Contrasting Patterns of Nucleotide Polymorphism at the Alcohol Dehydrogenase Locus in the Outcrossing *Arabidopsis lyrata* and the Selfing *Arabidopsis thaliana*

Outi Savolainen, C. H. Langley,† B. P. Lazzaro,†[1] and Hélène Fréville*[2]*

*Department of Biology, University of Oulu, Oulu, Finland; and †Center for Population Biology, University of California at Davis

Nucleotide variation at the alcohol dehydrogenase locus (*Adh*) was studied in the outcrossing *Arabidopsis lyrata*, a close relative of the selfing *Arabidopsis thaliana*. Overall, estimated nucleotide diversity in the North American ssp. *lyrata* and two European ssp. *petraea* populations was 0.0038, lower than the corresponding specieswide estimate for *A. thaliana* at the same set of nucleotide sites. The distribution of segregating sites across the gene differed between the two species. Estimated sequence diversity within an *A. lyrata* population with a large sample size (0.0023) was much higher than has previously been observed for *A. thaliana*. This North American population has an excess of sites at intermediate frequencies compared with neutral expectation (Tajima's $D = 2.3$, $P < 0.005$), suggestive of linked balancing selection or a recent population bottleneck. In contrast, an excess of rare polymorphisms has been found in *A. thaliana*. Polymorphism within *A. lyrata* and divergence from *A. thaliana* appear to be correlated across the *Adh* gene sequence. The geographic distribution of polymorphism was quite different from that of *A. thaliana*, for which earlier studies of several genes found low within-population nucleotide site polymorphism and no overall continental differentiation of variation despite large differences in site frequencies between local populations. Differences between the outcrossing *A. lyrata* and the selfing *A. thaliana* reflect the impact of differences in mating system and the influence of bottlenecks in *A. thaliana* during rapid colonization on DNA sequence polymorphism. The influence of additional variability-reducing mechanisms, such as background selection or hitchhiking, may not be discernible.

Introduction

Understanding of the genetic and ecological mechanisms that shape intraspecific DNA sequence polymorphism within and phylogenetic divergence among natural populations of related species is a fundamental goal of evolutionary biological research. While studies of *Drosophila* populations have provided a wealth of information on intraspecific nucleotide polymorphism (Moriyama and Powell 1996), investigations of plant populations have until now contributed little.

Most of the empirically testable population genetics theories related to the maintenance of variation assume a random mating population at mutation-drift equilibrium (Hudson 1990). Most research on DNA sequence polymorphims has been conducted on random-mating species, such as *Drosophila* and humans. Surveys of outcrossing plant populations allow testing of the theories. Furthermore, such data can also be readily compared with results from diverse random-mating taxonomic groups.

Many plant species, however, are partially or predominantly selfing. Mating system is a major determinant of the distribution of genetic variation within and between populations (e.g., Allard, Jain, and Workman 1968). Recent reviews have examined the role of the mating system in governing variation at marker loci (Hamrick and Godt 1990, 1996) and in quantitative traits (Charlesworth and Charlesworth 1995). Several factors may lead to reduced variation at neutral loci in species with selfing. Complete selfing alone is expected to decrease the effective population size and, thus, nucleotide diversity within a populations by one half (Pollak 1987) relative to a random-mating population. Due to the lowered within-population diversity, the relative proportion of between-populations variation (F_{st}) will be increased. The effects of selfing are often confounded with effects of other life history traits, such as weediness. Selfing and weediness may lead to frequent bottlenecks, which result in a further reduction in genetic diversity within populations due to founder effects and drift. Many selfing species may also be more likely than outcrossers to occur in a metapopulation system, in which high rates of extinction will lead to a lower overall effective population size and reduced nucleotide polymorphism (e.g., Wade and McCauley 1988; Barton and Whitlock 1997).

Moreover, predominantly selfing species have reduced effective recombination due to rarity of double heterozygotes. Linkage disequilibrium will extend over larger areas than in random mating species, and selected loci may influence large areas of the genome. Thus, hitchhiking due to advantageous mutations (Kaplan, Hudson, and Langley 1989) and background selection caused by deleterious mutations (Charlesworth, Morgan, and Charlesworth 1993; Charlesworth, Charlesworth, and Morgan 1995; Charlesworth, Nordborg, and Charlesworth 1997) may both be important in selfing species (Nordborg, Charlesworth, and Charlesworth 1996). Background selection could account for cases of

[1] Present address: Department of Biology, Pennsylvania State University.

[2] Present address: ISEM Génétique et Environnement CC065, Université de Montpellier-2, Montpellier, France.

Key words: *Arabidopsis lyrata*, *Arabidopsis thaliana*, alcohol dehydrogenase, nucleotide polymorphism.

Address for correspondence and reprints: Outi Savolainen, Department of Biology, University of Oulu, P.O. Box 3000, FIN-90401 Oulu, Finland. E-mail: outi.savolainen@oulu.fi.

Mol. Biol. Evol. 17(4):645–655. 2000

reduced variation within populations and for the relatively high F_{st} values of selfing populations (Nordborg 1997).

As discussed by Charlesworth, Morgan, and Charlesworth (1993), it is not easy to distinguish between the effects of these different factors. However, hitchhiking effects should be found only at some loci, bottleneck effects at all loci but only in some populations, and background selection only in specific regions of the genome.

Few studies comparing nucleotide diversity in closely related selfing and outcrossing species are currently available (but see Liu, Zhang, and Charlesworth 1998; Stephan and Langley 1998; Liu, Charlesworth, and Kreitman 1999).

For both practical and theoretical reasons, *Arabidopsis thaliana* L. (Heynh.) and its relatives are opportune species for examining the effects of mating system and life history on DNA sequence variation. *Arabidopsis thaliana* is a highly selfing annual species, with less than 1% outcrossing (Abbott and Gomes 1989). Patterns of historical recombination are also consistent with a low level of outcrossing (Innan et al. 1996). *Arabidopsis thaliana* is a weedy species and occurs in disturbed habitats. It seems to have spread recently from Asia to other parts of the world (Innan et al. 1996). Nucleotide polymorphism has recently been studied at the alcohol dehydrogenase locus (e.g., Hanfstingl et al. 1994; Innan et al. 1996) and several other loci (discussed below).

The close relative species *Arabidopsis lyrata* (L.) O'Kane and Al-Shehbaz consists of several subspecies: the North American ssp. *lyrata,* the European ssp. *petraea,* and the Far Eastern ssp. *kamchatica* and ssp. *kawasakiana.* The North American and European subspecies have previously been referred to as *Arabis lyrata* (L.) and *A. petraea* (L.) Lam., or *Cardaminopsis petraea* Hiit. O'Kane and Al-Shehbaz (1997) used rDNA sequences to justify moving these species to the genus *Arabidopsis.* Several studies suggest that *A. lyrata* and its relatives are the closest species to *A. thaliana* (Price, Palmer, and Al-Shehbaz 1994; O'Kane and Al-Shehbaz 1997; Koch, Bishop, and Mitchell-Olds 1999). Figure 1 shows the distributions of several of these species based on Hopkins (1937), Rollins (1941), Meusel, Jaeger, and Weinert (1965), Hultén and Fries (1986), and Jalas and Suominen (1994). *Arabidopsis lyrata* ssp. *lyrata* and ssp. *petraea* are diploid (2n = 16) (Jones 1963), perennial, and self-incompatible, with outcrossing rate estimates close to 1.0 (Schierup 1998; Kärkkäinen et al. 1999). While a low level of mating between relatives may occur in some populations, most populations exhibit genotype frequencies very close to Hardy-Weinberg equilibrium (Van Treuren et al. 1997; Schierup 1998). *Arabidopsis lyrata* occurs on sandy and rocky shores and serpentines.

The aim of this study was to survey the patterns of nucleotide polymorphism in the outcrossing *A. lyrata.* The comparison of *Adh* polymorphism between the outcrosser and the selfer *A. thaliana* and a comparison of polymorphism and divergence allows evaluation of the impact of different mating systems and life histories.

Materials and Methods

Populations

Arabidopsis lyrata ssp. *petraea* was represented by samples consisting of maternal seed families collected in 1993 from two populations. The Karhumäki, Russia (62°55′N, 34°25′E), population grows on the sandy riverbanks of the Kumsa river. The Iceland individuals were from families collected by Mikkel Schierup close to Reykjavik (Schierup 1998). *Arabidopsis lyrata* ssp. *lyrata* seeds were collected in 1992 from a population on the shore of Lake Michigan, Indiana Dunes, Ind. (41°41′N, 86°58′W). Thirty individuals derived from different maternal families were grown for this study. The Karhumäki and Indiana populations have previously been studied for isozymes and microsatellites (Jonsell, Kustås, and Nordal 1995; Van Treuren et al. 1997).

Molecular Methods for Studying the Alcohol Dehydrogenase Gene of *A. lyrata*

Adh has become the model gene for studies of sequence variation in plants (Gaut and Clegg 1993*a*, 1993*b*; Innan et al. 1996; Bergelson et al. 1998; Liu, Zhang, and Charlesworth 1998). For this reason, we chose this gene in *A. lyrata.*

Genomic DNA was isolated from leaves of mature plants by several published protocols (Rogers and Bendich 1985; Doyle and Doyle 1990). The *A. lyrata Adh* was PCR-amplified with primers designed using sequence information in Chang and Meyerowitz (1986). The nucleotide sites proximal to each primer of the various amplicons are indicated in table 2. Combinations of these primers were used to amplify parts of the gene for sequencing. Some of the data were derived from sequencing cloned PCR products (all of the ssp. *petraea* data and some of the ssp. *lyrata* individuals). In such cases, high-fidelity polymerases (Vent or Pfu) were used, and multiple clones from the same individuals were sequenced in most cases for both strands. Direct sequencing was also used for the ssp. *lyrata* plants. DNA sequences were determined from PCR fragments on the ABI 377 automated sequencer utilizing the manufacturer's chemistry. The single-strand conformational polymorphism (SSCP)/sequence analysis of the Indiana Dunes population was conducted as described in Aguadé et al. (1994).

Data Analysis

The *A. lyrata* sequences were aligned with the *A. thaliana* sequence in Chang and Meyerowitz (1986) using the CLUSTAL W multiple-sequence alignment program (Thompson, Higgins, and Gibson 1994). The aligned sequence corresponds to sites 1048–2699 of the *A. thaliana* sequence (GenBank M12196). The 11 sequences listed in table 1 have been submitted to EMBL and have accession numbers AJ251276–AJ25190. The Indiana Dunes sequences (except 34-1) had some missing nucleotides in exon 5 and were submitted as two entries. The alignment can be derived from table 1.

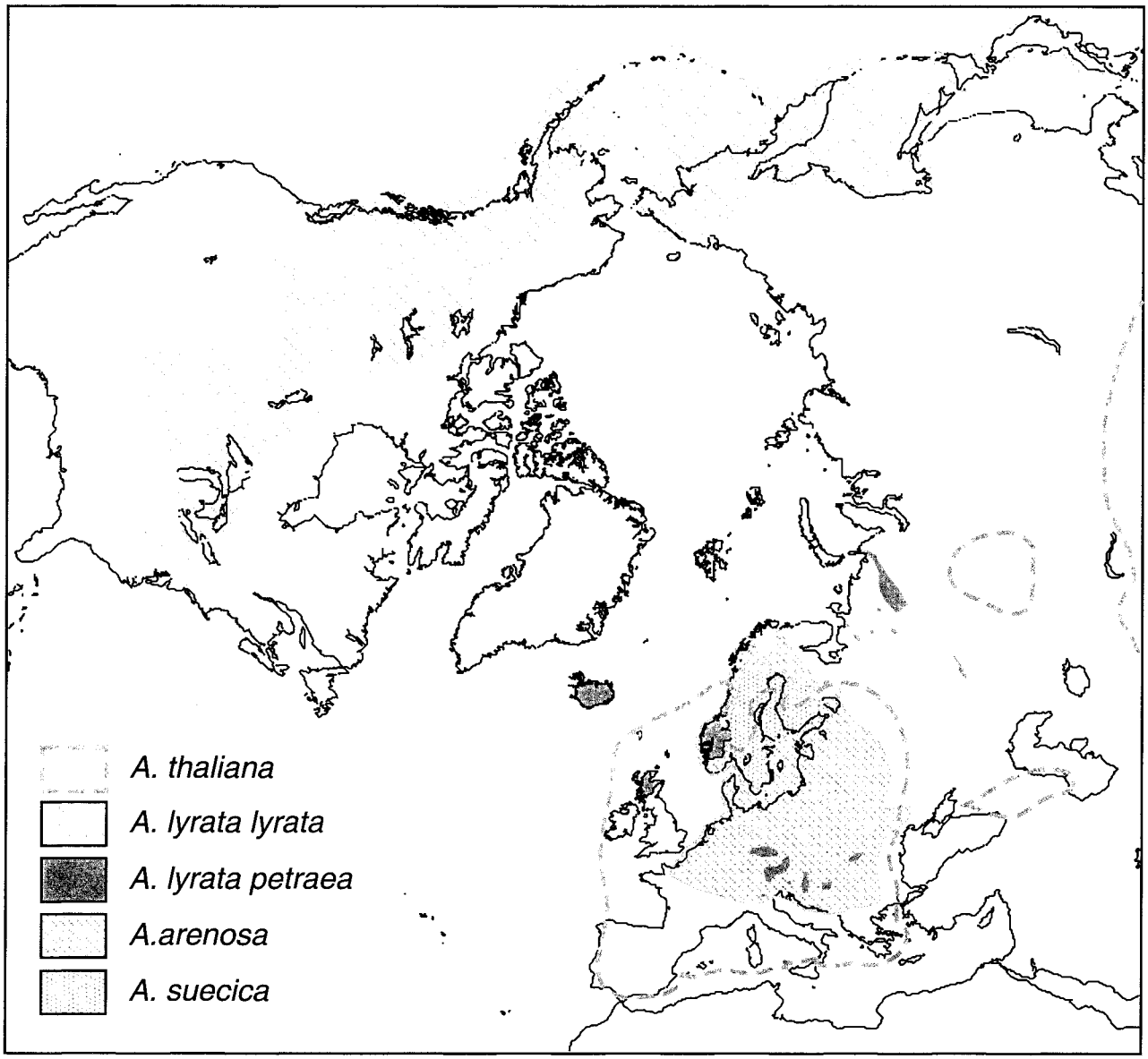

FIG. 1.—Geographic distributions of *Arabidopsis thaliana, Arabidopsis lyrata* ssp. *lyrata* and ssp. *petraea, Arabidopsis suecica,* and *Arabidopsis arenosa.*

The *A. lyrata* data consist of two separate sets. First, we have the sequences of clones of PCR products from six plants from *A. lyrata* ssp. *petraea* (two from Iceland and four from Russia), as described in table 1. For the North American *A. lyrata* ssp. *lyrata,* there are five individuals with sequence information on both alleles; thus, there is a set of 10 whole sequences. One allele from each individual is listed in table 1. Furthermore, there are 10 other North American *A. lyrata* ssp. *lyrata* individuals which have been typed for seven SSCP fragments (table 2). For this second set, full-length allele information is not available for those plants which are heterozygous for multiple SSCP fragments, as phase information for the fragments is missing. The sequence of each SSCP fragment class is also given in table 1. The distribution of the SSCP classes among the *A. lyrata* ssp. *lyrata* plants are given in table 2. The genotypes for each site of the *A. lyrata* plants can be derived from this information. For example, individual 33-1 in table 2 is inferred to be a heterozygote for classes 1.1/1.2 at SSCP fragment 1. The sequences of these fragment classes are derived from PCR sequencing of individuals 20-1, 13-2, and 34-1. In four SSCP assays (individual 4-2 for segment 1, individual 19-1 for segment 2, individuals 4-6 and 4-2 for segment 3), banding patterns initially scored as variant failed to be confirmed by sequencing. The sequence of each SSCP class could be determined in at least one homozygous individual. The sequence traces of segment 7 from individual 13-2 appeared to be a heterozygote of sequences determined from homozygotes 7.1 and 7.2, consistent with the SSCP pattern.

All analyses were based on analysis of individual sites without using the haplotype information (except for

Table 1
Sequences of Cloned *Adh* Alleles (Ice and Kar) and Sequences Deduced from SSCP Fragments (Ind) of *Arabidopsis lyrata*

SSCP:	1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	3	3	6	6	7	7	7
Region:						I1						I2		E3	I3				I5		E6		I6
Site:	1056.1	1058	1067.1–3	1078	1079	1099.10	1099.21	1099.22	1133	1136	1333	1411	1412.1–2	1431	1484	1532	1543.1–16	1544–5	2311	2325	2437	2444	2606
Plant																							
Ice-2	T_2	T	TGA	—	C	C	T	T	C	G		T	C	—	A		C	T	T_9G	TT	C	C			G	T	A
Ice-18	T_2	T	—	—	C	C	T	T	C	G		T	C	—	A		C	T	T_9G	TT	C	C			G	T	A
Kar-8	T_2	T	—	—	T	A	—	T	C	G		T	T	—	A		C	T	$T_{13}G$	TT	T	C			G	T	G
Kar-26	T_5	T	—	—	T	A	—	T	C	G		T	T	—	A		C	T	$T_{14}G$	TT	T	C			G	T	G
Kar-12	T_5	T	—	—	T	A	—	T	C	G		T	T	—	A		C	T	$T_{15}G$	TT	T	C			G	T	G
Kar-24	T_5	T	—	—	C	C	T	T	C	G		T	T	—	A		C	T	$T_{12}G$	TT	T	C			G	T	G
Ind 34-1	?	?	—	—	C	C	T	T	C	G		A	C	CT	C		T	—	—	—	C	C			C	C	G
Ind 19-1	T_2	A	—	—	T	T	A	—	C	T	T	A	C	CT	C		T	—	T_8G	TT	C	G			G	T	G
Ind 20-1	T_2	A	—	—	T	T	A	—	C	T	T	A	C	CT	C		T	—	—	—	C	G			G	T	G
Ind 4-2	T_2	A	—	—	T	T	A	—	C	T	T	A	C	CT	C		T	—	T_8G	TT	C	G			G	T	G
Ind 13-2a	T_2	A	—	—	T	T	A	—	C	T	T	A	C	CT	C		T	—	T_8G	TT	C	G			G	T	G
1.1	T_2	A	—	—	T	T	A	—	C	T	T	2.1 ...	A	CT	C	3.1 ...	T	—	T_8G	TT	G	G	7.1 ...		G	T	G
1.2	T_2	A	—	—	C	C	C	T	T	C	G					3.2 ...	T	—	—	—	G	C	7.2 ...		C	C	G

NOTE.—The last two rows are sequences of the SSCP classes listed in table 2. Numbering of sites is relative to the *Arabidopsis thaliana* sequence of Chang and Meyerowitz (1986). Insertions relative to the *A. thaliana* sequence are indicated with additional digits after the *A. thaliana* site number. Populations are from Iceland (Ice), Karhumäki (Kar), and Indiana Dunes (Ind).

Table 2
SSCP Classes in Seven Segments of the *Adh* Gene From a Sample from a Natural Population of *Arabidopsis lyrata* from Indiana Dunes

	SSCP FRAGMENT						
PLANT ID	1	2	3	4	5	6	7
4-2	**1**	**1**	**1**	**1**	**1**	**2**	**1**
11-2	1	1	1	1	1	2	1
28-2	1	1	1	1	1	2	1
13-2	**1**	**1**	**1**	**0**	**1**	**0**	**1, 2**
29-1	1	1	1, 2	1	1	1, 2	1, 2
20-1	**1**	**1**	**2**	**1**	**1**	**2**	**1**
19-1	**1**	**2**	**1**	**1**	**1**	**2**	**1**
15-1	1, 2	1	1, 2	1	1	1, 2	1
4-6	1, 2	1	1, 2	1	1	1, 2	1, 2
16-1	1, 2	1	1, 2	1	1	1, 2	1, 2
22-1	1, 2	1	1, 2	1	1	1, 2	1, 2
23-1	1, 2	1	1, 2	1	1	1, 2	1, 2
25-1	1, 2	1	1, 2	1	1	1, 2	1, 2
33-1	1, 2	1	1, 2	1	1	1, 2	1, 2
34-1	**2**	**1**	**2**	**1**	**1**	**1**	**2**
Primer F	1028	1267	1427	1673	1918	2150	2382
Primer R	1315	1473	1721	1964	2206	2438	2723

NOTE.—Bold numbers in the table indicate the segments that were sequenced. "0" indicates that no SSCP class was assigned but the fragment was sequenced. The numbers of the first and last base pairs proximal to the 5′ and 3′ primers (respectively) of each fragment relative to the *Arabidopsis thaliana* sequence (M12196) are in the last two rows.

analysis of recombination). Insertion/deletion (indel) variation was not included in the estimates of nucleotide diversity (π). Sites with missing data were excluded from individual estimates of nucleotide diversity. Thus, the within-population estimates were based on slightly different numbers of nucleotides. Standard deviations of nucleotide diversity were estimated following Nei (1987, without and including the stochastic variance using eqs. 10.7 and 10.9, respectively), even though the formulae do not take into account substructuring. The computations were carried out using DnaSP, version 3 (Rozas and Rozas 1999). For assessing between-populations variation, we estimated nucleotide diversity for the whole species and for *A. lyrata* ssp. *petraea* separately. As the sample sizes were quite different between populations, we used unweighted site frequencies to obtain these estimates, as described by Holsinger and Mason-Gamer (1996).

Estimates of divergence between populations of *A. lyrata* ssp. *lyrata*/ssp. *petraea*, measured as g_{st}, were obtained using standard methods, slightly modified by Holsinger and Mason-Gamer (1996). Indel differences between the consensus sequences of ssp. *lyrata* (Indiana Dunes) and ssp. *petraea* population Karhumäki and the sequence Ice-2 (since there were only two sequences) and *A. thaliana* were estimated.

Nucleotide polymorphism and divergence between *A. lyrata* and *A. thaliana* was estimated based on our set of 36 *Adh* sequences and the set of 17 *A. thaliana* sequences described by Miyashita, Innan, and Terauchi (1996) and Innan et al. (1996). We used the Jukes-Cantor distances for all sites (as in fig. 2) or separate estimates for synonymous and nonsynonymous sites, obtained by DnaSP, version 3 (Rozas and Rozas 1999). For

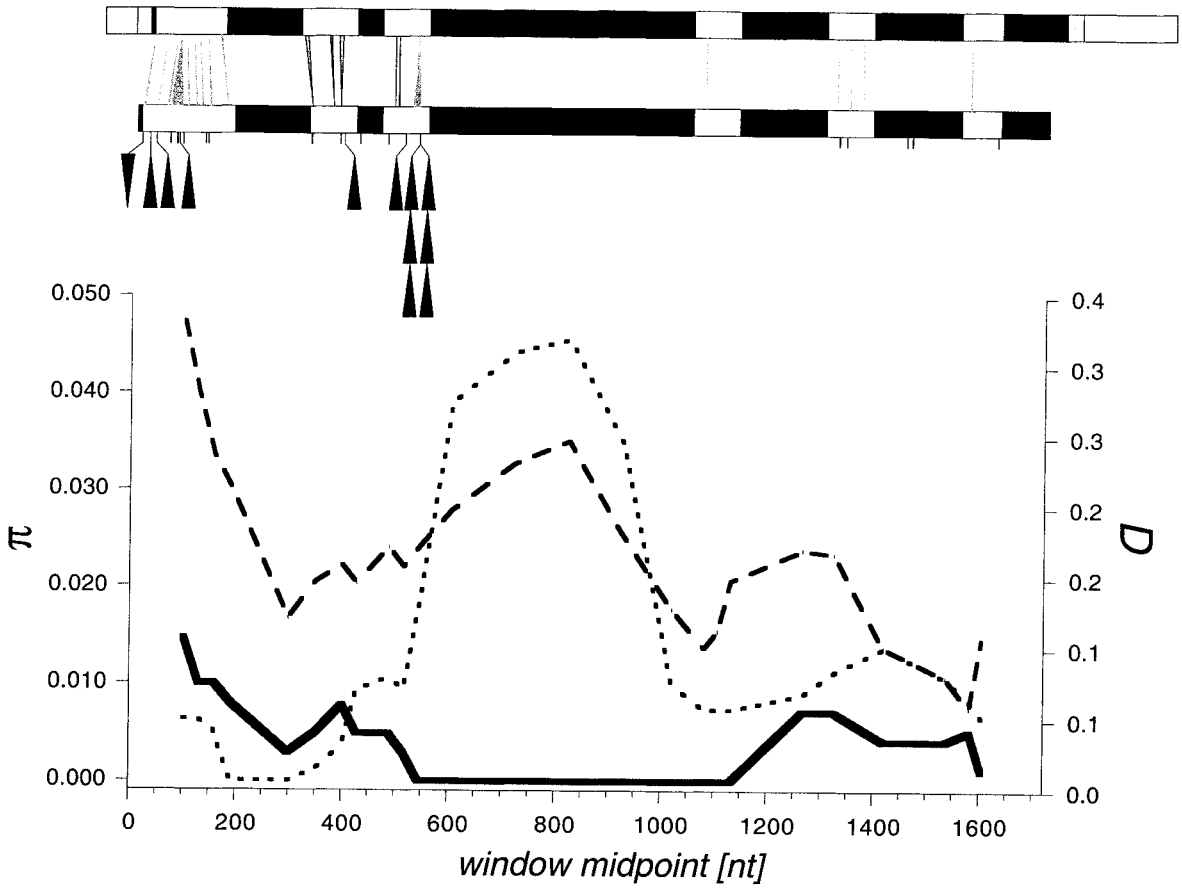

FIG. 2.—Above: Structure of the *Adh* gene of *Arabidopsis lyrata* in comparison to that of *Arabidopsis thaliana.* Black boxes represent exons, and white ones represent introns. Arrows show positions of indel polymorphism. Below: Estimated nucleotide diversity (π) of *A. lyrata* (black) and *A. thaliana* (gray) and divergence *D* of the two species (dashed) (Jukes-Cantor corrected). Sliding window of 100 nt with 25-nt steps.

comparisons between other species, sequences were obtained from GenBank as indicated.

The minimum number of exchanges required to explain the phylogenetic relationship among the sampled sequences (assuming no recurring mutation) was determined (Hudson and Kaplan 1985) using the data for which sequence information on full-length alleles was available.

A maximum-likelihood phylogenetic tree was constructed of *Adh* sequences using PAUP*, version 4.02Ba (Swofford 1999). The method was heuristic search (stepwise addition ''as-is'' option and TBR branch swapping) using the HKY85 substitution model.

Results

Arabidopsis lyrata has a single *Adh* locus that is homologous to the single *Adh* locus in *A. thaliana* (Chang and Meyerowitz 1986). The locus is genetically linked in *A. lyrata* to the microsatellite marker *ng111* (de Haan et al., personal communication), which is consistent with the *A. thaliana* linkage map (http://genome-stanford.edu/Arabidopsis/). It consists of seven exons (six introns) of similar size to those in *A. thaliana*. Twenty indels distinguish it from the *A. thaliana Adh*; the approximate locations and sizes of these indels are seen in figure 2. There are two larger insertions relative to *A. thaliana* (25 bp at site 1099 in intron 1 and 16 bp at site 1543 in intron 3), but the remaining indels distinguishing *A. lyrata* and *A. thaliana* are small (1–10 bp).

Nucleotide Polymorphism in the *Adh* Region

Among the 11 different ssp. *lyrata*/ssp. *petraea* sequence types shown in table 1, we found eight different indel polymorphisms. At site 1543, the sequences had a run of T repeats ranging in size from 8 to 15. The variable sizes of the runs of T-repeats are shown as nested insertions in figure 2. There were 15 sites segregating alternative nucleotides. Estimated nucleotide diversities within the populations of ssp. *petraea* and *ssp. lyrata* were 0 in the Icelandic population (but the two sequences differed by a 3-nt insertion), 0.0006 (SD = 0.0003/0.0006 without and including the stochastic variance) in Karhumäki, and 0.0023 (SD = 0.003/0.0013) in the Indiana Dunes population (table 3, diagonal). The overall specieswide estimate of nucleotide diversity for *A. lyrata* was 0.0038 (unweighted average over the three populations, SD = 0.0003/0.0021). In the *A. thaliana* worldwide sequences surveyed by Innan et al. (1996), the overall specieswide nucleotide diversity, estimated at

Table 3
Variation Within and Between Populations of *Arabidopsis lyrata* ssp. *lyrata* and ssp. *petraea* and Divergence from *Arabidopsis thaliana*

A. lyrata	Iceland	Karhumäki	Indiana Dunes	*A. thaliana*
Iceland	0	4	6	20
Karhumäki	0.80	0.0006	3	20
Indiana	0.67	0.61	0.0023	20
A. thaliana	0.150 (0.014)	0.152 (0.014)	0.150 (0.012)	0.0069

NOTE.—The diagonal shows substitution polymorphism (π) within the populations. Cells above the diagonal show the absolute numbers of independent insertions/deletions between consensus sequences of the populations. Cells below the diagonal show the proportion of nucleotide variation due to differences between the two ssp. *petraea* and ssp. *lyrata* populations (g_{st}). The last column shows the numbers of indels distinguishing the population consensus sequences from *A. thaliana.* The last row shows the divergence of the different populations from *A. thaliana* at synonymous (and nonsynonymous) sites. The lowest diagonal cell shows the specieswide estimate of nucleotide diversity for *A. thaliana,* based on sequences by Innan et al. (1996), analyzed over the same set of nucleotide sites as the *A. lyrata* sequences. All data relating to *A. thaliana* are shown in bold.

these same sites, was 0.0069 (SD = 0.001/0.0037). The difference in polymorphism may not be significant, but certainly *A. thaliana* is no less variable than *A. lyrata.*

As is evident in figure 2, the 15 substitutional polymorphisms were concentrated in the 5′ part of the gene, mainly in the first intron. Three polymorphisms were in exons, and two of them were replacement polymorphisms. In exon 3, the A/C polymorphism leads to a glutamine in ssp. *petraea* (and *A. thaliana*), whereas the North American ssp. *lyrata* has a histidine at that site. At site 2444, there is a C/T (leucine/phenylalanine) polymorphism in ssp. *lyrata,* whereas *A. thaliana* and ssp. *petraea* both have a phenylalanine at that site. Notice that the distribution of variation between the different areas of the gene is strikingly different between the two species: the numbers of segregating sites in the 3′ part, the middle part (from the beginning of exon 4 to the end of exon 5), and 5′ part are 10, 0, 5 and 19, 14, 7 in *A. lyrata* and *A. thaliana,* respectively ($\chi^2_{(2)} = 7.26$, $P < 0.05$).

All sites in *A. lyrata* had two alternative nucleotides, and none were unique. However, this was not true of indel variants. Theoretical expectations (Kimura 1983; Tajima 1989) and empirical experience indicate that the failure to observe unique polymorphisms is unusual. Such a deficiency of rare nucleotide substitution polymorphisms is reflected as positive values of Tajima's (1989) *D* statistic. This test assumes a sample from a single finite Wright-Fisher population with mutation to selectively neutral polymorphisms at equilibrium with genetic drift. Since there is significant interpopulation site frequency differentiation (see below), we tested the frequency spectra of the Indiana Dunes and Karhumäki populations separately. The probability of observing a Tajima's *D* of 2.31 in the Indiana Dunes sample of 30 alleles with eight segregating sites was determined by simulation (Braverman et al. 1995) to be less than 0.005, rejecting the neutral hypothesis. Even if we had missed one unique site in the SSCP analysis, Tajima's *D* would still be 1.84 ($P < 0.02$). Tajima's *D* for the small Karhumäki sample was 0.71, consistent with the distribution expected under neutral theory.

Differentiation Between Populations

Table 3 shows that there were four to six indel differences between consensus sequences, and two were fixed between ssp. *petraea* and ssp. *lyrata.* All populations were weighted equally to estimate the proportion of nucleotide variation due to differentiation between populations. Table 1 shows that there were five fixed differences between ssp. *petraea* and ssp. *lyrata.* The g_{st} values were high between all individual populations (0.72), between ssp. *lyrata* and individual ssp. *petraea* populations (0.61 for Karhumäki and 0.67 for Iceland), and between the combined ssp. *petraea* populations and ssp. *lyrata* (0.49). The Icelandic and Russian populations had two fixed differences. This low absolute level of divergence gave rise to a high differentiation estimate ($g_{st} = 0.80$), as these two populations had low estimates of within-population nucleotide diversity (see Charlesworth 1998).

Divergence from *A. thaliana*

There were 20 indel differences between *A. thaliana* and consensus sequences of *A. lyrata.* Nucleotide divergence between the species was 15% at silent sites and 1% at nonsynonymous sites (table 3). As indicated in figure 2, both indel and substitutional divergences were highest in the 5′ region of the gene.

Comparison of Polymorphism and Divergence

The sliding-window analysis of silent-site divergence and polymorphism (with a window length of 100 nt and a step size of 25 nt) in figure 2 shows that polymorphism within *A. lyrata* and divergence from *A. thaliana* are both highest in the 5′ part of the gene. There was no variation in *A. lyrata* in the middle part of the gene. We tested whether the divergence and polymorphism were consistent with neutral expectation using the HKA test (Hudson, Kreitman, and Aguadé 1987). Comparison of divergence and segregating sites between the 5′ part, the middle part (from the beginning of exon 4 to the end of exon 5), and the 3′ part of the gene did not deviate from neutral expectation ($\chi^2_{(2)} = 2.32$). Figure 2 also shows the variation in *A. thaliana* based on

the sequences by Miyashita, Innan, and Terauchi (1996) and Innan et al. (1996); the well-known peak of polymorphism is seen in exon 4.

Discussion

Outcrossing Versus Selfing Does Not Alone Account for Differences in Polymorphism Within Populations

Outcrossing versus selfing is known to be an important determinant of distribution of variation in plant populations, and selfing alone is expected to reduce the effective population size to 50% of the random-mating value. The North American ssp. *lyrata* had higher variation within populations (0.0023) at *Adh* than Bergelson et al. (1998) found at *Adh* of *A. thaliana* in North American and European populations (0.00052) based on four-cutter restriction fragment length polymorphisms (RFLPs). They found even lower values for the two other loci they studied in *A. thaliana* (average 0.0004). The small Karhumäki and Iceland ssp. *petraea* samples had quite low variation (table 3), perhaps due to the large variance expected (Hudson 1990). However, earlier studies with larger sample sizes have shown that for isozymes and microsatellites, the Karhumäki population was as variable as the Indiana Dunes population (Van Treuren et al. 1997). The Iceland population has high expected heterozygosity at isozyme loci; the average at nine loci for two populations was 0.32 based on data of Schierup (1998).

Hamrick and Godt (1990) showed in their review of isozyme literature that the within-population expected heterozygosity of selfers (0.074) is on average about 50% of that of outcrossers (0.136). However, the specieswide variation of selfing species (0.124) is not as much lower than that of wind-pollinated outcrossers (0.165). At the DNA level, such comparisons are scarce. In comparing predominantly outcrossing and predominantly selfing species of *Mimulus,* Awadalla and Ritland (1997) found a difference, but not as large as that for isozymes. For British *A. thaliana* populations, Abbott and Gomes (1989) found that the average expected within-population heterozygosity at seven isozyme loci was 0.057, slightly lower than the average for selfing species (Hamrick and Godt 1990). Earlier allozyme data have shown that many Scandinavian *A. thaliana* populations are monomorphic (Kuittinen, Mattila, and Savolainen 1997). Japanese populations and most Scandinavian populations have no variation at microsatellite loci (Todokoro, Terauchi, and Kawano 1996; Kuittinen, Mattila, and Savolainen 1997). Overall, these comparisons show that at both enzyme loci and microsatellites, *A. thaliana* seems to have lower within-population variation than the average selfing species. Thus, the effect of the selfing mating system alone is not sufficient to account for the low level of variation. At the nucleotide level, the only available comparison is *Leavenworthia,* for which no sequence variation was found within selfing populations at the *Adh* gene (Liu, Zhang, and Charlesworth 1998). Note also that the level of polymorphism in *A. thaliana* is quite variable between populations for isozymes, microsatellites, and sequence variation, as has been found for many selfing species (Schoen and Brown 1991).

Specieswide Polymorphism

The specieswide nucleotide polymorphism of *A. lyrata* could be estimated. Combining the North American and European populations resulted in estimated nucleotide diversity of 0.0038. The self-incompatible (Miyashita, personal communication) *Arabis gemmifera* had an estimate of about 0.008 for the same region of the gene (Miyashita et al. 1998). The overall estimate for *A. thaliana* for the same set of nucleotides was 0.0069 (table 3). Compared with the two outcrossing relatives, specieswide diversity is not reduced in *A. thaliana.* Other available data on specieswide *Adh* variation are from the outcrossing wild pearl millet, with nucleotide polymorphism of 0.0036 (measured by an estimate of θ) (Gaut and Clegg 1993*b*), and the cultivated maize, for which the estimate was 0.019 (Gaut and Clegg 1993*a*). The estimates for outcrossing *Leavenworthia* species are even higher, but they include indel variation (Liu, Zhang, and Charlesworth 1998). Compared with many other species, *A. lyrata* seems to have rather low nucleotide diversity, but we do not yet know whether this will hold for other loci as well.

Polymorphism in *A. lyrata* and Divergence from *A. thaliana*

Comparison of patterns of polymorphism and divergence can be used to detect the impact of natural selection (Hudson, Kreitman, and Aguadé 1987). Figure 2 shows the comparison for *A. lyrata* (based on the available 36 sequences) and *A. thaliana.* The middle area of the gene in *A. lyrata* had no variability, but the level of divergence was also relatively low in this area. The difference in polymorphism and divergence between different areas of the gene was not statistically significant. Based on these data, it is not necessary to invoke selective sweeps or other variability-reducing mechanisms in this area. The *A. thaliana* worldwide polymorphism has a different pattern, as there is a high peak of polymorphism in exon 4. This has been interpreted to be due to balancing selection by Hanfstingl et al. (1994). Innan et al. (1996) showed that the pattern could be well explained by the existence of two quite divergent haplotypes, found in different ecotypes around the world, and admixture and recombination between the two.

Pattern of Polymorphism

Arabidopsis lyrata exhibited 15 variable nucleotide sites, none of which were singletons. Tajima's *D* for the Indiana Dunes population (nucleotide substitutions) was 2.31 ($P < 0.005$). This result may be interpreted as evidence of balancing selection on this polymorphism or other nearby sites or as a result of a recent bottleneck (Tajima 1989). For the Karhumäki population, there was no deviation from neutral expectation. In contrast, the *Adh* locus of the self-incompatible *A. gemmifera* gave

rise to a negative Tajima's *D* (Miyashita, Innan, and Terauchi 1996). The *Adh* locus in *A. thaliana* had a large number of rare variants, but Tajima's test used by Innan et al. (1996) failed to find significance. The population samples of Bergelson et al. (1998) for *Adh* and the two other genes they studied would also yield nonsignificant Tajima's *D* values, as we evaluated from their tables 4–6. On the other hand, other *A. thaliana* genes (CAL, PI, AP3) have significant excesses of rare variants, as shown by negative values for Tajima's *D* (Miyashita, Innan, and Terauchi 1996; Purugganan and Suddith 1998, 1999). However, these surveys of *A. thaliana* (and *A. gemmifera*) were not based on population samples, but on collections of single representatives of individual populations. As the *A. thaliana* populations are quite diverged (see below), sampling of one sequence per population alone could generate such a pattern.

We found only two replacement polymorphisms, whereas there were six in the corresponding region of *A. thaliana*. Purugganan and Suddith (1998, 1999) found an excess of replacement polymorphisms (compared with divergence) in several *A. thaliana* genes, which they interpreted as evidence of fixation of slightly deleterious mutations in populations due to drift. Cummings and Clegg (1998) gave a similar interpretation for the excess replacement polymorphisms in *Adh* of wild barley. Such fixation should give rise to heterotic crosses between ecotypes. Some evidence for heterosis between crosses of *Arabidopsis* ecotypes is provided by the work of Pederson (1968) and Griffing and Zsiros (1971). The fixation of deleterious mutations may be a general explanation for heterosis in crosses between ecotypes, despite alternative overdominance interpretations (Mitchell-Olds 1995).

Variation Between Populations

Earlier work using isozymes and microsatellites has found that the *A. lyrata* populations show strong geographic differentiation and populations within Europe are quite diverged (Jonsell, Kustås, and Nordal 1995; Van Treuren et al. 1997). The *Adh* sequence data are consistent with this picture. Measured as g_{st}, about two thirds of the variation is between the North American and the European populations. This pattern of divergence suggests long-term isolation between the European ssp. *petraea* and the North American ssp. *lyrata.* Innan et al. (1996) pointed out that there is no overall geographic structure in *A. thaliana* populations. We analyzed separately their eight European and four North American sequences of *A. thaliana.* Based on this set of sequences, the g_{st} value was 0; thus, there was no geographic Europe/North America structure in *A. thaliana.* Furthermore, the dendrograms generated using variation at different loci are not congruent with one another or with geography (Innan et al. 1996; Purugganan and Suddith 1999). The same issue can be addressed based on the data of Bergelson et al. (1998). They reported a high divergence between their 11 populations, 5 from Europe, 5 from North America, and 1 from Asia, with an overall estimated g_{st} value of 0.63 for *Adh* and similar values for the two other genes they studied. The g_{st} estimate for the Europe/North America subdivision based on their data was 0.05 for the three genes and 0.06 for *Adh* (Bergelson et al. 1998). The pattern of high local differentiation is supported by the isozyme data of Abbott and Gomes (1989), who found that 62% of the isozyme variation was between the British populations. As pointed out by previous authors, the *A. thaliana Adh* sequence variation patterns suggest very rapid colonization (Innan et al. 1996).

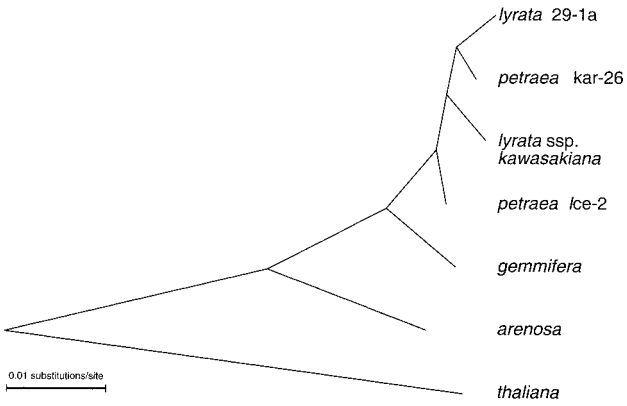

FIG. 3.—Maximum-likelihood tree of the *Adh* sequences of *Arabidopsis lyrata* ssp. *lyrata,* ssp. *petraea,* and ssp. *kawasakiana, Arabidopsis thaliana, Arabidopsis arenosa,* and *Arabis gemmifera.*

Divergence from *A. thaliana* and Other Relatives

These results on divergence at *Adh* agree with the view that *A. lyrata* is a close relative of *A. thaliana* (Price, Palmer, and Al-Shehbaz 1994; O'Kane and Al-Shehbaz 1997; Koch, Bishop, and Mitchell-Olds 1999). The level of divergence at *Adh* was slightly more than 1% for nonsynonymous substitutions and about 15% for synonymous substitutions. The same level of divergence between *A. thaliana* and *A. lyrata* of about 1%–2% of nonsynonymous sites and 15% at synonymous sites was also found for other genes, CAL, AP3, and PISTILLATA, by Purugganan and Suddith (1998, 1999). This level of divergence between the two species corresponds to about twice the level of differentiation between the sibling species *Drosophila melanogaster* and *Drosophila simulans* (Moriyama and Powell 1996).

Sequences are available for some close relatives of *A. lyrata.* The sequence of *A. lyrata* ssp. *kawasakiana* (tetraploid) is available from Miyashita et al. (1998) (GenBank accession number AB015501), as well as the sequence of the Japanese *A. gemmifera* (D63459) and a sequence from the European *Arabidopsis suecica* (GenBank accession number AB015507). A maximum-likelihood tree of the relatives is shown in figure 3. The tree was rooted at one half of the average of the distances between *A. thaliana* and the other sequences. The sequence labeled *arenosa* in the figure is derived from *A. suecica,* a hybrid between *A. thaliana* and *Arabidopsis* (*Cardaminopsis) arenosa.* Miyashita et al. (1998) concluded that the sequence represents the *A. arenosa* parent of the allotetraploid. This was confirmed by our

preliminary sequencing of *Adh* from *A. arenosa* (data not shown). The tree suggests that *A. arenosa* is more closely related to *A. lyrata* than to *A. thaliana.* (We have conducted initial studies on several *A. arenosa* populations, and all were found to be tetraploid and thus not easily amenable to population genetic studies.)

The earlier isozyme work suggested that the Russian Karhumäki population of ssp. *petraea* is equally diverged from the Swedish and North American populations (Van Treuren et al. 1997). The North American and European subspecies are interfertile, and an F_2 can be produced without difficulty (unpublished data). In our tree, *A. gemmifera* seems to be more diverged from ssp. *kawasakiana* than the *A. lyrata* ssp. *lyrata.* In fact, in the classification of O'Kane and Al-Shehbaz (1997), *A. gemmifera* is listed as a subspecies of *Arabidopsis halleri,* for which we do not yet have an *Adh* sequence available.

Conclusions

Arabidopsis lyrata and *A. thaliana* have very different patterns of variation, which are partly attributable to the differences in mating system. However, the variation within *A. thaliana* populations is reduced by more than the theoretically expected half compared with the outcrossing species. Furthermore, the reduction is larger than has been empirically found in comparisons of selfing and outcrossing species (Hamrick and Godt 1996). Other factors reducing variation must thus be invoked.

Additional possibilities include population subdivision and bottleneck effects, extinctions and recolonizations in a metapopulation setting, background selection, or hitchhiking.

Bottlenecks should reduce variation within populations due to drift, but overall neutral variation, compared with a panmictic situation, would not be reduced. It would be expected that populations would vary in level of variation. In a metapopulation setting, the extinctions and recolonizations would cause a large decline in both within-population and total nucleotide diversity (Pannell and Charlesworth 1999). There would likely also be increased differentiation between the populations in the metapopulation setting, but the effects depend on the biological details of migration and colonization (Wade and McCauley 1988; Barton and Whitlock 1997; Ingvarsson 1997; Pannell and Charlesworth 1999).

Many features of *A. thaliana* seem to fit with the bottleneck hypothesis. The overall level of variation is not reduced compared with outcrossing relatives. The within-population diversity is highly reduced compared with outcrossing related species. The level of diversity within populations is low and variable. The pattern of geographic differentiation, e.g, no overall difference between Europe and North America but high divergence within both areas, also agrees with this prediction. In the data of Bergelson et al. (1998), populations rather close to each were fixed for different haplotypes. Marginal populations in Scandinavia and Japan are devoid of variation.

The weedy characteristics of the species suggest that extinctions also would occur, which should lead to reduced specieswide variation (Pannell and Charlesworth 1999). Liu, Zhang, and Charlesworth (1998) also failed to find such a reduction in *Leavenworthia.* More extensive sampling of relative species may be needed for adequate comparisons of specieswide diversities.

Background selection due to deleterious alleles could also cause a major reduction in variation within populations (Charlesworth, Morgan, and Charlesworth 1993). It is not clear how background selection operates between populations. Charlesworth, Nordborg, and Charlesworth (1997) showed that background selection could reduce both the within-population and between-populations components of variation (their fig. 4). Liu, Zhang, and Charlesworth (1998, p. 299) suggested that background selection reduces only slightly the total nucleotide diversity. In most cases, background selection will increase F_{st}, but in very small populations with close linkage to deleterious genes, this will not always be the case (Pamilo, Pálsson, and Savolainen 1999). At the species level, *A. thaliana* is no less variable than the outcrossing relative, and background selection does not seem to have had an effect. (It is not clear what the correct sampling strategy should be to analyze variation in such a species.) The excess of rare alleles found as singletons in some populations of *A. thaliana* could also be caused by the sampling strategy of one individual per diverged ecotype, rather than genetic mechanisms reducing variation within populations.

The outcrossing populations of *A. lyrata* may be more stable and better fit the assumptions of population genetics models and the assumptions of tests of hypotheses about the role of natural selection.

Acknowledgments

We thank Oleg Kuznetsov for help in collecting in Karhumäki, Auli Karhu for help with sequencing ssp. *petraea,* and Laura Rose for help with the phylogeny in figure 3. O.S. acknowledges financial support from the Finnish Department of Education Center for International Mobility for support of the visit of Hélène Freville, and from the Environment and Natural Resources Research Council of Finland. C.H.L. acknowledges funding from the National Science Foundation (DEB-09548).

LITERATURE CITED

ABBOTT, R. J., and M. F. GOMES. 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Heredity **62**:411–418.

AGUADÉ, M., W. MEYERS, A. D. LONG, and C. H. LANGLEY. 1994. Single-strand conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster* X chromosome. Proc. Natl. Acad. Sci. USA **91**:4658–4662.

ALLARD, R. W., S. K. JAIN, and P. WORKMAN. 1968. The genetics of inbreeding species. Adv. Genet. **14**:55–131.

AWADALLA, P., and K. RITLAND. 1997. Microsatellite variation and evolution in the *Mimulus guttatus* species complex with

contrasting mating systems. Mol. Biol. Evol. **14**:1023–1034.

BARTON, N. H., and M. WHITLOCK. 1997. The evolution of metapopulations. Pp. 183–210 *in* I. A. HANSKI and M. E. GILPIN, eds. Metapopulation biology. Academic Press, New York.

BERGELSON, J., E. STAHL, S. DUDEK, and M. KREITMAN. 1998. Genetic variation within and among populations of *Arabidopsis thaliana.* Genetics **148**:1311–1323.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**:783–796.

CHANG, C., and E. M. MEYEROWITZ. 1986. Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA **83**:1408–1412.

CHARLESWORTH, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15**:538–543.

CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

CHARLESWORTH, B., M. NORDBORG, and D. CHARLESWORTH. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70**:155–174.

CHARLESWORTH, D., and B. CHARLESWORTH. 1995. Quantitative genetics of plants: the effect of the breeding system on genetic variability. Evolution **49**:911–920.

CHARLESWORTH, D., B. CHARLESWORTH, and M. T. MORGAN. 1995. The pattern of neutral molecular variation under the background selection model. Genetics **141**:1619–1632.

CUMMINGS, M. P., and M. T. CLEGG. 1998. Nucleotide sequence diversity at the alcoholdehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc. Natl. Acad. Sci. USA **95**:5637–5642.

DOYLE, J. J., and J. L. DOYLE. 1990. Isolation of plant DNA from fresh tissue. BRL Focus **12**:13–15.

GAUT, B. S., and M. T. CLEGG. 1993*a.* Molecular evolution of the *Adh1* locus in the genus *Zea.* Proc. Natl. Acad. Sci. USA **90**:5095–5099.

———. 1993*b.* Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135**:1091–1097.

GRIFFING, B., and E. ZSIROS. 1971. Heterosis associated with genotype-environment interactions. Genetics **68**:443–455.

HAMRICK, J. L., and M. J. GODT. 1990. Allozyme diversity in plant species. Pp. 43–63 *in* A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER, and B. S. WEIR, eds. Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland, Mass.

———. 1996. Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B Biol. Sci. **351**:1291–1298.

HANFSTINGL, U., A. BERRY, E. A. KELLOGG, J. T. COSTA, W. RUEDIGER, and F. M. AUSUBEL. 1994. Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics **138**:811–828.

HOLSINGER, K. E., and R. J. MASON-GAMER. 1996. Hierarchical analysis of nucleotide diversity in geographically structured populations. Genetics **142**:629–639.

HOPKINS, M. 1937. Arabis in eastern and central North America. Rhodora **39**:63–98, 106–148, 155–186.

HUDSON, R. R. 1990. Gene genealogies and the coalescent process. Pp. 1–44 *in* D. FUTUYMA and J. ANTONOVICS, eds. Oxford surveys in evolutionary biology. Vol. 7. Oxford University Press, Oxford, England.

HUDSON, R. R., and N. L. KAPLAN. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**:147–164.

HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

HULTÉN, E., and M. FRIES. 1986. Atlas of North European vascular plants north of tropic of cancer. Koeltz Scientific Books, Königstein, Germany.

INGVARSSON, P. K. 1997. The effect of delayed population growth on the genetic differentiation of local populations subject to frequent extinctions and recolonizations. Evolution **51**:29–35.

INNAN, H., F. TAJIMA, R. TERAUCHI, and N. T. MIYASHITA. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana.* Genetics **143**:1761–1770.

JALAS, J., and J. SUOMINEN. 1994. Atlas Florae Europaea. Distribution of vascular plants in Europe. 10. Cruciferae (Sisymbrium to Aubrieta). Helsinki University Printing House, Helsinki, Finland.

JONES, B. M. G. 1963. Experimental taxonomy of the genus *Arabis.* Ph.D. thesis, University of Leicester, England.

JONSELL, B., K. KUSTÅS, and I. NORDAL. 1995. Genetic variation in *Arabis petraea*, a disjunct species in northern Europe. Ecography **18**:321–332.

KAPLAN, N. L., R. R. HUDSON, and C. H. LANGLEY. 1989. The 'hitchhiking' effect revisited. Genetics **123**:887–899.

KÄRKKÄINEN, K., H. KUITTINEN, R. VAN TREUREN, C. VOGL, S. OIKARINEN, and O. SAVOLAINEN. 1999. Genetic basis of inbreeding depression in *Arabis petraea.* Evolution **53**: 1354–1365.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, Mass.

KOCH, M., J. BISHOP, and T. MITCHELL-OLDS. 1999. Molecular systematics and evolution of *Arabidopsis* and *Arabis.* Plant Biol. **1**:529–537.

KUITTINEN, H., A. MATTILA, and O. SAVOLAINEN. 1997. Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana.* Heredity **79**: 144–152.

LIU, F., D. CHARLESWORTH, and M. KREITMAN. 1999. The effect of mating system differences on nucleotide diversity at the phosphoglucose isomerase locus in the plant genus Leavenworthia. Genetics **151**:343–357.

LIU, F., L. ZHANG, and D. CHARLESWORTH. 1998. Genetic diversity in *Leavenworthia* populations with different inbreeding levels. Proc. R. Soc. Lond. B Biol. Sci. **265**:293–301.

MEUSEL, H., E. JAEGER, and E. WEINERT. 1965. Vergleichende Chorologie der zentral-europäischen Flora, Vol. 2. Maps. Gustav Fischer Verlag, Jena, Germany.

MITCHELL-OLDS, T. 1995. Interval mapping of viability loci causing heterosis in *Arabidopsis.* Genetics **140**:1105–1109.

MIYASHITA, N. T., H. INNAN, and R. TERAUCHI. 1996. Intra- and interspecific variation of the Alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana.* Mol. Biol. Evol. **13**:433–436.

MIYASHITA, N. T., A. KAWABE, H. INNAN, and R. TERAUCHI. 1998. Intra- and interspecific DNA variation and codon bias of alcohol dehydrogenase (Adh) locus in *Arabis* and *Arabidopsis* species. Mol. Biol. Evol. **15**:1420–1429.

MORIYAMA, E., and J. R. POWELL. 1996. Intraspecific nuclear DNA variation in *Drosophila.* Mol. Biol. Evol. **13**:261–277.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NORDBORG, M. 1997. Structured coalescent processes on different time scales. Genetics **146**:1501–1514.

NORDBORG, M., B. CHARLESWORTH, and D. CHARLESWORTH. 1996. Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. R. Soc. Lond. B Biol. Sci. **263**:1033–1039.

O'KANE, S. L., and I. A. AL-SHEHBAZ. 1997. A synopsis of *Arabidopsis* (Brassicaceae). Novon **7**:323–327.

PAMILO, P., S. PÁLSSON, and O. SAVOLAINEN. 1999. Deleterious mutations can reduce differentiation in small, subdivided populations. Hereditas **130**:257–264.

PANNELL, J. R., and B. CHARLESWORTH. 1999. Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution **53**:664–676.

PEDERSON, D. G. 1968. Environmental stress, heterozygote advantage and genotype-environment interaction in *Arabidopsis.* Heredity **23**:127–138.

POLLAK, E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117**:353–360.

PRICE, R. A., J. D. PALMER, and I. A. AL-SHEHBAZ. 1994. Systematic relationships of *Arabidopsis.* Pp. 7–19 *in* E. M. MEYEROWITZ and C. SOMERVILLE, eds. Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

PURUGGANAN, M. D., and J. SUDDITH. 1998. Molecular population genetics of the *Arabidopsis* CAULIFLOWER regulatory gene: non-neutral evolution and wild variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95**: 8130–8134.

———. 1999. Molecular population genetics of floral homeotic loci: departures from the equlibrium-neutral model at the APETALA3 and PISTILLATA genes of *Arabidopsis thaliana.* Genetics **151**:839–848.

ROGERS, S. O., and A. J. BENDICH. 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol. Biol. **5**:69–76.

ROLLINS, R. C. 1941. A monographic study of *Arabis* in western North America. Rhodora **43**:289–325, 348–411, 425–481.

ROZAS, J., and R. ROZAS. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**:174–175.

SCHIERUP, M. H. 1998. The effect of enzyme heterozygosity on growth in a strictly outcrossing species, the self-incompatible *Arabis petraea* (Brassicaceae). Hereditas **128**:21–31.

SCHOEN, D. J., and A. H. D. BROWN. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl. Acad. Sci. USA **88**:4494–4497.

STEPHAN, W., and C. H. LANGLEY. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. Genetics **150**:1585–1593.

SWOFFORD, D. L. 1999. PAUP*: Phylogenetic analysis using parsimony (and other methods). Version 4.02 Ba. Sinauer, Sunderland, Mass.

TAJIMA, F. 1989. Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and wight matrix choice. Nucleic Acids Res. **22**:4673–4680.

TODOKORO, S., R. TERAUCHI, and S. KAWANO. 1996. Microsatellite polymorphism in natural populations of *Arabidopsis thaliana* in Japan. Jpn. J. Genet. **70**:543–554.

VAN TREUREN, R., H. KUITTINEN, K. KÄRKKÄINEN, E. BAENA-GONZALEZ, and O. SAVOLAINEN. 1997. Evolution of microsatellites in *Arabis petraea* and *A. lyrata,* outcrossing relatives of *Arabidopsis thaliana.* Mol. Biol. Evol. **14**:220–229.

WADE, M. J., and D. E. MCCAULEY. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution **42**:995–1005.

WOLFGANG STEPHAN, reviewing editor

Accepted December 21, 1999