North, with significant differences between populations from Central Europe (average $H_e=0.57$) and Northern populations (average $H_e=0.28$; Table 4). The population diversity in Mayodan was among the lowest ($H_e=0.21$).

Within Europe, alleles detected in the North were often a subset of those observed in the South. The populations from Central Europe (Plech and Bohemia) had the highest private allelic richness and ratio of private allelic richness to total allelic richness. The Scandinavian populations showed the lowest values of these statistics. Interestingly, among Northern European populations, Karhumäki showed the highest ratio of private allelic richness to allelic richness.

Among 138 tests of Hardy–Weinberg Equilibrium, 23 were significant, revealing an excess of homozygotes. Only three remained significant after Bonferroni correction: one in Bohemia (F19K23) and two in Plech (F19G10 and ICE3). Over all loci, F_{IS} values were significant in the populations Bohemia, Plech, Reykjavik and Mayodan. The same significance results were obtained in all 20 subsets of loci (Table 4).

All tests of mutation-drift equilibrium (BOTTLENECK) detected highly significant heterozygosity excess within Bohemia, on the whole data set as well as on the 20 subsets, and using the IAM as well as the SMM model. In Mayodan and Reykjavik, no deviation from drift-mutation equilibrium was detected, and the T2 statistic was negative or positive depending on the subset of loci. In the other populations, the results were less consistent; under IAM, significant heterozygosity excess was detected in Lom, Stubbsand and Plech in the whole data set. In the subsets of the data, only five subsets were significant for Lom, and none for the others. Overall in these populations, there were always more loci with heterozygosity excess. Accordingly positive T2 values were observed (Table 4) and tests were often marginally significant. Under SMM, no test was significant, and T2 was most often negative.

Population differentiation and relationship with diversity levels

Across populations, F_{ST} was highly significant ($F_{ST}=0.458$. Average over loci, $P<0.0001$) and highly variable across loci. Among subsets of independent loci, the average F_{ST} was 0.471. Among European populations only, F_{ST} value was equal to 0.42.

The overall F_{ST} values for individual loci were significantly negatively correlated with the average within population gene diversity (H_S; Figure 3, $r=-0.719$, $P=0.0002$). One locus, ICE2, did not follow this relationship and had a lower F_{ST} value than expected based on its diversity. When F_{ST} values where standardized to the maximum value they can reach given H_S, the correlation with H_S disappeared (F'_{ST}; Figure 3).

Pairwise F_{ST} values and population relationships

All populations of the data set were significantly differentiated from each other, with highly variable F_{ST} estimates over loci (Table 5). Again a negative relationship was detected between average within-population diversity and F_{ST} values (Figure 4). Some pairs of populations did not follow this relationship: the comparisons between populations belonging to the same region (that is, Scandinavian populations, all showing a low level of diversity), and the comparisons between the population Reykjavik and the Scandinavian populations (Figure 4). The negative relationship was still present when using the standardized values of F_{ST} (F'_{ST}; Figure 4), and for other genetic distances (M Nei and LL Cavalli-Sforza, not shown).

Population relationships are depicted in Figure 5. For both matrices of genetic distances, the same topology was obtained for all 20 subsets of loci, but with variable branch lengths. These trees showed that Scandinavian populations were closely related to each other and to some extent to Reykjavik. The Russian population Karhumäki was more closely related to the North American population, than to Scandinavian populations. The same topology was obtained for trees built based on pairwise F_{ST} or F'_{ST} values.

Discussion

Genome-wide and worldwide pattern of diversity in *A. lyrata*

We analyzed the genetic diversity and structure of *A. lyrata* in a broad-scale sample, based on microsatellite polymorphism. We found wide differences in the levels of diversity among populations and among loci, as well as a high divergence between populations. This general pattern extends and explains partial and sometimes conflicting results previously reported (for example, extent of the difference of diversity between ssp. *lyrata*

Table 3 Diversity statistics per microsatellite locus and per population

Locus	*Bohemia*	*Plech*	*ICE*	*Karhumaki*	*St_Swe*	*Lom_NOR*	*Spiterstulen*	*NC_USA*
AthACS								
R_a	4.715	5.78	2.365	2.948	2	2.88	3.602	1.625
He	0.776	0.799	0.160	0.593	0.312	0.566	0.474	0.063
F_{IS}	0.293**	−0.001	−0.041	−0.192	−0.176	0.058	−0.018	0.000
AthCDPK9								
R_a	2.956	2.999	2.81	1	2.996	1.999	2	1
H_e	0.515	0.664	0.541	0.000	0.450	0.381	0.457	0.000
F_{IS}	0.199	0.098	0.078	—	0.200	0.039	−0.338	—
AthDET1								
R_a	2	3.261	2	1.357	1	1	1	1
H_e	0.503	0.404	0.500	0.0357	0	0	0	0
F_{IS}	0.456*	0.111	0.000	0.000	—	—	—	—
AthSO191								
R_a	2.345	3.762	1.81	1.993	3	1.333	1	1
H_e	0.492	0.553	0.108	0.299	0.610	0.033	0	0
F_{IS}	0.232	0.280	−0.030	0.286	−0.203	0.000		0.000
AthZFPG								
R_a	4.884	7.115	5.454	1	2	2	2.646	2.625
H_e	0.780	0.740	0.789	0	0.526	0.506	0.195	0.434
F_{IS}	0.296**	0.082*	0.230*	—	0.250	−0.055	−0.060	−0.010
ATTSO392								
R_a	6.492	8.263	2.556	2.971	4	3.922	3.983	1.867
H_e	0.807	0.874	0.522	0.502	0.768	0.680	0.718	0.121
F_{IS}	0.223*	0.039	0.47531*	−0.154	0.229	0.060	−0.009	−0.034
ELF3								
R_a	6.111	8.437	4.476	2	4.727	3.423	4.308	3.819
H_e	0.818	0.872	0.632	0.468	0.597	0.586	0.562	0.506
F_{IS}	0.217**	−0.103	0.124	−0.227	0.556**	−0.199	0.113	0.514**
F19G10								
R_a	2	2.875	2.938	2	1	1	1	1
H_e	0.479	0.422	0.381	0.457	0	0	0	0
F_{IS}	−0.008	0.531***	−0.172	−0.338	—	—	—	—
F19K23								
R_a	2.929	1.833	1	1	1	1	1	1.956
H_e	0.479	0.082	0.000	0	0	0	0	0.175
F_{IS}	0.858***	−0.011	—	—	—	—	—	−0.071
F20D22								
R_a	3.97	3.578	3	1	2	2.813	3.822	2.823
H_e	0.721	0.597	0.641	0	0.485	0.549	0.661	0.284
F_{IS}	0.093	0.199	0.048	—	0.259	0.214	0.009	−0.103
ICE12								
R_a	1.998	2.4	1	1.999	1	1	1	2.955
H_e	0.354	0.484	0.000	0.382	0.000	0.000	0	0.486
F_{IS}	−0.073	0.177	—	−0.125	—	—	—	0.234
ICE13								
R_a	2	6.882	3.111	2.837	2	2.372	2.575	2
H_e	0.460	0.787	0.491	0.447	0.507	0.188	0.503	0.444
F_{IS}	0.102	0.035	0.096	0.284	0.107	−0.067	−0.168	−0.132
ICE14								
R_a	3.927	3.962	1	1.84	2	2.787	2.357	1
H_e	0.666	0.735	0.000	0.135	0.100	0.394	0.468	0.000
F_{IS}	0.306*	0.076	—	−0.059	0.000	−0.223	−0.147	—
ICE2								
R_a	2	4.233	1.81	1	2	2.952	2.572	1
H_e	0.407	0.580	0.108	0	0.312	0.599	0.387	0
F_{IS}	−0.018	0.106	−0.030	—	−0.176	0.168*	0.201	0.000

Table 3 Continued

Locus	*Bohemia*	*Plech*	*ICE*	*Karhumaki*	*St_Swe*	*Lom_NOR*	*Spiterstulen*	*NC_USA*
ICE3								
R_a	4.715	8.28	5.741	1.591	4.905	3.653	5.069	1
H_e	0.715	0.850	0.756	0.070	0.753	0.582	0.553	—
F_{IS}	0.231*	0.346***	0.008	1*	−0.091	−0.032	0.066	—
nga106								
R_a	2.885	4.57	5.144	2.591	2	2.119	1.991	3.942
H_e	0.440	0.445	0.710	0.533	0.485	0.129	0.290	0.692
F_{IS}	0.21897	0.285**	0.301*	−0.279	−0.132	−0.036	0.051	−0.088
nga112								
R_a	3.197	6.098	5.055	1	2	2.416	2.69	2
H_e	0.577	0.759	0.667	0	0.337	0.414	0.543	0.387
F_{IS}	0.136	−0.157	−0.152	—	−0.200	−0.007	−0.082	0.362
nga1139								
R_a	5.261	4.955	2.983	1	1.909	1.97	1.826	1.986
H_e	0.754	0.709	0.558	0	0.091	0.235	0.131	0.226
F_{IS}	0.217	0.099	0.239	—	0.000	−0.137	−0.057	0.455
nga151								
R_a	1.345	2	1	1	2	2	2	1
H_e	0.035	0.481	0.000	0	0.455	0.427	0.509	0
F_{IS}	0	−0.083	—	—	−0.429	−0.094	−0.020	—
nga248								
R_a	2.991	2.978	3.111	1.964	1.909	1.893	2.575	3.333
H_e	0.629	0.288	0.562	0.223	0.091	0.098	0.493	0.435
F_{IS}	0.0694	0.170	0.611**	−0.125	0.000	0.322	−0.265	0.701**
nga280								
R_a	4.318	3	2.922	1.591	2	1	1.826	1.867
H_e	0.752	0.679	0.570	0.070	0.507	0	0.131	0.121
F_{IS}	0.038	0.119	0.225*	−0.019	0.474	—	0.477	−0.034
SOC1								
R_a	3.549	3.044	4.476	1.742	1.996	3.25	2.333	1
H_e	0.639	0.460	0.575	0.103	0.173	0.474	0.220	0.000
F_{IS}	−0.026	0.396*	−0.066	−0.038	−0.053	0.016	0.091	—
Over loci								
n	29	25	18	28	11	30	29	16
R_a	3.423	4.382	2.858	1.706	2.121	2.149	2.291	1.943
H_e	0.575	0.591	0.405	0.202	0.324	0.298	0.321	0.208
F_{IS}	0.197***	0.112***	0.135**	−0.093	0.071	0.001	−0.033	0.174*

Significance of deviation from Hardy–Weinberg disequilibrium.
*$P<0.05$.
**$P<0.01$.
***$P<0.001$.

Table 4 Summary statistics of within population microsatellite diversity

	Bohemia	*Plech*	*Reykjavik*	*Lom*	*Spiterstulen*	*Stubbsand*	*Karhumaki*	*Mayodan*
n	29	25	18	30	29	11	28	16
H_e	0.562 a	0.571 a	0.364 b	0.283 b	0.305 b	0.314 b	0.202 b	0.210 b
R_a	3.421 ab	4.133 a	2.776 bc	2.120 cd	2.216 cd	2.159 cd	1.701 d	1.986 cd
A_{pr}	26.1	43.4	13.2	6.4	7.8	8.8	10.6	13.0
% private alleles	0.363	0.472	0.220	0.142	0.163	0.198	0.295	0.318
F_{IS}	0.201***	0.09**	0.138*	−0.033	−0.024	0.058	−0.127	0.228**
T2 IAM	3.46***	*1.18*	0.46	0.76*	0.82	*1.04*	*0.79*	−0.04

Abbreviations: *n*, sample size per population; H_e, gene diversity; R_a, allelic richness.
H_e and R_a are the averages of the estimates obtained for the different subsets of independent loci, excluding ICE3. Values with the same letter (a, b, c…) are not significantly different (Wilcoxon sign-test). A comparison was declared significant if it was significant for more than 75% of the subsets of loci. A_{pr}: sum over 21 loci of the private allelic richness estimated after rarefaction. % of private alleles: ratio of A_{pr} on the sum over loci of the allelic richness (estimated after rarefaction). ICE3 is not included in the calculation. F_{IS}: The estimate and its significance level are averages of values obtained for the different subsets of independent loci. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$.
T2 IAM: T2 statistics computed under the infinite allele model of mutation. The value shown is the average over the 20 subsets of independent loci. ***Sign test and Wilcoxon test significant over all subsets. *Sign test significant for five subsets; italics: consistent positive T2 and majority of loci in heterozygosity excess, although not significant.

and *petraea*, Ramos-Onsins *et al*., 2004; Balana-Alcaides *et al*., 2006), and shows that in this species, a wide sample is necessary to adequately describe the genetic structure.

Within Europe, intrapopulation diversity decreased from Central Europe toward Northern Europe and populations were strongly differentiated. A recent microsatellite study had found a decreased diversity in Iceland, relative to Central Europe (Clauss and Mitchell-Olds, 2006), and had estimated similar H_E and F_{ST} values than ours within Central Europe. In Sweden, an isozyme-based analysis of population structure (Kärkkäinen *et al*., 2004) resulted in F_{ST} values similar to ours between Scandinavian populations. Averaging on a large number of different loci thus leads to robust conclusions, despite the wide variability among loci.

At the subspecies level, we also found a high divergence between *A.l. lyrata* and *A.l. petraea*. Diversity in the *A.l. lyrata* population analyzed for microsatellites (Mayodan) was as low as in the less diverse European populations. However, earlier studies (Van Treuren *et al*., 1997; Balana-Alcaides *et al*., 2006) suggest that North American populations have also widely variable levels of polymorphism. Therefore, general conclusions cannot be drawn on comparisons between *A.l.* ssp. *lyrata* and

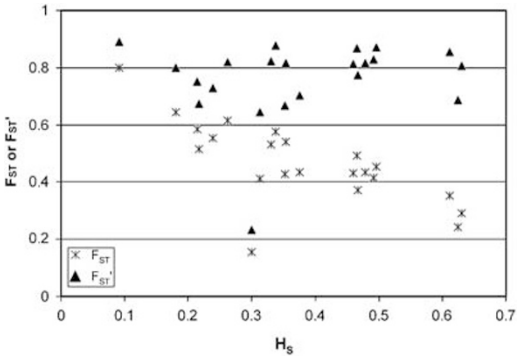

Figure 3 Plot of among-population F_{ST} and F'_{ST} measures against average within-population heterozygosity H_S for individual loci.

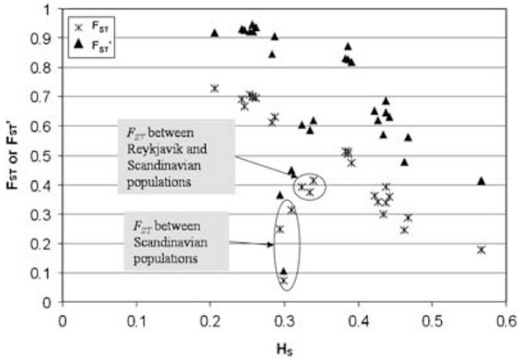

Figure 4 Plot of pairwise F_{ST} and F'_{ST} values against average within-population heterozygosity H_S. All plotted values are average over estimates obtained in the 20 subsets of independent loci.

Table 5 Matrix of pairwise genetic distances between populations

	Bohemia	*Plech*	*Reykjavik*	*Karhumäki*	*Lom*	*Spiterstulen*	*Stubbsand*	*Mayodan*
Bohemia	—	0.346	0.361	1.013	0.634	0.590	0.482	1.332
Plech	0.177	—	0.479	0.993	0.521	0.522	0.618	1.044
Reykjavik	0.244	0.286	—	0.924	0.428	0.433	0.416	1.362
Karhumäki	0.514	0.513	0.614	—	1.372	1.540	1.311	1.224
Lom	0.393	0.360	0.414	0.699	—	0.034	0.219	1.592
Spiterstulen	0.362	0.342	0.393	0.691	0.072	—	0.171	1.556
Stubbsand	0.300	0.337	0.375	0.706	0.315	0.249	—	1.858
Mayodan	0.504	0.474	0.630	0.728	0.694	0.668	0.702	—

Above diagonal: Nei's genetic distance; below diagonal: F_{ST}. The values are averages over 20 subsets of independent loci.

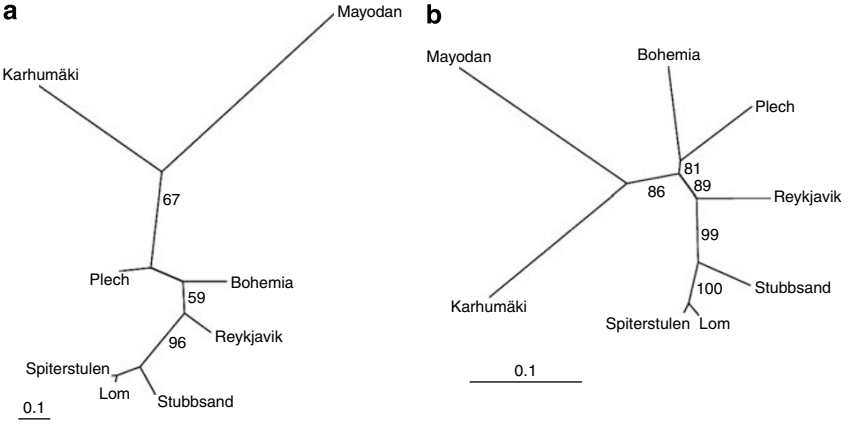

Figure 5 Neighbor-joining trees based on the mean Nei's genetic distance (**a**) and on the mean Cavalli-Sforza's genetic distance (**b**), estimated for 20 subsets of loci. Percentages on each branch indicate the proportion of bootstrap replicates in which the two sets separated by that branch appear.

A.l. ssp. *petraea* before a broad survey of diversity across North America is conducted.

Evolutionary history of *A. lyrata*: postglacial colonization in Europe and worldwide differentiation

The patterns of diversity and differentiation detected within Europe suggest that Northern European populations probably derive from two distinct refugia. First, *A. lyrata* ssp. *petraea* may have persisted in Central Europe during the Last Glacial Maximum (Clauss and Mitchell-Olds, 2006). Expansion from this refugium would have given rise to the Iceland and Scandinavian populations, as shown by decreasing levels of diversity from South to North and patterns of pairwise differentiation. In contrast, the Russian population may originate from the Eastern part of the distribution of ssp. *petraea* (Siberia; Jonsell *et al*., 1995). This hypothesis is supported by high pairwise F_{ST} values between Karhumäki and the other European populations and by a high relative amount of private microsatellite alleles. Karhumäki had already been shown to differ from other European populations, with estimates of population differentiation at markers or sequences (Van Treuren *et al*., 1997; Balana-Alcaides *et al*., 2006), leaf morphological variation (Jonsell *et al*., 1995) or segregation distortion in crosses with a Swedish population (Kuittinen *et al*., 2004).

The founder effects induced by the postglacial colonization from Central Europe are very strong, with a twofold difference in diversity statistics. For comparison, in tree species, the magnitude of variation of within-population diversity is lower (H_e varies between 0.85 and 0.7 for microsatellite loci in common ash; Heuertz *et al*., 2004). For *Dryas octopetala*, an Arctic–Alpine species widely present in Northern habitats, the diversity is even reduced in Southern compared to Northern populations (Skrede *et al*., 2006). The fate of genetic diversity during colonization indeed strongly depends on the life-history traits of the species (Austerlitz *et al*., 2000), and on the fate of the different routes of colonization from the refugia (Taberlet *et al*., 1998). Two routes of colonization of Fennoscandinavia have been recognized; in some cases, they have met, resulting in admixture, as in Northern Sweden for *Cerastium alpinum* (Nyberg Berglund and Westerbergh, 2001), which can increase local population diversity. For *A.l.* ssp. *petraea*, our results suggest that two separate routes of recolonization exist but they have remained distinct. Moreover, the species has a fragmented distribution, with populations distant from each other (Figure 1): *A. lyrata* has probably disappeared from the regions between Central Europe and its present locations in Northern Europe. Its seeds are only dispersed by gravity (Clauss and Mitchell-Olds, 2006). All these characteristics result in a low level of gene flow, increasing genetic drift between populations, and preventing the genetic homogenization between regions and populations. This contrasts for instance with the high levels of pollen flow and continuous populations described for trees (Austerlitz *et al*., 2000) or *D. octopetala* (Skrede *et al*., 2006).

Overall, only weak evidence of bottlenecks was detected in the North. As the species appears to be a pioneer cold-tolerant species, and as it persisted in Central Europe during the Ice Ages, colonization of Northern Europe may have occurred rapidly after the retreat of the ice. The excess of heterozygosity induced by bottlenecks and founder effects is detectable only in a window of time of 0.5 to 5 times the effective population size of the bottlenecked population (Cornuet and Luikart, 1996). Consequently, for rapid recolonization and small population sizes, the microsatellite markers may have had time to recover a pattern resembling drift-mutation equilibrium.

At the whole species level, we showed a strong differentiation between three groups, North America, Russia and Western Europe, but the first two regions are only represented by one population each. The genealogical relationships between the three subspecies are not well resolved (Al-Shehbaz and O'Kane, 2002; Beck *et al*., 2007) and no information is available on the population structure of ssp. *kamchatica* and on ssp. *petraea* in the Eastern part of its distribution. The subspecies ssp. *lyrata* could be directly derived from Western European populations, through Trans-Atlantic dispersal (Abbott and Brochmann, 2003). It could also represent the extreme part of a worldwide distribution across Siberia and Alaska. These contrasting hypotheses have been debated for many circumboreal species (Abbott and Brochmann, 2003). Distinguishing them, as well as understanding the consequences of climatic oscillations within each region requires sampling extensively in Northwestern America, Alaska, Eastern Asia and Russia.

We observed or inferred homoplasy in the microsatellite polymorphism, but we estimated statistics designed for markers mutating according to the IAM. Homoplasy generally results in lower power to detect population differentiation and to underestimation of genetic distances (Queney *et al*., 2001). Here, it would be conservative and not alter our general conclusions. However, its impact is probably stronger in comparison between continents and other kinds of markers, such as nucleotide sequences have to be used for a more precise description of genetic and genealogic relationships at the species level (Queney *et al*., 2001).

Local population diversity and structure

Northern European populations are at Hardy–Weinberg equilibrium, but significant positive F_{IS} values have been estimated in Central Europe and in North America. Similar results were obtained by Balana-Alcaides *et al*. (2006) on two genes. *A. lyrata* is strictly self-incompatible even if a breakdown of the self-incompatibility system is observed in some populations (for example, Mable *et al*., 2005). Gene flow can however be restricted to short distances within population (Schierup *et al*., 2006), and high differentiation has been detected even within the continuous distribution of the species in Iceland (Schierup, 1998). Population substructure may therefore explain significant F_{IS} values. Indeed, in Plech, Clauss and Mitchell-Olds (2006) estimated an F_{ST} of 0.092 between different patches of the population, a value identical to the F_{IS} we estimated for this population (where the location of the individuals is unknown). The Bohemia sample is from laboratory propagation, and it serves us only as a reference for allelic composition, without further demographic interpretation. Finally, within Mayodan, probably null alleles at two loci are mainly responsible for the significant deviation from Hardy–Weinberg equilibrium, (ELF3 and nga 248).

Markers with different mutation rates and populations with different sizes: proceed with caution!

The microsatellite loci surveyed showed highly different diversity levels, and thus probably had different mutation rates. A strong negative correlation was detected between average intrapopulation diversity (H_S) and overall population differentiation estimated for individual loci (Figure 3). This relationship has been investigated by Hedrick (1999, 2005); from Hedrick's graph (Figure 2 in Hedrick, 2005), we can see that in a model of divergence between populations, the negative correlation between F_{ST} and H_S increases with the time of divergence. Hedrick showed that F_{ST} cannot be higher than a maximum value dependent on H_S and proposed a standardized value (F'_{ST}) where F_{ST} is divided by this maximum value. However, F'_{ST} is not independent of H_S: in the same model of population divergence, there is a positive correlation between F'_{ST} and H_S; this correlation decreases with the time of divergence and progressively disappeared as F_{ST} reaches its asymptotic value (Hedrick, 2005). In our case, the negative correlation was strong between overall F_{ST} and H_S, and absent between F'_{ST} and H_S (Figure 3), suggesting that the divergence between the whole set of populations is so old that the asymptotic F_{ST} values are almost reached. When calculated among Scandinavian populations, which have probably diverged more recently, the per-locus F_{ST} values are not correlated with H_S (not shown).

This effect of diversity on F_{ST} estimates is more complex when considering pairwise population comparisons. Indeed, populations of *A. lyrata* have highly different levels of diversity, possibly reflecting different population sizes. We observed a negative correlation between pairwise F_{ST} among loci and average H_S in the compared populations (Figure 4). This relationship can be explained both by the inherent effect of H_S on F_{ST} (see above) and by the fact that genetic drift is stronger in smaller populations. Moreover, it is also observed for the standardized measure proposed by Hedrick; this is expected as pairs of populations with smaller effective sizes reach their asymptotic F_{ST} value faster. These confounding effects of population size and time of divergence raise the question of whether our different pairwise F_{ST} estimates can be interpreted with respect to different migration rates or divergence times (Figure 5) or only in terms of different population sizes. Recall that we also observed a negative correlation of other measures of genetic distance with H_S. Additionally, Scandinavian populations did not follow the common correlation, and were clearly less differentiated from each other than from the Russian population (which had a similar low level of diversity). This suggests that Karhumäki's special status is not only due to a small population size.

Implication for the study of local adaptation in *A. lyrata*

Detecting the loci and nucleotide sites responsible for adaptive variation relies on detecting deviations from the standard neutral model (Wright and Gaut, 2005), or on detecting deviations from a genome wide pattern of polymorphism (for example, Beaumont and Balding, 2004). The demographic history of *A. lyrata* alone has given rise to large reductions in variation and high divergence, as shown by the genome-wide pattern detected from microsatellites in this study. It is possible that population structure effects may even be important at the within population level (Clauss and Mitchell-Olds, 2006; Schierup *et al*., 2006). Thus, any tests for the adaptive significance of nucleotide variation will need to be based on deviation from this pattern. The results also imply that any association tests will need to take this population structure into account.

Acknowledgements

We acknowledge financial support from the Academy of Finland. We thank CH Langley, Mikkel Schierup, Maria Clauss, Marc MacNair for seeds, Helmi Kuittinen and Anne Tsitrone for constructive comments on this paper.

References

Abbott RJ, Brochmann C (2003). History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Mol Ecol* **12**: 299–313.

Al-Shehbaz IA, O'Kane SL (2002). Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). In: Somerville C, Meyerowitz E (eds). *The Arabidopsis Book*. American Society of Plant Biologists: Rockville, MD, USA.

Alsos IG, Engelskjon T, Gielly L, Taberlet P, Brochmann C (2005). Impact of ice ages on circumpolar molecular diversity: insights from an ecological key diversity. *Mol Ecol* **14**: 2739–2753.

Austerlitz F, Mariette S, Machon S, Gouyon P-H, Godelle B (2000). Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* **154**: 1309–1321.

Balana-Alcaides D, Ramos-Onsins SE, Boone Q, Aguadé M (2006). Highly structured nucleotide variation within and among *Arabidopsis lyrata* populations at the *FAH1* and *DFR* gene regions. *Mol Ecol* **15**: 2059–2068.

Beaumont M, Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**: 969–980.

Beck JB, Al-Shehbaz IA, O'Kane SL, Schaal BA (2007). Further insights into the phylogeny of *Arabidopsis* (Brassicaceae) from nuclear *Atmyb2* flanking sequence. *Mol Phylogenet Evol* **42**: 122–130.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001). *GENETIX 4.02, Logiciel Sous Windows™ Pour La GéNéTique Des Populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II: Montpellier, France.

Bell CJ, Ecker JR (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**: 137–144.

Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* **19**: 233–257.

Clauss MJ, Cobban H, Mitchell-Olds T (2002). Cross-species microsatellites markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Mol Ecol* **11**: 591–601.

Clauss MJ, Koch MA (2006). Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci* **11**: 449–459.

Clauss MJ, Mitchell-Olds T (2006). Population genetics structure of *Arabidopsis lyrata* in Europe. *Mol Ecol* **15**: 2753–2766.

Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.

Felsenstein J (2005). *PHYLIP (Phylogeny Inference Package) Version 3.6*, Distributed by the author. Department of genome sciences, University of Washington: Seattle.

Hedrick PW (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.

Hedrick PW (2005). A standardized genetic differentiation measure. *Evolution* **59**: 1633–1638.

Heuertz M, Hausman J-F, Hardy OJ, Vendramin GG, Frascaria-Lacoste N, Vekemans X (2004). Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southeastern European populations of the common ash (*Fraxinus excelsior* L.). *Evolution* **58**: 976–988.

Hewitt GM (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci* **359**: 183–195.

Jalas J, Suominen J (1994). Atlas florae Europaeae, Vol. 10, *Distribution of Vascular Plants in Europe*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki, Finland.

Jonsell B, Kustås K, Nordal I (1995). Genetic variation in *Arabis petraea*, a disjunct species in northern Europe. *Ecography* **18**: 321–332.

Kalinowski ST (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.

Kärkkäinen K, Loe G, Agren J (2004). Population structure in *Arabidopsis lyrata*: evidence for divergent selection on trichome production. *Evolution* **58**: 2831–2836.

Kivimäki M, Kärkkäinen K, Gaudeul M, Loe G, Agren J (2007). Gene, phenotype and function: *GLABROUS1* and resistance to herbivory in natural populations of *Arabidopsis lyrata*. *Mol Ecol* **16**: 453–462.

Kuittinen H, Aguadé M, Charlesworth D, de Haan AA, Lauga B, Mitchell-Olds T *et al*. (2002). Primers for 22 candidate genes for ecological adaptations in Brassicaceae. *Mol Ecol Notes* **2**: 258–262.

Kuittinen H, de Haan AA, Vogl C, Oikarinen S, Leppälä J, Koch M *et al*. (2004). Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics* **168**: 1575–1584.

Mable BK, Robertson AV, Dart S, Di Berardo C, Witham L (2005). Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* **59**: 1437–1448.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia Univeristy Press: New York.

Nyberg Berglund A-B, Westerbergh A (2001). Two post-glacial immigration lineages of the polyploid *Cerastium alpinum* (Caryophyllaceae). *Hereditas* **134**: 171–183.

Petit RJ, El Mousadik A, Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* **12**: 844–855.

Piry S, Luikart G, Cornuet J-M (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Heredity* **90**: 502–503.

Ponce MR, Robles P, Micol JL (1999). High-throughput genetic mapping in *Arabidopsis thaliana*. *Mol Gen Genet* **261**: 408–415.

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001). Stationary distribution of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Mol Biol Evol* **18**: 2169–2178.

Ramos-Onsins SE, Stranger B, Mitchell-Olds T, Aguadé M (2004). Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**: 373–388.

Raymond M, Rousset F (1995). Population genetics software for exact test and ecumenicism. *J Heredity* **86**: 248–249.

Rice WR (1989). Analyzing tables of statistical tests. *Evolution* **43**: 223–225.

Riihimäki M, Podolsky RH, Kuittinen H, Koelewijn H, Savolainen O (2005). Studying genetics of adaptive variation in model organisms: flowering time variation in *Arabidopsis lyrata*. *Genetica* **123**: 63–74.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenic trees. *Mol Biol Evol* **4**: 406–425.

Savolainen O, Langley CH, Lazzaro BP, Fréville H (2000). Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol Biol Evol* **17**: 645–655.

Schierup MH (1998). The effect of enzyme heterozygosity on growth in a strictly outcrossing species: the self-incompatible *Arabis petrae* (Brassicaceae). *Hereditas* **128**: 21–31.

Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB (2006). Selection at work in self-incompatible *Arabidopsis lyrata*: mating patterns in a natural population. *Genetics* **172**: 477–484.

Skrede I, Bronken Eidesen P, Pineiro Portela R, Brochmann C (2006). Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L). *Mol Ecol* **15**: 1827–1840.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* **7**: 453–464.

Van Treuren R, Kuittinen H, Kärkkäinen K, Baena-Gonzalez E, Savolainen O (1997). Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol Biol Evol* **14**: 220–229.

Weir BS (1996). *Genetic Data Analysis II*. Sinauer Associates Inc.: Sunderland.

Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Looseley M, Gaut BS *et al*. (2006). Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* **174**: 1421–1430.

Wright SI, Gaut BS (2005). Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol* **22**: 506–519.