Molecular Ecology (2005) **14**, 137–148 doi: 10.1111/j.1365-294X.2004.02359.x

Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe

HANS K. STENØIEN,* CHARLES B. FENSTER,† ANNI TONTERI‡§ and OUTI SAVOLAINEN‡
*Uppsala University, Department of Ecology and Evolution, Evolutionary Biology Centre, SE-752 36 Uppsala, Sweden, †University of Maryland, Department of Biology, College Park, Maryland 20742, USA, ‡University of Oulu, Department of Biology, PO Box 3000, Fin-90014 Oulu, Finland

Abstract

Ten populations of the model plant *Arabidopsis thaliana* were collected along a north–south gradient in Norway and screened for microsatellite polymorphisms in 25 loci and variability in quantitative traits. Overall, the average levels of genetic diversity were found to be relatively high in these populations, compared to previously published surveys of within population variability. Six of the populations were polymorphic at microsatellite loci, resulting in an overall proportion of polymorphic loci of 18%, and a relatively high gene diversity for a selfing species (H_E = 0.06). Of the overall variability, 12% was found within populations. Two of six polymorphic populations contained heterozygous individuals. Both F_{ST} and phylogenetic analyses showed no correlation between geographical and genetic distances. Haplotypic diversity patterns suggested postglacial colonization of Scandinavia from a number of different sources. Heritable variation was observed for many of the studied quantitative traits, with all populations showing variability in at least some traits, even populations with no microsatellite variability. There was a positive association between variability in quantitative traits and microsatellites within populations. Several quantitative traits exhibited Q_{ST} values significantly less than F_{ST}, suggesting that selection may be acting to retard differentiation for these traits.

Keywords: F_{ST}, microsatellites, migration, population differentiation, Q_{ST}, selfing

Received 11 June 2004; revision received 27 August 2004; accepted 27 August 2004

Introduction

Marker gene variability is often found to be very low within field-collected populations, as well as within stock centre accessions in the model plant *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) (Abbott & Gomes 1989; Todokoro *et al*. 1995; Kuittinen *et al*. 1997; Bergelson *et al*. 1998; Loridon *et al*. 1998; Breyne *et al*. 1999; Vander Zwan *et al*. 2000; Clauss *et al*. 2002; Kuittinen *et al*. 2002; Le Corré *et al*. 2002). The observed variability levels have been explained by severe population subdivision in this highly selfing annual species, consisting of populations subjected to repeated colonization and extinction events (Clauss *et al*. 2002). Selfing is known to reduce the effective recombination rate, and inbreeding taxa such as *A. thaliana* are thought to be particularly vulnerable to the effects of hitchhiking or background selection (Nordborg *et al*. 1996; Nordborg & Donnelly 1997; Nordborg 2000), as well as drift associated metapopulation dynamics (Ingvarsson 2002). Furthermore, worldwide levels of variability have been found to be high, and this variability has mostly been found among populations (Todokoro *et al*. 1995; Innan *et al*. 1997; Bergelson *et al*. 1998; Breyne *et al*. 1999; Miyashita *et al*. 1999). Most studies have not found a correlation between the geographical origin of worldwide accessions or populations, and the levels of genetic distance, which is possibly explained by a recent and rapid spread worldwide (e.g. Todokoro *et al*. 1995; Bergelson *et al*. 1998; Miyashita *et al*. 1999; Provan & Campanella 2003), and metapopulation dynamics (Wakeley & Aliacar 2001). However, Sharbel *et al*. (2000), using a large number of AFLP (amplified fragment length polymorphisms) markers, found low but significant isolation by distance among *A. thaliana* accessions from

Correspondence: H. K. Stenøien. Fax: + 46 18 55 34 19; E-mail: hans.stenoien@ebc.uu.se
§Present address: University of Turku, Department of Biology, Fin-20014 Turku, Finland

Eurasia and southern Europe, and concluded that after the last ice age northern European populations may have been colonized by populations originating from southwestern Europe.

Here we quantify patterns of genetic variability in Norwegian populations of *A. thaliana*, chosen for several reasons. Scandinavian populations of *A. thaliana* represent the periphery of this species' distribution range. Such marginal populations may experience more extreme ecological conditions than populations in the central parts of the distribution, representing the climatic range boundaries of the species (Hoffmann 2002). This could in turn affect levels of genetic variability and structuring in these populations, either by depleting variability in an extreme environment, or increasing variability in environments with strong and rapidly changing directional selection (Vavilov 1926; Lewontin 1974; Schwartz *et al*. 2003). Also, few studies of *A. thaliana* have employed field-collected populations, especially populations with no major human influence. Natural Norwegian populations may be found in many different vegetation types, some of which have relatively little cultural impact, e.g. coastal rocks (Fremstad 1997).

Therefore, we focused on quantifying levels of genetic variability and structure in natural or seminatural populations of *A. thaliana* within a limited range of approximately 900 km. Populations were collected along a north–south gradient in Norway, representing the northernmost European distribution range of this species. All populations within this range should be the result of postglacial colonization. We sampled populations found in different sets of vegetation types in Norway described as natural for this plant by Fremstad (1997). Patterns of population genetic variation for both single locus markers and quantitative traits were quantified. Therefore, we are able to quantify whether patterns of variability are the same at local and regional geographical scales compared to the global or continental scale often employed in population genetic studies of this species.

There are several predictions concerning the expected patterns of genetic variability in these populations. First, if Scandinavian populations have been formed by migrants originating from southwestern parts of Europe, then one might predict that variability levels are lower overall in Norwegian populations compared to Mediterranean and western European populations, which represent regions of maximum genetic variability in this species (Hoffmann *et al*. 2003). Furthermore, if postglacial migration occurred in a stepwise fashion from south to north, then northern haplotypes should simply represent subsamples of haplotype variants found in the south, containing subsequent diversification as a result of drift, and possibly mutations and/or recombination. In this case, one should also expect northern Norwegian populations to contain less variability than southern Norwegian populations. The stepwise migration hypothesis implies that one should expect significant isolation by distance among populations, and at least some southern haplotypes should be present in northern populations. Alternatively, the peripheral populations could have arisen from a larger pool of possible immigrants and not only southwestern ones, causing no obvious isolation by distance, and not necessarily a genetic depletion towards the north. Stenøien *et al*. (2002) documented clinal variability in two quantitative traits in studied Norwegian *A. thaliana* populations, and inferred that this could be the result of natural selection. Here we compare the structuring in neutral markers with a range of quantitative traits, in order to test whether quantitative variability has been shaped by adaptive forces.

The specific questions we address are: (i) what is the pattern of neutral microsatellite variability in Norwegian *A. thaliana* populations; (ii) how does variability in several quantitative traits compare to patterns of microsatellite variability; (iii) what are the roles of migration and recombination in shaping observed genetic patterns; and (iv) whether northern Norwegian populations originated by colonization from the southern Norwegian populations.

Materials and methods

Earlier studies have quantified variability in hypocotyl responses to various light regimes in the same populations (Stenøien *et al*. 2002). Here we quantify patterns of population genetic differentiation for additional quantitative traits based on seeds collected from 104 maternal families sampled along the same latitudinal gradient (see Table 1 for detailed description of the populations, and Fig. 1 for the geographical location of the various populations in Norway). Because germination was limiting for adjacent populations 1 and 2 (situated *c*. 50 m apart), these families were combined into one population, resulting in nine populations in the analyses of quantitative variation. Seeds from a maternal family were sown into two different pots, the pots were randomized and cold stratified at 2–4 °C for 3 weeks. Pots were rotated throughout the cold stratification. Seedlings were then allowed to germinate at approximately 20 °C and 18-hour day supplemented light, again with periodic rotation of pots on the bench. At the 2–4 leaf stage, seedlings were transplanted into a randomized block design, consisting of 26 blocks (each block corresponding to a flat of approximately 20 pots), with three to six replicates per family, for a total of 104 families and 504 seedlings. The following four traits were measured: (i) day of first flowering, plants were censused twice daily for the first day of flowering; (ii) stem length, the main stem length of each plant that bolted was measured at the end of the flowering stage; (iii) trichome number, a one cm length of leaf margin of the largest leaf, centred at the middle of the leaf, was surveyed under 10 × magnification for trichome

Table 1 Habitat description, vegetation types (*sensu* Fremstad 1997), geographical location, metres above sea level (m a.s.l.), estimated census population size (*N*) and number of maternal families sampled (*n*) of the 10 *Arabidopsis thaliana* population employed in this study

Population	Locality	Habitat	Vegetation	Latitude, longitude	m a.s.l.	*N*	*n*
1	Hvaler	Rocky site 0.5 × 0.5 m, close to seashore and beach. Sun exposed. Thin sandy soil, few species: *Viola tricolor*, *Sedum acre*, *Poa* sp.	F3	59′01″, 11′01″	1	30	8
2	Hvaler	Rocky site, west facing locality, close to seashore and public beach. Thin sandy soil, few species present: *Plantago* sp., *Sedum acre*.	F3	59′01″, 11′01″	2	20	11
3	Hvaler	Flat open meadow, sun exposed, public camping area close to beach. Sandy soil with rich seashore vegetation, many herbaceous species.	W2	59′01″, 11′01″	1	200–300	7
4	Moss	Shallow sandy soil, flat rocks. Near highway, clear-cutting for expansion of highway. *Poa* sp. and bryophyte rich vegetation.	I2	59′28″, 10′42″	30	1000	10
5	Jessheim	Railroad track, small individuals with many flowers. Gravel rich, sandy soil with no other species present.	I2	60′07″, 11′10″	50	500	13
6	Stange	Picnic area near motorway, locality facing towards south. Bryophyte-rich and sandy substrate.	I2	60′30″, 11′14″	30–40	100–300	18
7	Trondheim	Farm area, tractor road. Sandy soil with few other species present.	I2	63′25″, 10′08″	50	50–100	12
8	Skatval	Edge of private garden, beneath large rocks, sandy soil close to seashore, open and sun exposed.	I2	63′28″, 10′50″	2	500–1000	13
9	Skatval	Meadow, recently idle agricultural field with rich soil, many herbaceous species.	W2	63′28″, 10′52″	10	500	22
10	Meløy	Locality a few meters below a scree in steep landscape. Birch forest, rich vegetation with *Ulmus glabra*, *Populus tremula*. Relatively closed canopy.	C2	66′53″, 13′41″	100	30	12

These are the same populations as those employed by Stenøien *et al*. (2002). Vegetation types are classified according to Fremstad (1997), and include coastal rocks (F3), dune meadows (W2), exposed railways and roadsides (I2), and birch forest (C2). Population 10 was collected in early June 1999, while populations 1–9 were collected in May 2000. The approximate area sampled ranged from less than a square meter (population 1) to approximately 400 square meter (population 4), but most populations covered less than 100 square metres.

number; and (iv) specific leaf area, the largest basal rosette leaf was collected, dried, weighed to the nearest 0.0001 mg, and measured for leaf area to the nearest 0.01 cm^2. Specific leaf area is the area per unit mass.

A nested analysis of variance (ANOVA) was conducted to partition patterns of genetic variation to between and within population components using the model

$$y_{ijkl} = \mu + L_j + \rho_k + \alpha_{l(k)} + \varepsilon_{ijkl}, \quad \text{(eqn 1)}$$

where y_{ijkl} is the response of the ith seed from the lth maternal family from the kth population and the jth block, and μ is the mean of the response. The maternal family effect (α) was considered random and nested within population (k). The block effect was also considered random. For all analyses, the interaction between block and population, and block and family nested within population were nonsignificant and consequently dropped from further analyses. To meet assumptions of ANOVA, analyses were conducted on transformed data for the following three traits: log for stem length, square root for day to first flower and trichome number. For our design, using samples of families from a selfing species, the genetic variance within populations, V_G, is estimated by the among family nested within population component of variance. From this output, levels of genetic variation for each population were computed in terms of coefficient of genetic variation

$$(CV_G = 100(V_G^{0.5}/X)), \quad \text{(eqn 2)}$$

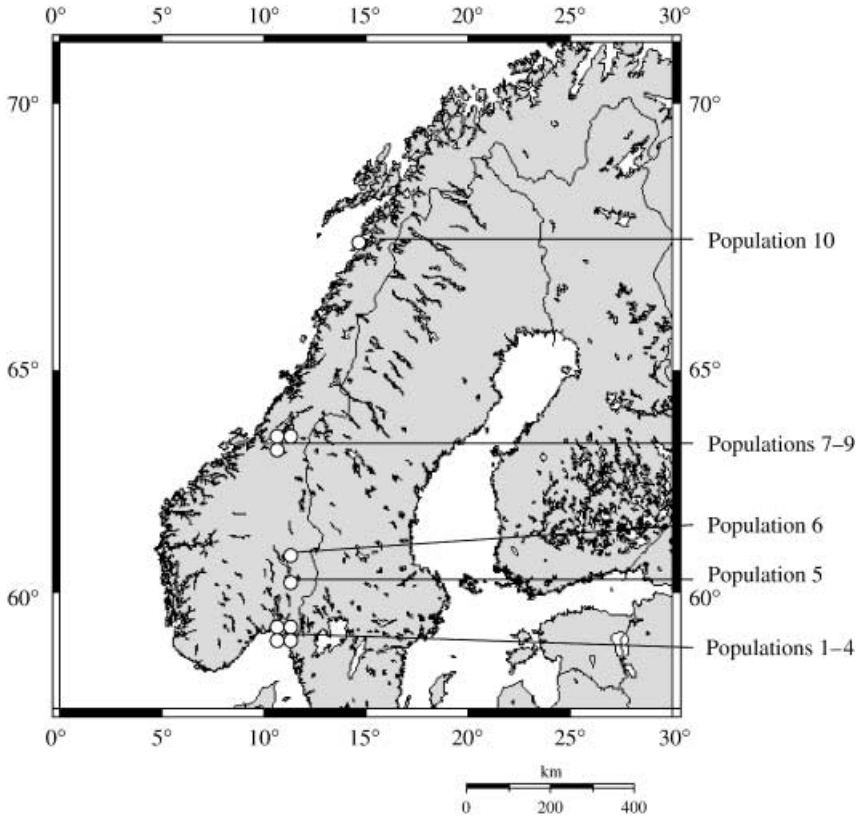

Fig. 1 Overview of study sites. Populations of *Arabidopsis thaliana* sampled along a latitudinal gradient in Norway and subjected to analyses of variation in quantitative traits and neutral molecular markers.

where X is the population mean of the transformed or nontransformed data. Total broad-sense heritabilities were estimated for each trait by estimating the variance as a result of maternal family nested within population.

Prior to genetic analysis on single-locus markers, seeds from 127 maternal families representing the 10 populations above were sown and vernalized for 7 days and then grown in constant daylight. The same maternal families were mostly employed for microsatellite screening as for quantitative assessments in populations 2, 3, 9 and 10. In the other populations, different maternal families were mostly employed for the molecular and quantitative assessments. DNA was extracted from *c*. 200 mg of plant leaves using the FastDNA® Kit of Qbiogene. A total of 25 microsatellite markers were used in the analysis: *nga12*, *nga208*, *nga128*, *nga168*, *nga158*, *AthCTR1A*, *nga162*, *nga129*, *AthCHIB*, *nga59*, *nga111*, *nga225*, *nga248*, *nga151* and *nga249* (Bell & Ecker 1994); *nga1139*, *AthPHYC* and *nga1145* (Ponce *et al*. 1999); and *AthCDPK9*, *ICE8*, *AthDET1*, *F20D22*, *F21M12*, *ICE7* and *ICE4* (Clauss *et al*. 2002). For two loci (*nga168* and *AthCHIB*), newly designed primers by Ponce *et al*. (1999) were used. To some of the loci (*nga158*, *AthCTR1A*, *nga111* and *nga225*), a new forward or reverse primer was designed in order to obtain a longer PCR (polymerase chain reaction) product (Table 2). To facilitate genotyping, a GTTT 'tail' was added to the 5′ end of each

Table 2 Summary of the primers newly designed in the present study (developed from Bell & Ecker 1994) and their analysis conditions

Locus	Forward primer	Reverse primer	Chrom.	Fluor. label	Primer amount (mM)	Pooling volume (mL)	Size range (bp)
nga158	ACCTGAACCATCCTCCGTC	GTTTGTAGCAAAGACAACAGAGGAAA*	5	HEX	0.5	1.5	289–309
AthCTR1A	GTTTGGTGACGGAAACTGACACCT	TATCAACAGAAACGCACCGAG	5	TET	0.5	2	327–337
nga111	GTTTCATCGACCCATGTTTCTTGAG	TGTTTTTTAGGACAAATGGCG	1	HEX	0.5	2	174
nga225	GAAATCCAAATCCCAGAGAGG	GTTTATTTGGCCCCAAAGACAAC	5	TET	0.5	2	220–237

*In order to enhance the 3′ adenylation, a GTTT tail was added to the 5′ end of nonlabelled primers as indicated by the underlining.

nonlabelled primer. This way, the nontemplated adenylation of the 3′ end of the DNA strand being synthesized should be enhanced to almost 100% (Brownstein *et al*. 1996).

The amplification reactions consisted of 10 ng of DNA, 0.2–0.7 μM of each primer, one of which was fluorescently labelled with TET, HEX or FAM, 1 × reaction buffer (Promega), 2.5 mM $MgCl_2$, 0.2 mM dNTP and 0.25 U of *Taq* DNA polymerase (Promega). Sterile H_2O was added to make the final reaction volume 10 μL. All amplifications were carried out on a Mastercycler gradient (Eppendorf) thermal cycler. Electrophoresis was performed on an ABI Prism™ 377 genetic analysis instrument. The GENESCAN 3.1 and GENOTYPER 2.0 softwares (Applied Biosystems) were used to analyse the DNA fragments and to score the genotypes. Haplotypes could be determined directly in individuals heterozygous at less than two loci. Only one individual found in population 6 contained multiple heterozygous loci. At four of the loci, one of the alleles was the only representative of its kind in this population. We assigned the haplotypes for this individual by assuming that one gamete represented a rare immigrant type containing the singleton alleles at four loci and the rarer allele at the fifth heterozygous locus. Immigration seems possible, as this population was found near a picnic site. The other haplotype was assumed to contain the common alleles at all loci.

Analyses of molecular diversity at locus and haplotypic levels were conducted, including estimation of expected heterozygosity (gene diversity, H_E; Nei 1987), frequency of polymorphic sites (P_p), and number of different haplotypes. The inbreeding fixation coefficient, F_{IS}, was estimated using analysis of molecular variance (AMOVA), as described in Excoffier *et al*. (1992). In an equilibrium mixed-mating model one can show that $F_{IS} = S/(2 - S)$ when F_{IS} is the inbreeding coefficient and S is the fraction of a population undergoing selfing (Allard *et al*. 1968). This relationship was used to estimate the degree of selfing in the populations.

The presence of linkage disequilibrium (LD) between all pairs of loci was tested using an exact test of linkage disequilibrium (Slatkin 1994; Schneider *et al*. 2000). The percentage of loci in linkage disequilibrium in relation to the maximum possible number of loci in linkage disequilibrium (P_d) was estimated for each population (Stenøien & Såstad 1999). We used an adjusted measure of P_d where significance values for the exact tests of LD were corrected for multiple comparisons by Bonferroni correction. In order to reveal genetic relationships, a neighbour-joining tree (Saitou & Nei 1987) was constructed for the detected different haplotypes, using the shared allele distance (Jin & Chakraborty 1993) as distance measure. Statistical support for the tree constructed was investigated by conducting 1000 bootstraps. The tree must be interpreted with caution, as there is evidence of recombination between loci (see succeeding discussion).

The fixation index F_{ST} (Weir & Cockerham 1984) and partitioning of variability by AMOVA were calculated for all pairs of populations. The significance of F_{ST} was tested using the nonparametric permutation approach described by Excoffier *et al*. (1992). Raymond & Rousset's (1995a) exact test of population differentiation was performed. Relative measures of population divergence, like F_{ST} may be inappropriate in systems with different intensities of forces that reduce within population variability, i.e. variation in outcrossing and migration rates (Charlesworth 1998). Therefore, at the haplotypic level, the average number of pairwise allelic differences between populations was also included as an absolute measure of between population divergences. This measure was corrected for the average number of pairwise differences between haplotypes within populations. Thus, tests of isolation by distance were performed in two ways: either by regressing $F_{ST}/(1 - F_{ST})$ on geographical distances for populations, both assuming linear and two-dimensional habitats (Rousset 1997), or by regressing corrected number of pairwise differences between populations on geographical distance. Phylogenetic relationships among populations were studied by the constructed neighbour-joining tree.

Pearson correlation analyses were performed to test for associations between levels of microsatellite diversity and variability in quantitative traits. Significance level was set to 0.05, and adjusted for multiple comparisons by Bonferroni corrections. Partitioning of quantitative variability within and among populations was summarized by the F_{ST} equivalent of polygenic traits, i.e. the Q_{ST} statistics (Wright 1951). Following Bonnin *et al*. (1996) Q_{ST} for a predominantly selfing species was calculated as

$$Q_{ST} = V_B/(V_B + V_G) \qquad \text{(eqn 3)}$$

where V_B is the variance among populations and V_G is the variance among selfed full-sib families nested within populations, assuming that the differences among populations are strictly genetic. Because the experiment was conducted in a randomized common environment, it is reasonable to assume that the among population differences that we detect were indeed genetic. However, maternal effects may have contributed to the among family component, thus inflating both the estimate of genetic variation within and between population level. Thus, our results should be considered with caution. Approximate SE (standard error) of Q_{ST} was obtained by bootstrapping estimates (850 samples) using the individual progeny as the level of sampling.

The degree of population divergence in single locus markers and neutral quantitative traits (as measured by F_{ST} and Q_{ST}, respectively) are expected to be similar and independent of mutation rates, and depend only on effective population size and migration rates which affect all traits equally (Prout & Barker 1989; McKay & Latta 2002). Consequently,

all else being equal, any difference between F_{ST} and Q_{ST} indices estimated for the same set of populations should be attributed to the effects of natural selection (Merilä & Crnokrak 2001). These predictions are general and do not depend on whether an island model or other population structure is assumed (Whitlock 1999), but it is assumed that populations are in drift-migration equilibrium. The relationships between F_{ST} and Q_{ST} were analysed for every quantitative trait measured in order to test for putative adaptive basis of quantitative variability.

Variance components analyses were performed using the VARCOMP procedure in SAS (SAS 2000). Heritabilities significantly greater than zero were determined by the significance of family effects using the random option of PROC GLM (SAS 2000). Various population genetic analyses were conducted with the aid of the ARLEQUIN ver. 2.000 software (Schneider *et al*. 2000) and the GENEPOP ver. 3.3 software (Raymond & Rousset 1995b). POPULATIONS ver. 1.2.28 (Langella, O. 1999 unpublished) was used for phylogenetic analyses, and trees were visualized with the TREEVIEW program (Page 1996). Other statistical analyses were done with the SPSS package (Windows version 11.0.0, SPSS Inc., Chicago, U.S.A.).

Results

A summary of within population variability measures is presented in Table 3. Six out of the 10 surveyed populations exhibited microsatellite polymorphisms. Across all populations, the average gene diversity per population (± SD) is $H_E = 0.06$ (± 0.08), the average frequency of polymorphic loci $P_p = 0.18$ (± 0.22), and average number of different haplotypes per population is 2.6 (± 3.1), given the assumption of haplotype assignment described above. When considering only polymorphic populations, then $H_E = 0.10$ (± 0.08), $P_p = 0.30$ (± 0.22), and the number of haplotypes per population is 4.3 (± 2.9). Population 3 has the highest gene diversity (0.21), 50% of the studied loci are polymorphic, and > 70% of the maternal families contain different haplotypes. Relatively few locus pairs are found to be in significant linkage disequilibrium in this population (35% of maximum possible number), and the population is highly genetically differentiated from other populations residing only 150–200 meters away (see below).

There is no correlation of microsatellite variability with estimated population size, or latitude. There is no vegetation type or habitat that seems to contain higher levels of variability than others. There is significant LD between pairs of polymorphic loci in the studied populations with P_d values ranging from 35% to 100%, suggesting lack of complete disequilibrium within *A. thaliana* populations. A total of 31 different haplotypes is found for all the populations, and genealogical relationships among the haplotypes are shown in Fig. 2. It is clear that all populations contain unique variants, and there are no shared haplotypes among the populations. Haplotypes from various geographical regions tend to group together, but the bootstrap support for the various nodes is mostly low, and the results are thus uncertain.

The average inbreeding coefficient within population 6 is $F_{IS} = 0.92$, and in the northernmost population 10 is $F_{IS} = 0.97$, and there are no heterozygotes in the four other populations containing genetic variability. In population 6, there are three individuals with at least one heterozygous locus, while in the northernmost population 10, one individual is heterozygous in one locus. In all, 11 haplotypes were observed in population 6, and observed heterozygosity in the various heterozygous loci ranges from 0.059 to 0.071 with a mean observed heterozygosity of 0.062 (± 0.005).

Average F_{ST} values range from 0.62 to 1 between pairs of populations, i.e. the populations are highly differentiated (all P values < 0.00001). Average F_{ST} over all populations

Table 3 Overview of within population microsatellite variability in studied *Arabidopsis thaliana* populations. Populations are numbered according to their geographical position with population 1 being the southernmost and population 10 being the northernmost population sampled in Norway

`	Pop 1 Hvaler	Pop 2 Hvaler	Pop 3 Hvaler	Pop 4 Moss	Pop 5 Jessheim	Pop 6 Stange	Pop 7 Skatval	Pop 8 Skatval	Pop 9 Trondheim	Pop 10 Meloy
Polymorphic loci (P_p)	—	—	0.50	—	0.04	0.60	0.13	—	0.20	0.32
Mean gene diversity (H_E)	—	—	0.21	—	0.01	0.17	0.06	—	0.05	0.11
Frequency heterozygotes	—	—	0	—	0	0.17	0	—	0	0.08
F_{IS}	—	—	1	—	1	0.92	1	—	1	0.97
Outcrossing frequency	—	—	0	—	0	4%	0	—	0	2%
Number of haplotypes	1	1	4	1	2	11	3	1	4	3
P_d	—	—	35%	—	—	52%	100%	—	60%	100%

The table shows frequency of polymorphic loci (P_p), mean gene diversity (H_E), frequency of heterozygote maternal families observed, inbreeding coefficient (F_{IS}), outcrossing frequency, number of different haplotypes per population, and percentage of locus pairs in significant linkage disequilibrium (LD) compared to maximum number of possible locus pairs in LD (P_d). '—' = no variability.

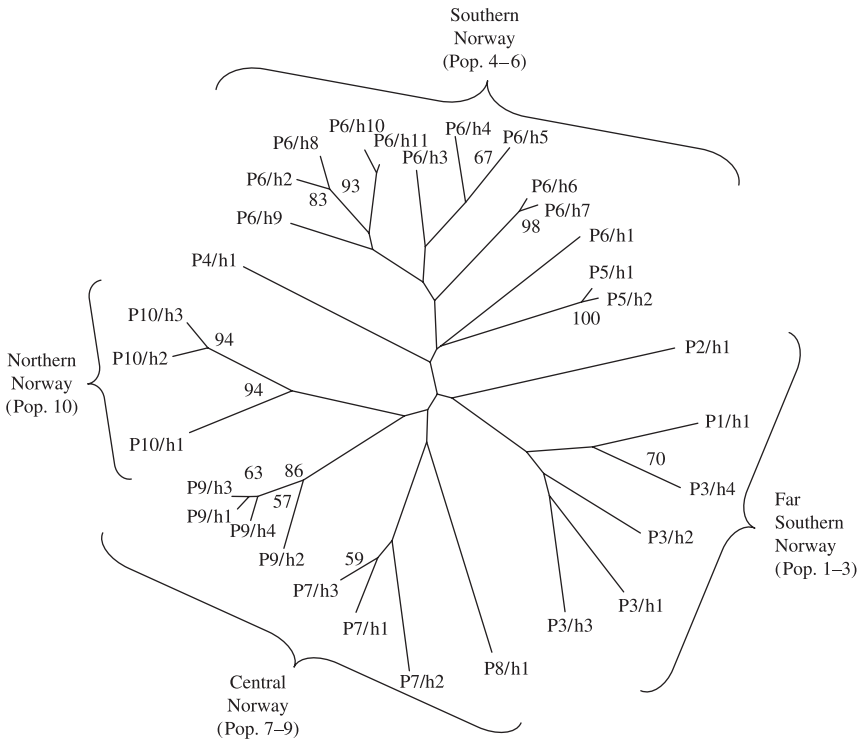

Fig. 2 Unrooted neighbour-joining tree of the 31 different *Arabidopsis thaliana* haplotypes detected in the 10 studied Norwegian populations. 'P' indicates the population the haplotype is sampled from and 'h' indicates haplotype number in a given population. The support from 1000 bootstraps are indicated at each node, but only bootstrap values > 50 are shown. Haplotypes from the same regions in Norway group together and there seems to be an association between geographical and genetic distances. However, bootstrap support is very low for the majority of nodes, and the tree cannot be viewed as phylogenetically informative.

Table 4 Coefficients of genetic variation for studied quantitative traits in populations of *Arabidopsis thaliana*

	Pop 1 & 2	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6	Pop 7	Pop 8	Pop9	Pop 10
Specific leaf mass	39.1			5.2	—	—	—	13.7	19.4	—	—
Flowering time	10.8			4.8	—	—	9.0	3.2	8.3	—	5.0
Trichome number	1.0			—	4.1	—	4.3	6.4	4.6	—	—
Stem length	16.0			—	—	—	14.6	4.9	4.7	6.2	—
Mean	16.7			2.5	1.0	—	7.0	7.1	9.3	1.6	1.3
Hypocotyl, blue		38.9	6.0	16.3	—	—	40.0	—	38.2	—	66.3
Hypocotyl, red		2.6	—	2.3	3.2	5.9	6.2	8.1	6.2	5.5	2.5
Hypocotyl, white		—	—	16.6	7.6	25.1	18.2	3.3	—	—	23.4
Hypocotyl, darkness		—	—	—	—	—	5.4	—	5.3	—	—
Hypocotyl, far red		4.0	13.6	7.6	0	3.1	12.5	—	3.4	—	2.7
Mean across treatments		9.1	3.9	8.6	2.2	6.8	16.5	2.3	10.6	1.1	19.0
Mean total CG_V				5.9	1.7	3.8	12.2	4.4	10.0	1.3	11.1

Closely situated populations 1 and 2 have been merged in the analysis of some traits because of limited germination. All populations exhibit significant genetic variability in all quantitative traits studied, even populations with no microsatellite variability. — no variability.

is found to be 0.88 ($P < 0.00001$). Tests of population differentiation are also all highly significant for all population pairs ($P < 0.00001$). From AMOVA, 12% of the variation is on average found among individuals within populations ($P < 0.001$), 76% of the variation is on average found among populations within the three geographical regions (i.e. southern, central and northern Norway, $P < 0.001$), and 12% is found among the three regions ($P = 0.003$). There is no indication of isolation by distance, based on F_{ST} estimates when regressing $F_{ST}/(1 - F_{ST})$ on geographical distances for populations ($P = 0.203$ and $P = 0.936$ when using distance and log of distance as independent variables, respectively). A neighbour-joining tree indicates isolation by distance, but it is not significant, as there is less than 50% bootstrap support for the majority of nodes (Fig. 2). Geographic distance and number of pairwise differences between populations is significantly associated according to a linear regression analysis ($R^2 = 0.161$, $P = 0.007$), but this is not the case if rightmost outliers (i.e. population pairs separated by > 500 km) are excluded from regression analysis ($R^2 = 0.083$, $P = 0.075$).

Tables 4 and 5 summarize the results of the analyses on quantitative traits. All populations display genetic variability in one or several quantitative traits, even populations

Table 5 Estimated partitioning of quantitative genetic variability within and among populations (Q_{ST}) for studied *Arabidopsis thaliana* populations

Quantitative trait	Q_{ST}	95% confidence interval
Blue	0.61	0.39–0.83
Red	0.56	0.36–0.76
White	0.84	0.54–1.00
Darkness	0.95	0.64–1.00
Far red	0.83	0.66–0.97
Specific leaf mass	0.98	0.68–1.00
Trichome number	0.99	0.80–1.00
Stem length	0.84	0.71–0.98
Flowering time	0.76	0.60–0.97

The 95% standard error (SE) estimates were obtained by bootstrapping. Quantitative traits include hypocotyl responses to various light regimes (see Stenøien *et al*. 2002), specific leaf mass, trichome number, stem length and days to flowering.

with no microsatellite polymorphisms. Significant heritabilities of 0.17 ($P < 0.005$) and 0.18 ($P < 0.02$) are observed for flowering time and stem length, respectively. Also, hypocotyl responses to blue and red light treatments have significant heritabilities of 0.19 ($P < 0.05$) and 0.34 ($P < 0.0001$), respectively. No significant genetic correlations are observed, based on the correlation of family means and following sequential Bonferroni corrections. There is a positive correlation between the frequency of polymorphic microsatellite loci and the average quantitative variability within populations, when considering the five hypocotyl traits (Pearson correlation coefficient 0.55, $P = 0.050$ in a one-tailed test). For the other quantitative traits, the within population variability cannot be estimated for populations 1 and 2. If these populations are excluded, then there is no significant correlation between the frequency of polymorphic microsatellite loci and the overall variability in all quantitative traits (Pearson correlation coefficient 0.51, $P = 0.097$ in a one-tailed test). On the individual trait level, there is also a significant positive correlation between within-population variability in hypocotyl responses both to white and blue light, and variability in both gene diversity and frequency of polymorphic loci (correlation coefficients 0.71–0.82, with P values ranging from 0.012 to 0.048), but no association is significant following Bonferroni correction for multiple comparisons. A significant positive correlation between within population variability in hypocotyl responses to far red light and number of polymorphic loci is also observed (correlation coefficient 0.73 with $P = 0.041$), but neither this latter association is significant following Bonferroni correction. The coefficients of genetic variation within populations are high for merged populations 1 and 2, as expected, given the high microsatellite differentiation between these two populations (see above). Populations 6, 8 and the northernmost population 10 have the highest levels of variability in quantitative traits. Estimated Q_{ST} values are significantly lower than 1 for the traits days to first flower, stem length and hypocotyl response to far red, red, and blue light, based on the nonoverlap of the *c*. 95% confidence intervals with 1.0. However, only for the traits hypocotyl response to red and blue light, are the Q_{ST} values significantly different from F_{ST}, based on the nonoverlap of the 95% confidence intervals with the estimated F_{ST} of 0.888.

Discussion

We quantified relatively high levels of microsatellite variability within studied Norwegian populations compared to many previous studies (Abbott & Gomes 1989; Todokoro *et al*. 1995; Kuittinen *et al*. 1997; Bergelson *et al*. 1998; Loridon *et al*. 1998; Breyne *et al*. 1999; Vander Zwan *et al*. 2000; Clauss *et al*. 2002). For instance, no microsatellite polymorphisms were found within studied Japanese populations, based on relatively small samples (Todokoro *et al*. 1995), and average gene diversity over all 10 Norwegian populations is six times higher than what was reported in Clauss *et al*.'s (2002) study. Our study is not the first to report high variability in northernmost *Arabidopsis thaliana* populations. Berge *et al*. (1998) found high levels of isozyme variability in one Norwegian *A. thaliana* population, Kuittinen *et al*. (1997) reported high levels of gene diversity in one Danish and one Finnish population, and Fedorenko *et al*. (2001) found high isozyme variability within Karelian populations. Together, these results demonstrate that natural populations of a selfing annual plant may be highly variable, at least in parts of the distribution range. Other studies have found variable levels of within population genetic diversity in several selfing plants (Schoen & Brown 1991). Our study thus confirm results from studies of other vascular and nonvascular plants that life history characteristics associated with reduced effective population sizes, such as inbreeding, asexuality, haploidy, and extensive metapopulation dynamics, do not necessarily completely exhaust within population genetic variability for either markers or quantitative traits (e.g. Wyatt *et al*. 1989; Fenster & Ritland 1992; Carr & Fenster 1994; Stenøien & Såstad 1999).

In peripheral parts of a species' distribution range, populations could be subject to extreme ecological conditions and maximum stress, and this may cause population variability to be reduced, i.e. Vavilov's (1926) gene centre theory (e.g. Schwartz *et al*. 2003). Following these predictions, Kuittinen *et al*. (1997) explained the low variability in some Japanese and northern European *A. thaliana* populations as caused by the fact that these are at the outer range of the species' distribution, whereas the east and west European populations are more central. They argued that effects of drift may be pronounced in populations in northern Europe because of small population sizes, repeated

bottlenecks and low founder numbers in unfavourable environmental conditions. Our findings do not support this hypothesis, because the general levels of variability do not seem lower in the northernmost compared to the central European populations. Furthermore, northern Norwegian populations are not more genetically depleted than populations from southern Norway. The northernmost population studied (at 67′ N) has a seemingly low census population size, but still manifests high levels of variability. In fact, this northernmost population has the highest mean coefficient of genetic variation for quantitative traits, as a result of high within population variability particularly in hypocotyl responses to different light regimes. The data suggest that there is a rich source of genetic variation from which the Norwegian populations have been drawn.

Although several studies have reported low levels of variability in populations from central parts of the species' distribution range, one cannot conclude from our data that peripheral populations of *A. thaliana* contain higher levels of variability than central populations. For instance, Bergelson *et al*. (1998) found more than one haplotype within half of their studied local populations in the USA, England, Kazakhstan and Czechoslovakia, i.e. both central and peripheral populations. Nevertheless, there are several examples of species harbouring higher genetic variability in the periphery compared to central parts (see Schwartz *et al*. 2003 for an overview). More studies are needed to see whether this is also the case for *A. thaliana*.

The observed pattern of relatively high variability in *A. thaliana* is possibly caused by several different factors, each factor possibly affecting different populations in different ways. Some variable *A. thaliana* populations may have experienced high levels of immigration in recent generations from a set of unknown source populations. For instance, in population 6, three individuals contain one or more heterozygous loci. One individual is heterozygous at five loci, containing alleles not found in other individuals from that population. This highly heterozygous individual may be the result of an outcross with a genotype that initially arose from a different population.

Nevertheless, ongoing gene flow is unlikely to explain the overall distribution of variability and structuring in the studied populations. All populations are highly differentiated, as measured by F_{ST} statistics, tests of population differentiation and absolute measures of number of pairwise differences, as found in earlier studies (e.g. Bergelson *et al*. 1998). There are no shared haplotypes among the populations, and even populations separated by a few hundred meters show F_{ST} values close to or equal to one, reflecting the selfing nature of this species. An unrooted neighbour-joining tree could indicate that closely situated populations are more genetically similar, as measured by the number of shared alleles (Fig. 2), but bootstrap support is very low for this tree. The tree has a distinct star-shaped topology indicating recent spread and expansion of the studied populations. Furthermore, there is lack of association between F_{ST} and geographical distance, which may indicate little local and regional migration for most populations. There is a significant association between the corrected number of pairwise differences and the geographical distance between pairs of populations, indicating isolation by distance. However, if comparisons between southernmost populations and the northernmost population 10 are removed from the regression analysis (i.e. population pairs separated by more 500 km), then the analysis becomes nonsignificant, suggesting a too large pairwise differentiation on a local level compared to expectations based on geographical distances. Thus, comparison of geographical distance and pairwise differentiation does not give an unambiguous support to the theory of isolation by distance. Moreover, there is no evidence of a progressive colonization pattern from south to north with constant divergence of haplotypic variants (Fig. 2). This lack of progressive divergence of haplotypes does not indicate a wave-like colonization of northern sites by migrants entirely from adjacent southern populations. Rather, the variability pattern may indicate colonization from a number of different sources, maybe even from more eastern populations, or from different glacial refuges, as observed, e.g. in European common ash (Heuertz *et al*. 2004).

In the absence of gene flow, within population variability is mainly generated by novel mutations, or recombination in the presence of several founding genotypes. It is difficult to determine the relative roles of mutation and recombination in generating haplotypic diversity in the present samples. However, it seems that variability in population 5 could be resulting from the establishment of one haplotype and then a subsequent mutation event at one locus, giving rise to a second allele at this locus with almost similar repeat copy number (data not shown). In populations 3, 7 and 9, two immigrating haplotypes could have given rise to the other haplotypes present by recombination. In population 6, more than two haplotypes must have been present to account for the haplotypic variability, i.e. either many mutations and/or several immigration events must have taken place. Because population 6 is situated near a picnic area, the occurrence of rare immigration events as a result of human mediated transport seem possible. If just two haplotypes came to population 10, then there must have been some recombination, or else more haplotypes must have immigrated. More studies of microsatellite variability in Karelian populations, e.g. populations from Finland and westernmost parts of Russia (e.g. Onega Lake coast), may shed light on alternative immigration routes for the northernmost populations.

Patterns of haplotypic diversity indicate historical recombination events in studied populations. There are also other reasons for believing that recombination may

take place relatively often within at least some of the variable populations studied. In general, one may assume that patchy subdivided and highly inbred populations should yield high LD levels, and extensive LD have indeed been demonstrated within local populations of *A. thaliana*, even between distant loci (Kuittinen *et al*. 2002; Nordborg *et al*. 2002). For instance, Nordborg *et al*. (2002) found little decay in LD over 50–100 cM (centimorgans) within studied Michigan populations of *A. thaliana*. There is a strong haplotypic structure in the Norwegian polymorphic populations, as revealed by significant LD. Nevertheless, there is still evidence of some recombination, e.g. only 35% of all locus pairs have significant LD in one of the polymorphic populations. However, these results should be interpreted with caution as the sample sizes are relatively low, and the statistical power is thus limited.

Despite the observation that recombination may be important for shaping haplotypic variability, we only find heterozygous individuals in two of the populations, with three individuals with at least one heterozygous locus found in population 6. Abbott & Gomes (1989) estimated outcrossing to take place in less than 1% of *A. thaliana* matings, and we find outcrossing on average to occur in the same order of magnitude based on our overall F_{IS} estimates The rates are also consistent with insect visitation data (Hoffman *et al*. 2003). We cannot rule out that outcrossing frequencies vary among variable populations. Specific environmental conditions, e.g. associated with the frequency of insect visitation, may yield outcrossing in one generation, followed by inbreeding in some or many generations.

The traits studied here were chosen because their intraspecific pattern of expression has often been shown to have adaptive significance, e.g. day of first flower frequently corresponds to the length of the growing season (Galloway & Fenster 2001); variable stem length is associated with different herbivory patterns (Ehrlén *et al*. 2002; Aspi *et al*. 2003) as is variation in the trichome number in *A. thaliana* (Mauricio 1998), while specific leaf area is often negatively correlated to drought resistance (Dudley 1996; Knight & Ackerly 2003). We used the hypocotyl responses to the various light regimes as a proxy to quantifying variation in the photoperiodic pathway, which has been implicated in contributing to, e.g. flowering time (Weinig *et al*. 2002), and adaptation to light regimes (Dorn *et al*. 2000). Parallel to our findings with molecular markers, Norwegian population of *A. thaliana* harbour significant genetic variation for these life history and morphological traits. Significant heritabilities for date of first flower, stem length and hypocotyl response to various light conditions indicate the potential for these traits to respond to selection within the population level.

F_{ST} values for neutral markers close to 1 limit our ability to infer selection acting on the measured morphological and life history traits. We expect F_{ST} values to be high for a predominantly selfing organism, but our values and those reported by Kuittinen *et al*. (1997) also for Scandinavian populations of *A. thaliana* are approximately twice those reported as average for selfing plant species (Hamrick & Godt 1989). In other studies, Q_{ST} is commonly larger than F_{ST} (Merilä & Crnokrak 2001), suggesting selection for different optima in different populations. Because we observed F_{ST} values close to one, we cannot distinguish between selection and drift as causal agents for the differentiation for traits where we observe Q_{ST} also close to 1. Thus, it may very well be that selection and drift are both contributing to, e.g. differentiation of trichome number and specific leaf area in studied populations. Field-based reciprocal transplant studies or tests based on QTL (quantitative trait loci) investigation of between population differentiation (Orr 1998; Rieseberg *et al*. 2002) are necessary to document the role of selection in the population differentiation of these traits. On the other hand, where we did observe significantly lower Q_{ST} values than the overall F_{ST} value of 0.888, is suggestive of selection retarding differentiation. These traits certainly have the ability to respond further to selection as they exhibit significant genetic variation within population.

Stenøien *et al*. (2002) found clinal variability for hypocotyl response to red and far red light in the studied *A. thaliana* populations, suggesting that hypocotyl responses could be an adaptation to the various light regimes found along the studied gradient. Clinal variation in genetically-based traits can provide evidence for spatially varying selection across an environmental gradient (Haldane 1948; Endler 1977). However, clinal variation in allelic frequencies at genes underlying a particular trait can result from gene flow between partially isolated populations that have diverged in genetic composition via drift, or admixture between two or more genetically-differentiated founding populations (e.g. Storz 2002). Clines can thus be just as well explained by a nonselective isolation by distance model as selection by correlated environmental factors (Gould & Johnston 1972). If the observed cline in trait values is solely attributable to isolation by distance, then Q_{ST} for the trait and F_{ST} for neutral markers should exhibit concordant patterns of increase with geographical distance (Merilä & Crnokrak 2001; Storz 2002). Furthermore, the strong LD in a predominantly selfing population would also contribute to similar patterns of microsatellite and quantitative variation under the neutral model. In the absence of natural selection, there should therefore be a strong association between F_{ST} and Q_{ST}. The significantly lower Q_{ST} compared to F_{ST}, and the observed lack of correlation between F_{ST} values and geographical distance among populations thus support the hypothesis of selection for hypocotyl elongation in natural populations of *A. thaliana*. Therefore, our results support the suggestion of Stenøien *et al*. (2002) that a hypocotyl response at least to red and also blue light regimes is indeed an adaptation.

Acknowledgements

The authors thank Joy Bergelson, Helmi Kuittinen and Jon Ågren for their comments on the manuscript and Marianne Hansen for technical assistance. This study was financially supported by the Research Council of Norway (grant no. 134800/410), and the Swedish Research Council (grant no. 621-2002-5896). The authors acknowledge financial support from the Academy of Finland and the University of Oulu.

References

Abbott RJ, Gomes MF (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity*, **62**, 411–418.

Allard RW, Jain SK, Workman P (1968) The genetics of inbreeding species. *Advances in Genetics*, **14**, 55–131.

Aspi J, Jäkäläniemi A, Tuomi J, Siikamäki P (2003) Multilevel phenotypic selection on morphological characters in a metapopulation of *Silene tatarica*. *Evolution*, **57**, 509–517.

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

Berge G, Nordal I, Hestmark G (1998) The effect of breeding systems and pollination vectors on the genetic variation of small plant populations within an agricultural landscape. *Oikos*, **81**, 17–29.

Bergelson J, Stahl E, Dudek S, Kreitman M (1998) Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics*, **148**, 1311–1323.

Bonnin I, Prosperi JM, Oliveri I (1996) Genetic markers and quantitative genetic variation in *Medicago trunculata* (Leguminosae): a comparative analysis of population structure. *Genetics*, **143**, 1795–1805.

Breyne P, Rombaut D, Van Gysel A, Van Montagu M, Gerats T (1999) AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes. *Molecular and General Genetics*, **261**, 627–634.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of nontemplated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques*, **20**, 1004–1010.

Carr DE, Fenster CB (1994) Levels of genetic variation and covariation for *Mimulus* (Scrophulariaceae) floral traits. *Heredity*, **72**, 606–618.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Clauss MJ, Cobban H, Mitchell-Olds T (2002) Cross-species microsatellite markers for elucidating differences in population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Molecular Ecology*, **11**, 591–601.

Dorn LA, Pyle EH, Schmitt J (2000) Plasticity to light cues and resources in Arabidopsis thaliana: testing for adaptive value and costs. *Evolution*, **54**, 1982–1994.

Dudley SA (1996) The response to differing selection on plant physiological traits: evidence for local adaptation. *Evolution*, **50**, 103–110.

Ehrlén J, Käck S, Ågren J (2002) Pollen limitation, seed predation and scape length in *Primula farinosa*. *Oikos*, **97**, 47–51.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fedorenko OM, Savushkin AI, Olimpienko GS (2001) Genetic diversity in natural populations of *Arabidopsis thaliana* (L.) Heynh. from Karelia. *Russian Journal of Genetics*, **37**, 223–229.

Fenster CB, Ritland K (1992) Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *American Journal of Botany*, **79**, 1440–1447.

Fremstad E (1997) Vegetation types in Norway (Vegetasjonstyper i Norge). *Norsk Institutt for Naturforskning Temahefte*, **12**, 1–279.

Galloway LF, Fenster CB (2001) Nuclear and cytoplasmic contributions to intraspecific divergence in an annual legume. *Evolution*, **55**, 488–497.

Gould SJ, Johnston RF (1972) Geographic variation. *Annual Review of Ecology and Systematics*, **3**, 457–498.

Haldane JBS (1948) The theory of a cline. *Journal of Genetics*, **48**, 277–284.

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates, Sunderland, MA.

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

Hoffmann MH (2002) Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *Journal of Biogeography*, **29**, 125–134.

Hoffmann MH, Glass AS, Tomiuk J, Schmuths H, Fritsch RM, Bachmann K (2003) Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with geographical information systems (GIS). *Molecular Ecology*, **12**, 1007–1019.

Ingvarsson PK (2002) A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution*, **56**, 2368–2373.

Innan H, Terauchi R, Miyashita T (1997) Microsatellite polymorphisms in natural populations of the wild plant *Arabidopsis thaliana*. *Genetics*, **146**, 1441–1452.

Jin L, Chakraborty R (1993) Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. *Molecular Biology and Evolution*, **11**, 120–127.

Knight CA, Ackerly CC (2003) Evolution and plasticity of photosynthetic thermal tolerance, specific leaf area and leaf size: congeneric species from desert and coastal environments. *New Phytologist*, **160**, 337–347.

Kuittinen H, Mattila S, Savolainen O (1997) Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity*, **79**, 144–152.

Kuittinen H, Salguero D, Aguade M (2002) Parallel patterns of sequence variation within and between populations at three loci of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **19**, 2030–2034.

Le Corré V, Roux F, Reboud X (2002) DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution*, **19**, 1261–1271.

Lewontin RC (1974) *The Genetic Basis of Evolutionary Change*. Columbia Univeristy Press, New York.

Loridon K, Cournoyer B, Goubely C, Depeiges A, Picard G (1998) Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*. *Theoretical and Applied Genetics*, **97**, 591–604.

Mauricio R (1998) Cost of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *American Naturalist*, **151**, 20–28.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution*, **17**, 285–291.

Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.

Miyashita NT, Kawabe A, Innan H (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified length polymorphism analysis. *Genetics*, **152**, 1723–1731.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics*, **154**, 923–929.

Nordborg M, Borevitz JO, Bergelson J *et al.* (2002) The extent of linkage disequilibrium in the highly-selfing species *Arabidopsis thaliana*. *Nature Genetics*, **30**, 190–193.

Nordborg M, Charlesworth B, Charlesworth D (1996) The effect of recombination on background selection. *Genetical Research*, **67**, 159–174.

Nordborg M, Donnelly P (1997) The coalescent process with selfing. *Genetic*, **146**, 1185–1195.

Orr HA (1998) Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, **149**, 2099–2104.

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357–358.

Ponce MR, Robles P, Micol JL (1999) High-throughput genetic mapping in *Arabidopsis thaliana*. *Molecular and General Genetics*, **261**, 408–415.

Prout T, Barker J (1989) Ecological aspects of the heritability of body size in *Drosophila buzzatii*. *Genetics*, **123**, 803–813.

Provan J, Campanella JJ (2003) Patterns of cytoplasmic variation in *Arabidopsis thaliana* (Brassicaceae) revealed by polymorphic chloroplast microsatellites. *Systematic Botany*, **28**, 578–583.

Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.

Raymond M, Rousset F (1995b) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2002) Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences USA*, **99**, 12242–12245.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evoution*, **4**, 406–425.

SAS Institute (2000) *SAS/STAT user's guide, Version 8*. SAS Institute, Cary, NC.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN ver 2.000: A software for population genetics analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. http://lgb.unige.ch/arlequin/.

Schoen DJ, Brown AHD (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences U.S.A.*, **88**, 4494–4497.

Schwartz MK, Mills LS, Ortega Y, Ruggiero LF, Allendorf FW (2003) Landscape location affects genetic variation of Canada lynx (*Lynx Canadensis*). *Molecular Ecology*, **12**, 1807–1816.

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109–2118.

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. *Genetics*, **137**, 331–336.

Stenøien HK, Fenster CB, Kuittinen H, Savolainen O (2002) Quantifying latitudinal clines to light responses in natural populations of *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany*, **89**, 1604–1608.

Stenøien HK, Såstad SM (1999) Genetic structure in three haploid peat mosses (*Sphagnum*). *Heredity*, **82**, 391–400.

Storz JF (2002) Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Molecular Ecology*, **11**, 2537–2551.

Todokoro S, Terauchi R, Kawano S (1995) Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. *Japanese Journal of Genetics*, **70**, 543–554.

Vander Zwan C, Brodie SA, Campanella JJ (2000) The intraspecific phylogenetics of *Arabidopsis thaliana* in worldwide populations. *Systematic Botany*, **25**, 47–59.

Vavilov NI (1926) Studies on the origin of cultivated plants. *Bulletin of Applied Botany Leningrad*, **16**, 1–248.

Wakeley J, Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics*, **159**, 893–905.

Weinig C, Ungerer MC, Dorn LA *et al.* (2002) Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics*, **162**, 1875–1884.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock MC (1999) Neutral additive genetic variance in a metapopulation. *Genetical Research*, **74**, 215–221.

Wright S (1951) The genetic structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wyatt R, Odrzykoski IJ, Stoneburner A (1989) High levels of genetic variability in the haploid moss *Plagiomnium ciliare*. *Evolution*, **43**, 1085–1096.

Hans Stenøien is an assistant professor at Uppsala University. His research on evolution at the molecular level spans from explaining evolutionary stasis in primitive land plants (mosses), to understanding the molecular basis of adaptive differentiation in angiosperms (*Arabidopsis*). Charlie Fenster is in the Department of Biology at the University of Maryland. His research focuses on quantifying the modes of microevolutionary process, including gene flow, natural selection and mutation, using plants as model systems as well as determining the genetic architecture underlying fitness differentiation and floral evolution. Anni Tonteri is currently a Ph.D. student at the Department of Biology, University of Turku, where she is working on the population genetic structure of north European Atlantic salmon. Outi Savolainen is at the Oulu University. She and her group study the genetics of adaptation in plants, in *Arabidopsis* species and Scots pine.
