Molecular Ecology (2003) **12**, 2073–2085 doi: 10.1046/j.1365-294X.2003.01898.x

Genetic structure and gene flow in a metapopulation of an endangered plant species, *Silene tatarica*

N. TERO,* J. ASPI,* P. SIIKAMÄKI,† A. JÄKÄLÄNIEMI† and J. TUOMI*
**Department of Biology, PO Box 3000, FIN-90014 University of Oulu, Finland, †Oulanka Biological Station, Liikasenvaarantie 134, FIN-93999 Kuusamo, Finland*

Abstract

We investigated the distribution of genetic variation within and between seven subpopulations in a riparian population of *Silene tatarica* in northern Finland by using amplified fragment length polymorphism (AFLP) markers. A Bayesian approach-based clustering program indicated that the marker data contained not only one panmictic population, but consisted of seven clusters, and that each original sample site seems to consist of a distinct subpopulation. A coalescent-based simulation approach shows recurrent gene flow between subpopulations. Relative high F_{ST} values indicated a clear subpopulation differentiation. However, AMOVA analysis and UPGMA-dendrogram did not suggest any hierarchical regional structuring among the subpopulations. There was no correlation between geographical and genetic distances among the subpopulations, nor any correlation between the subpopulation census size and amount of genetic variation. Estimates of gene flow suggested a low level of gene flow between the subpopulations, and the assignment tests proposed a few long-distance bidirectional dispersal events between the subpopulations. No apparent difference was found in within-subpopulation genetic diversity among upper, middle and lower regions along the river. Relative high amounts of linkage disequilibrium at subpopulation level indicated recent population bottlenecks or admixture, and at metapopulation levels a high subpopulation turnover rate. The overall pattern of genetic variation within and between subpopulations also suggested a 'classical' metapopulation structure of the species suggested by the ecological surveys.

Keywords: AFLP, dispersal, gene flow, genetic structure, metapopulation, *Silene*

Received 29 January 2003; revision received 2 May 2003; accepted 2 May 2003

Introduction

Many natural species are subdivided into local breeding units, a condition that can lead to the genetic differentiation of local demes. Much of the theory of population structure contrasts the effects of finite local population size and the rate of the gene flow among populations on genetic structure. Gene flow strongly influences the spatial scale over which genetic differentiation will be observed (Slatkin 1985a).

In addition to population structure, metapopulation dynamics may have significant consequences on the amount of genetic variation within demes and also at the population level (Pannell & Charlesworth 1999, 2000). Population structure is affected critically by nonequilibrium dynamics and the properties of new populations; in particular, founding number, probability of common origin and kin structure are vital in understanding genetic variation (Slatkin 1977; Whitlock & McCauley 1990; Hanski 1999). In plant metapopulations, which are subject to catastrophic forms of disturbance, long-distance seed dispersal is essential for metapopulation to persist (Cain *et al*. 2000).

In the case of an aggregate of linearly arranged subpopulations, five basic hypotheses of migration patterns with contrasting predictions can be derived (Fig. 1). In the first three models the patches or subpopulations are principally permanent. In the first of these models, despite the apparent spatial subpopulation structure, the patches form a single panmictic unit with free gene flow throughout the

Correspondence: Jouni Aspi, Department of Biology, FIN-90014 University of Oulu, Finland. Fax: + 38 8 5531061; E-mail: jouni.aspi@oulu.fi

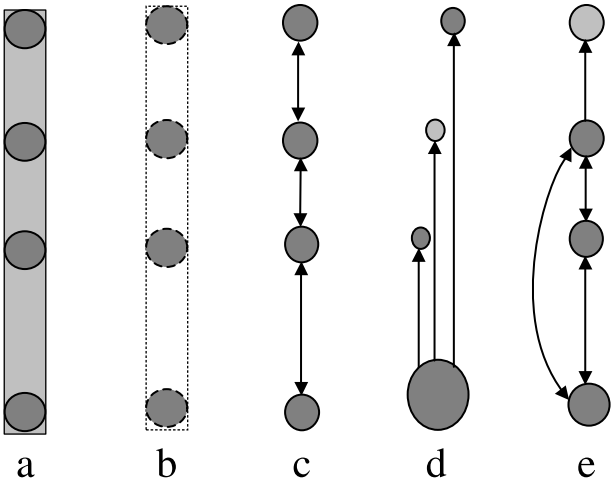

Fig. 1 Schematic representation of the different possible models of population structure and migration between linearly arranged patches or subpopulations: (a) genetically uniform population with a free gene flow across the population; (b) fragmentized population without recurrent gene flow between the subpopulations; (c) stepping-stone population; (d) source-sink population; (e) 'classical' metapopulation.

genetically uniform population (Fig. 1a). Second, the patches may represent only fragments of formerly continuous population without any present-day migration between the subpopulations (Fig. 1b). The third hypothesis is the classical stepping-stone model with migration between only the adjacent temporally persistent subpopulations (Fig. 1c). The last two models are metapopulation models, in which the subpopulations are more or less ephemeral. Two patterns of migration between subpopulations in a metapopulation may be distinguished. In the first one, the subpopulations form a source-sink metapopulation (Fig. 1d), in which more permanent, high-quality source subpopulations are the net exporters and ephemeral low quality subpopulations net importers of the migrants. The sink subpopulations will go extinct without migration from source populations. In the last 'classical' metapopulation model (Fig. 1f) all the subpopulations are ephemeral, and the source of migrating individuals may be more diverse.

These different models of population structure provide different testable hypotheses about the distribution of genetic variation within and between subpopulations. Genetic information permits inferences about how microevolutionary forces have interacted throughout the history of subpopulations and the amount of gene flow that has occurred between them. According to the panmixia hypothesis, there should be no population structure among the patches. On the contrary, the vicariance or fragmentation hypothesis assumes no recurrent gene flow between the subpopulations.

In contrast to the panmixia and metapopulation models, the stepping-stone models assume a population structure where migration occurs only between adjacent local subpopulations (Kimura 1953; Kimura & Weiss 1964). Hence at equilibrium, genetic distance between subpopulations should increase monotonically with increasing geographical distance between them (Kimura & Weiss 1964). According to Hutchison & Templeton (1999), one can reject the null hypothesis that the populations of a region are at equilibrium if there is an insignificant association between genetic and geographical distances separating pairs of populations in a region, and a scatterplot of the genetic and geographical distances separating each pairwise combination of populations fails to reveal a positive and monotonic relationship over all distance values of a region. Moreover, the scatter of the pairwise points should increase outward from narrow at the origin to wider at further geographical distances of separation. In the stepping-stone model the amount of genetic variation in subpopulations is also determined by mutation-drift equilibrium. Loss of heterozygosity because of genetic drift in these subpopulations is predicted to be related inversely to effective subpopulation size (e.g. Nei 1987). Accordingly, we should find significant association between subpopulation size and the amount of genetic diversity within the investigated subpopulations.

Both the source-sink and stepping-stone models predict significant differences in genetic diversity between subpopulations. In the source-sink model, within-population heterozygosity will be much reduced in sink subpopulations relative to source subpopulation. Genetic variation can be maintained in a small sink subpopulation only through the continued influx of genes from a large source (Gaggiotti & Smouse 1996). A recent coalescent-based stepping-stone migration model also suggested differences in genetic diversity between subpopulations. However, the genetic diversity in this model is expected to be higher near the centre of the habitat than near the ends (Wilkins & Wakeley 2002).

A theoretical prediction that is shared with source-sink and metapopulation dynamics is that recent population bottlenecks can increase linkage disequilibrium between loci considerably within subpopulations (McVean 2002). Extinction–replacement dynamics should also generate a relatively high level of linkage-disequilibrium at total population level (Ohta 1982).

Population structure and pattern of migration between subpopulations contribute significantly to the choice of the method used to estimate the amount of gene flow between subpopulations. Indirect methods of gene flow estimation characterize the spatial distribution of genotypes by some parameter and then apply a population genetic model to ask what level of gene flow would produce a distribution with the same parameter value. Traditional *F*-statistic

(Wright 1969) and private allele (Slatkin 1985b) approaches provide indirect estimates of the average effective number of migrants exchanged per generation among a set of populations. These methods are often inappropriate, because natural sets of populations may not be at equilibrium (Bossart & Prowell 1998; Hutchison & Templeton 1999; Whitlock & McCauley 1999). When immigration is sufficiently low, assignment analysis can be used to estimate dispersal rates in natural populations by computing the proportion of individuals that are identified as immigrants (Cornuet *et al*. 1999; Cain *et al*. 2000).

Silene tatarica is an endangered perennial plant growing along periodically disturbed riverbanks and shores of two rivers in Finland. The species has spatially structured populations with subpopulations or patches scattered widely in a more or less continuous area of suitable habitat along riverside. The plants grow from open sand, gravel shores and erosion banks to more densely vegetated shores and riverbanks (Aspi *et al*. 2003). The establishment of new patches and the expansion of old patches occur exclusively by seed dispersal in *S. tatarica*. The seeds are dispersed presumably by gravity, wind and water, and perhaps animals and man. Short-distance primary dispersal may lead to fine-scaled spatial (Tilman & Kareiva 1997) and genetic structure within and among populations (McCauley 1997; Giles *et al*. 1998; Ingvarsson & Giles 1999; Richards *et al*. 1999). Hydrochory or zoochory, however, may lead to effective long-distance dispersal (Kudoh & Whigham 1997), which may weaken the regional association between genetic and geographical distances over the entire metapopulation. If gene movements are driven by hydrochoric dispersal, the movements should be unidirectional from the upper toward the lower drainage. On the other hand, zoochory and anthropochory may also lead to bidirectional dispersal.

The aim of this study is to reveal whether the riverbank population of *S. tatarica* is consistent with the abovementioned population structure and migration models. The specific questions addressed are:

1 Is there any genetic structuring in *S. tatarica* and, if so, are the subpopulations completely isolated, or are there any migration between subpopulations?
2 Are there correlations between genetic and geographical distances and also between subpopulation size and heterozygosity?
3 Is there significant variation in genetic diversity among different subpopulations or parts of the distribution area?
4 Do the linkage disequilibrium within subpopulations and the whole population level indicate population bottlenecks and high rate of extinction–replacement during the recent history?
5 What is the level of gene flow between subpopulations, and is it uni- or bidirectional?

Materials and methods

Plant material

The main distribution area of *S. tatarica* is on the Russian steppes, with disjunctive occurrences in Hungary, Germany, Lithuania and northwest Russia (Ulvinen 1997). The northwestern range of the distribution area of the species is in northern Finland, where it has invaded naturally the riverside habitats of the Oulankajoki and the Kitinen Rivers. These rivers are over 100 km apart and separated by a watershed. In addition to these natural populations a few secondary populations are found from railway- and roadsides, indicating successful anthropochoric and/or zoochoric dispersal.

The study area is situated in the Oulankajoki River valley in northern Finland and in northwestern Russia (66°21′–66°16′ N and 29°24′–29°21′ E). The Oulankajoki River is an unregulated river with natural dynamics. Ice jams and dramatic seasonal fluctuations in water level are typical, causing frequent disturbances at the riverside (Koutaniemi 1984).

We collected plant samples from *c*. 30 individuals from seven different sites along the Oulankajoki River (Fig. 2, Table 1) in summer 1999. Five of the sites were in Finland and two in Russia. The number of individuals in each subpopulation was estimated by counting all fertile plants during the flowering period. The straight-line distance between subpopulations was estimated directly on the basis of the site coordinates. Distances between subpopulations were also measured along the river.

DNA extraction and amplified fragment length polymorphism (AFLP) analysis

Genomic DNA was isolated from leaves of individual *Silene* plants using the slightly modified cetyltrimethyl ammonium bromide (CTAB) method (Rogers & Bendich 1985). The starting material was 0.1 g frozen leaves.

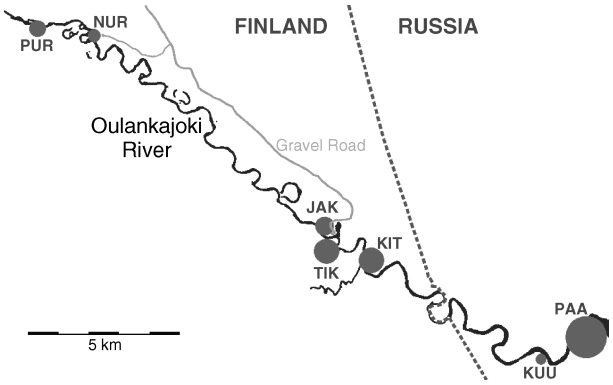

Fig. 2 Schematic map of the sample sites. The subpopulation name abbreviations refer to Table 1.

Table 1 Genetic diversity within subpopulations in *Silene tatarica*. N is the size of a subpopulation, N2 is the number of individuals analysed, *P* is the percentage of polymorphic loci, H'_{eN} is Nei's gene diversity assuming Hardy–Weinberg equilibrium, H'_{eH} is the Bayesian gene diversity (without assuming Hardy–Weinberg equilibrium), π is the nucleotide diversity, *F* is the probability that the first event in the genealogy of two chromosomes chosen at random in a subpopulation is a coalescence rather than an immigration or founder event and M is the amount of gene flow (*Nm*) in a given subpopulation

	Abbr.	N	N2	*P* (%)	H'_{eN}	H'_{eH}	π (× 1000)	*F*(± SE)	M
Upper course									
Purkuputaanniemi	PUR	198	30	25.9	0.075	0.131	8.49	0.655 ± 0.040	0.133
Nurmisaari	NUR	151	30	51.8	0.155	0.178	14.53	0.274 ± 0.032	0.677
Middle course									
Tikkuniemi	TIK	312	30	41.5	0.149	0.178	13.51	0.463 ± 0.038	0.294
Jäkälämutka	JAK	241	30	54.9	0.176	0.190	13.91	0.300 ± 0.033	0.592
Kitkajokisuu	KIT	341	28	37.3	0.092	0.154	9.66	0.548 ± 0.039	0.209
Lower course									
Kuusinkijokisuu	KUU	50	24	42.5	0.125	0.171	15.18	0.479 ± 0.037	0.275
Paanajärvi	PAA	750	30	25.9	0.114	0.152	11.54	0.501 ± 0.043	0.253
Mean		291.9	28.9	40.0	0.127	0.165	12.40	0.460	0.348

For this study we chose the amplified fragment length polymorphism (AFLP) marker system, which is a technique through which selected fragments from the digestion of total DNA are amplified by polymerase chain reaction (PCR) (Vos *et al*. 1995). The AFLP method has many advantages: there is no need for prior sequence knowledge, the repeatability is generally good, the quantity and quality of DNA requirements are small and the resulting DNA fingerprints provide a large number of genetic markers. The drawback of this method is that markers are dominant, so that heterozygotes cannot be distinguished from homozygotes.

AFLP reactions were conducted using the AFLP™ Plant Mapping kit supplied by Perkin-Elmer Applied Biosystems (USA). A kit optimized for a genome sized 500–6000 Mb was used according to the manufacturer's instructions, with slight modifications. The preselective and selective amplifications were carried out in a Perkin Elmer GeneAmp® PCR System 9700 thermocycler. From the 64 possible primer combinations six first primer pairs, which amplified a reasonable number of bands in *S. tatarica*, were used. The used primer pairs are given in Table 2. Only the selective nucleotides are shown. For testing selective amplification primer pairs reactions were performed according to the manufacturer's instructions. After the primer pairs were chosen, the total volume of the selective amplification reactions was reduced by half from the manufacturer's instructions.

Formamide (1.5 μL) was mixed with 0.5 μL of a GeneScan 500 ROX internal lane standard (Applied Biosystems) and 1.2 μL of that solution was mixed with 1 μL selective amplification product. Samples were denatured at 95 °C for 5 min. The denatured samples were analysed on an automated DNA sequencer (ABI Prism 377™). AFLP electrophenograms were analysed by GeneScan software (Applied Biosystems). AFLP genotypes were scored for presence or absence of certain fragments between 50 and 500 bp using Genotyper software (Applied Biosystems).

Table 2 Six used primer combinations, the number of bands and degree of polymorphism

Primer pairs	Total bands	Polymorphic bands	% polymorphism
E-AAG/M-CAC	85	45	52.9
E-ACT/M-CTA	57	26	45.6
E-ACA/M-CTT	72	47	65.3
E-ACA/M-CAC	50	26	52.0
E-ACA/M-CTG	48	25	52.1
E-AAG/M-CTC	49	24	49.0
Total	361	193	53.5

One plant from the KIT subpopulation was removed due to ambiguous amplification results, and five plants from the KUU subpopulation were removed because amplification was unsuccessful with a large number of primer pairs.

Genetic diversity within subpopulations and linkage disequilibrium

The estimation of allele frequencies for dominant markers presents some statistical difficulties (Lynch & Milligan 1994; Zhivotovsky 1999). Traditionally, the frequency of the null allele is estimated by taking the square root of the frequency of the null homozygote, or the absence of a band. Earlier studies have suggested that this approach may lead to statistically biased results of average estimates

of heterozygosity and population differentiation if low number of polymorphic loci were used, and at least two alternative approaches to estimate allele frequencies have been suggested (Lynch & Milligan 1994; Zhivotovsky 1999). However, Krauss (2000) has shown recently that biases are largely eliminated in highly polymorphic dominant marker data sets, and all suggested procedures to estimate allele frequencies seem to provide accurate estimates of heterozygosity with data sets with large numbers of AFLP loci and about 30 individuals per population. Accordingly, we used the traditional method to estimate allele frequencies.

Genetic diversity was quantified as (i) the percentage of within-population polymorphic loci and (ii) Nei's unbiased expected gene diversity assuming Hardy–Weinberg equilibrium (see Population structure). The genetic diversity indices were estimated separately for each locus and averaged by using the POPGEN program (Yeh *et al*. 1995). Heterozygosities within populations were also estimated (iii) using the Bayesian approach of Holsinger *et al*. (2002), which does not assume Hardy–Weinberg equilibrium within populations (see below). We also estimated (iv) the nuclear diversity (π) for each subpopulation, and the average for the total population (π_T) and subpopulations (π_S) (Innan *et al*. 1999). The nucleotide diversity indices were estimated separately for each primer pair and averaged. The sampling variance of nucleotide diversity was computed by the jackknife method following Nei & Miller (1990). Linkage disequilibrium between AFLPs was tested by using the χ^2 test following Miyashita *et al*. (1999).

Population structure

We used Pritchard *et al*.'s (2000) software 'structure' (version 2.0) to infer population structure and assign individuals to subpopulations. The analysis consists of two phases. First, the issue of model choice (i.e. how many subpopulations are most appropriate for interpreting the data) is considered without prior information of the number of locations at which the individuals were sampled and into which location each individual belongs. Second, the program assigns probabilistically the individuals in the sample to the subpopulations.

Because it is not possible to distinguish all the genotypes with AFLP data, each class of genotypes was treated as being, effectively, a haploid allele (see Pritchard *et al*. 2000). We assumed the model to be of no population admixture and that the allele frequencies are correlated within populations (D. Falush *et al*. unpublished). We conducted a series of independent runs for each value of K (the number of subpopulations) between 1 and 9. After conducting numerous runs to investigate the behaviour of the program we chose to use a burn-in period of 30 000 iterations and collect data 10^5 iterations. We ran three independent simulations of this length for each K, and found that the independent runs produced highly consistent results.

The analysis assumes that the subpopulations are in Hardy–Weinberg equilibrium, and that that the markers are unlinked. The first assumption in our subpopulations is probably appropriate (see Results). However, there seems to be more linkage disequilibrium than expected by chance alone within populations (see Results). To test the robustness of the model, we split the data randomly into two subsets with equal amounts of loci, and also ran the program with these subsets. Additionally, we ran the program with a smaller subset of 30 loci, which were unlinked in each subpopulation.

To assign individuals into subpopulations and to identify individuals who are immigrants, or who have recent immigrant ancestry in the last G generations (where $G = 0$ is the present generation) we used the model with prior information of the membership of subpopulations. To apply the method, we must first specify a value for v, the probability that an individual is an immigrant to a given subpopulation (the individual is considered to be purely from a given subpopulation with a probability of $1 - v$). We performed analyses for $v = 0.05$, 0.1 and 0.3, and with $G = 1$.

To analyse whether there had been complete isolation between subpopulations or some gene flow between them, we used the program '2mod' (Ciofi *et al*. 1999) to estimate the relative likelihoods of immigration drift equilibrium vs. drift models since a certain time. The gene-flow model assumes that the gene frequencies within subpopulations are determined by a balance between genetic drift and immigration. In the drift model it is assumed that an ancestral panmictic population separated into several independent units which start diverging purely by genetic drift. In the drift-only model it is assumed that the mutation rate u is much smaller than the reciprocal of t, the number of generations since isolation; under the gene-flow model, the mutation rate u should be much rarer than migration m, so that the descendents of each immigrant are identical by descent. These assumptions should be met in the case of *S. tatarica* if u is of the order 10^{-4} or less. The program uses a coalescent theory based on Markov's chain Monte Carlo simulation approach with Metropolis–Hastings sampling to explore the alternative models, as described in Ciofi *et al*. (1999). The program also simulates the posterior density of F, which is the probability that the first event in the genealogy of two chromosomes chosen at random in a subpopulation is a coalescence rather than an immigration or founder event. Two independent simulations with 100 000 iterations were carried out to check the convergence of the posterior probabilities of the models and F values were checked for convergence by comparing the means and time-series standard errors for the two runs. The means and standard errors for the combined output were reported.

We dropped the first 10% of points in the output to avoid dependence on initial starting values.

The presence of genetic structure in the total population was analysed using the Bayesian method suggested by Holsinger *et al*. (2002). The method allows direct estimates of F_{ST} from dominant markers without assuming previous knowledge of the degree of within-population inbreeding, and that genotypes within populations are in Hardy–Weinberg proportions. The method also provides some information about the degree of within-population inbreeding (F_{IS}), given that there are large numbers of loci and populations. We used the program HICKORY (version 0.7) with a full model and using noninformative priors for f (estimate of F_{IS}) and θ_B (estimate of F_{ST}). We conducted several runs with default sampling parameters (burn-in = 50 000, sample = 250 000, thin = 50) to ensure that the results were consistent.

F_{ST} values among subpopulations were also estimated on the basis of AFLP allele frequencies (e.g. Weir 1990), assuming that the populations were in Hardy–Weinberg equilibrium (random mating) (i.e. $F_{IS} = 0$) and (ii) assuming total selfing within subpopulations (i.e. $F_{IS} = 1$). F_{ST} estimates using these two assumptions were obtained with the POPGENE program (Yeh *et al*. 1995). Moreover, F_{ST} were also estimated on the basis of nucleotide diversities assuming Hardy–Weinberg equilibrium as:

$$F_{ST} = \frac{\pi_T - \pi_S}{\pi_T}$$

in which π_S is average pairwise diversities per nucleotide site in samples taken from within subpopulation and π_T from the total metapopulation (Charlesworth 1998).

The presence of genetic structure on subpopulation and regional levels was also tested by an analysis of variance framework using analysis of molecular variance (AMOVA; Excoffier *et al*. 1992), which is based basically on hierarchical variance of gene frequencies. However, in the case of dominant markers AMOVA partitions the genotypic variance, and not the variance of allele frequencies. When using AMOVA with dominant markers it has to be assumed that there is the same mating pattern in all subpopulations. We examined two different population structure models with AMOVA. In the first model we defined three regional groups of subpopulations: upper (consisting of PUR and NUR subpopulations), middle (TIK, JAK and KIT) and lower course (KUU and PAA) (Fig. 2). In the second group no grouping of subpopulations was defined, and the hierarchical analysis of variance partitions the total variance into covariance components due to interindividual and interpopulation differences. We used program ARLEQUIN (version 2.00; Schneider *et al*. 2000) to perform AMOVA analysis. We used this software to estimate genetic distances (Φ_{ST}) and their significances between subpopulation pairs. We also constructed an UPGMA dendrogram between the subpopulations using Miller's TPGA software (Miller 1997) and Nei's unbiased genetic distance (Nei 1978) assuming Hardy–Weinberg equilibrium. A total of 1000 permutations were used to construct multiple new data sets by resampling with replacement over loci, and the proportion of permuted data sets that result in the formation of a node seen in the original data set are reported.

We used the approach of Hutchison & Templeton (1999) to test the regional equilibrium between drift and migration. The pairwise F_{ST} values, used as genetic distances between subpopulations, were estimated using the approach of Holsinger *et al*. (2002) and assuming that $F_{IS} = 0$. Pairwise F_{ST} values and geographical distances were used to construct scatterplots of F_{ST} on geographical distances and to calculate the correlation coefficient describing the relationship between them. Because pairwise elements of distance matrices are not independent and thus violate the basic assumptions associated with standard tests of significance, a Mantel test (e.g. Manly 1985) was used to assign an estimate of the 95% upper tail probabilities for each matrix correlation coefficient. To determine whether the degree of scatter increased with the geographical distance as expected under the null hypothesis of equilibrium, the residual obtained from a standard linear regression of F_{ST} value on geographical distances were correlated with geographical distance. Significance was evaluated by subjecting the residual and geographical distance matrices to a Mantel test.

Results

AFLP analysis

In AFLP analysis from the 64 possible primer combinations six first primer pairs were used, which resulted in an amplification in *S. tatarica*. The used primer pairs, total number of bands and the degree of polymorphism are given in Table 2. Only the selective nucleotides are shown. Different primer pairs amplified variable number of bands, from 48 to 85. The percentage of polymorphic bands between primer pairs varies from 45.6% to 65.3%. Total number of bands was 361, of which 193 were polymorphic (53.5%).

Genetic diversity within subpopulations and linkage disequilibrium

Each examined individual exhibited an individual AFLP pattern. The proportion of polymorphic loci varied between 25.9% and 54.9% (Table 1). The number of private alleles was 68 (35.2%) and private alleles were found from each subpopulation. However, there was a significant heterogeneity between subpopulations in the number of

private alleles (χ^2 test with a correction for a sample size: $\chi^2 = 36.50$, d.f. = 5, $P < 0.001$). There appears to be also significant regional differences in the mean number of private alleles (χ^2 test with a correction due to different number of individuals in a sample: $\chi^2 = 32.60$, d.f. = 2, $P < 0.001$). More private alleles were found in subpopulations of the middle region (mean 17.66, SD = 1.16) than in the upper (2.5, SD = 2.12) or lower course (5.0, SD = 4.24).

The total population gene diversity (π_T) was 2.95×10^{-2} ($\pm 1.97 \times 10^{-3}$) and the average among subpopulations (π_S) was 1.24×10^{-2} ($\pm 4.04 \times 10^{-4}$). Gene and nucleotide diversities within subpopulations are reported in Table 1. The mean genetic diversity within populations assuming Hardy–Weinberg equilibrium (H_{eN}) was 0.127; the Bayesian estimate (H_{eH}) was slightly higher (0.165). However, the correlation between these estimates among subpopulations was high ($r = 0.96$, $n = 7$, $P = 0.001$). There was no significant difference in nucleotide diversity (π) between the subpopulations (Kruskal–Wallis test: $\chi^2 = 11.85$, d.f. = 6, $P = 0.066$). However, Nei's gene diversity (H_{eN}) was significantly different among subpopulations (Kruskall–Wallis test: $\chi^2 = 45.03$, d.f. = 6, $P < 0.001$). The amount of genetic variation was largest in JAK and smallest in PUR subpopulations. However, there appeared to be no consistent regional pattern in gene diversity. When the different geographical regions (upper, middle and lower course) were compared, there was no significant difference between them in genetic diversity (H_{eN}) (Kruskal–Wallis test: $\chi^2 = 4.59$, d.f. = 2, $P = 0.101$).

No significant association was detected between the estimated subpopulation size and genetic variation. The regression coefficient between the estimated subpopulation size and polymorphic loci was very low ($b = -0.021$) and statistically not significant ($t = -1.406$, d.f. = 6, $P = 0.209$). There was no significant association between either the number of private alleles and subpopulation size ($b = -0.004$, $t = -0.258$, d.f. = 6, $P = 0.807$), between subpopulation size and the average nucleotide diversity ($b = -0.004$, $t = -0.821$, d.f. = 6, $P = 0.449$) and subpopulation size and Nei's gene diversity ($b = -0.000$, $t = 0.389$, $P = 0.713$) (Fig. 3), suggesting that there was no equilibrium between drift and mutation in the subpopulations. The result was similar when the Bayesian estimate of gene diversity was used, and it is not reported here.

Linkage disequilibrium was examined by χ^2 test between all the polymorphic bands within each subpopulation. There were 26 290 combinations of which 3713 were significant, at least at the 5% level. The proportion of significant pairs (14.1%) was higher than expected by chance. However, the linkage disequilibrium appears to be sporadic. There was no specific locus pair, which would have been in disequilibrium in each or even in most of the subpopulations. No significant difference in the proportion of significant linkage disequilibrium values between populations was detected (χ^2 test: $\chi^2 = 0.000$, d.f. = 6, $P > 0.1$). At total population level there where 18 336 combinations of which 3684 (20.1%) were significant, at least at the 5% level. Accordingly, there were about three times more significant linkage disequilibrium tests at population level and four times more at whole population level than expected by change alone.

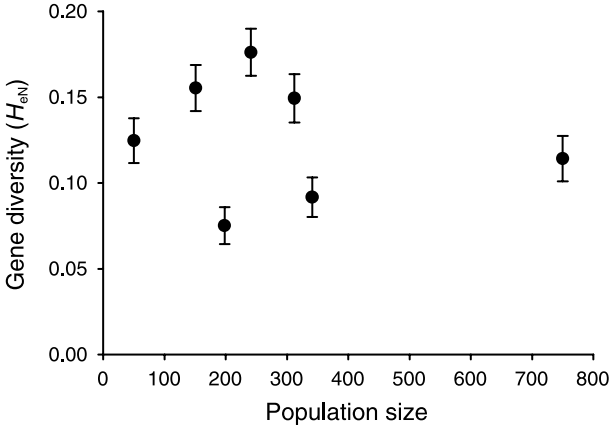

Fig. 3 The relationship between census subpopulation size (number of flowering individuals) and Nei's gene diversity.

Population structure

The estimates of the posterior probabilities of values of K (the number of subpopulations) with different values of v ($v = 0.05$, 0.1 and 0.3) using all loci gave similar results, and only the results with $v = 0.05$ are reported here. The only model which explained the data sufficiently ($P \approx 1.000$) was the model with $K = 7$. The other models ($K = 1$–6 and $K = 8$–9) were completely insufficient to model the data ($P < 0.001$ in each case). Accordingly, the data did not contain one panmictic population only, and the most probable number of clusters was seven. The highest probabilities for K using only half the loci, and also unlinked loci, were similar to those of the whole data set, suggesting that the result was robust and not due to spurious linkage disequilibrium within the subpopulations.

The likelihoods of the gene-flow model and pure drift models were 1 and 0, respectively, i.e. none of the simulations supported the drift model. This result suggests that the subpopulations of *S. tatarica* are not totally isolated units, but there is some migration between the populations. In the gene-flow model the number of migrants per generation (M) may be estimated as $(1 - F)/(4F)$ (Table 1). In most subpopulations the probabilities of common ancestry (F) were relatively high within subpopulations (0.655–0.274) and immigration into each subpopulation was low (0.133–0.677).

Table 3 Analysis of molecular variance (AMOVA) for 202 *Silene tatarica* individuals with and without regional structure

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	*P*
Regional structure					
Among groups	2	253.38	−0.420	−3.35	0.764
Among subpopulations within groups	4	620.56	4.976	39.77	> 0.001
Within subpopulations	196	1591.02	7.955	63.58	> 0.001
Total	202	2464.96	12.512		
No regional structure					
Among subpopulations	6	873.93	4.657	36.92	> 0.001
Within subpopulations	196	1591.02	7.955	63.08	> 0.001
Total	202	2464.96	12.512		

In the Bayesian analysis of population structure using the free model the posterior mean estimate for F_{IS} (f) was 0.0036 ± 0.012 (95% credible intervals 0.000 and 0.014), and the Bayesian factor analysis suggested that there was only weak evidence that f was different from 0. In fact, the criterion of model choice for model $f = 0$ was slightly smaller (*DIC* = 13126.2) than for the full model (*DIC* = 13128.4), suggesting that there is no particular reason to prefer the full model to $f = 0$. This is consistent with our preliminary analysis of parent–offspring AFLP data in a subpopulation of *S. tatarica*, in which we have found only very low levels of self-crossing (unpublished results).

Because the estimate of F_{IS} was very low the analysis gave us almost exactly similar posterior mean estimates of F_{ST} (θ_B), whether we were using the full model (0.287 ± 0.012) or the model with $f = 0$ (0.286 ± 0.012) (95% credible intervals 0.265 and 0.310 for both cases). This was lower than the traditional estimate of the overall F_{ST} value between all subpopulations (F_{ST} = 0.390 ± 0.018 and F_{ST} = 0.425 ± 0.018 when assuming total outbreeding or total inbreeding, respectively) or AMOVA estimate (ϕ_{ST} = 0.369). The F_{ST} value estimated on the basis of nucleotide diversities gave even higher values for the differentiation between the subpopulations (F_{ST} = 0.580 ± 0.020).

A summary of the AMOVA analysis is shown in Table 3. In the first model analysed with regional grouping into upper, middle and lower courses, the variance component among groups was slightly negative (–3.35%), suggesting an absence of genetic structure at this hierarchical level. The total variation among groups was not significant either (P = 0.752; 1023 permutations), confirming that the model was not adequate. In contrast, in the model with only two hierarchical levels both variance components between subpopulations and within subpopulations were highly significant ($P < 0.001$; 1023 permutations). The variation between subpopulations explained 36.9% of the total variation, and the largest amount of variation (63.1%) was within subpopulations. The pairwise ϕ_{ST} values estimated from the genotype frequencies between JAK and NUR subpopulations were significant at level P = 0.0017 (3024 permutations); all the other pairwise values were significant at level $P < 0.001$, suggesting that the genotype frequencies in different subpopulations were dissimilar.

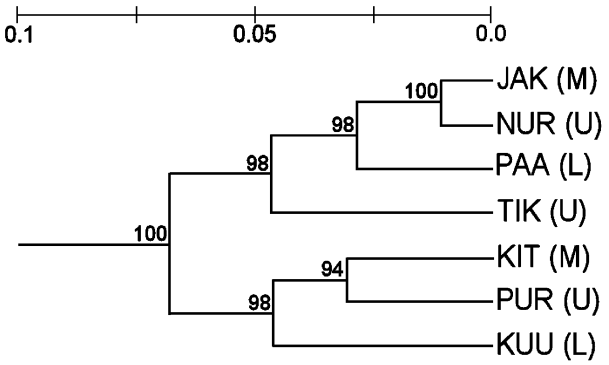

Fig. 4 UPGMA dendrogram of the subpopulations based on Nei's unbiased genetic distance. Bootstrap values (based on 1000 permutations) are indicated in each node. The subpopulation abbreviations refer to Table 1 and the region abbreviations (upper = U, middle = M and lower = L) are given in parentheses.

The UPGMA dendrogram (Fig. 4) was supported with high bootstrap values (94–100%), suggesting a high reliability of the pattern found. However, the way the subpopulations tended to cluster seemed to be independent of their geographical origin. For example, the dendrogram seems to be divided into two highly supported (100%) monophyletic clusters, and both of them consisted of subpopulations from all three regions. Moreover, geographically closely located subpopulations were not clustered together in any case.

The average pairwise F_{ST}-value estimated using the Holsinger *et al.* (2002) method was 0.310 ± 0.079. Neither the scatterplot (Fig. 5) nor the correlation analysis showed any association between pairwise F_{ST} values and geographical distance along the river ($r_M = -0.139$, P = 0.248, 1000 permutations). Similarly, there was no indication that the degree of scatter was associated with geographical

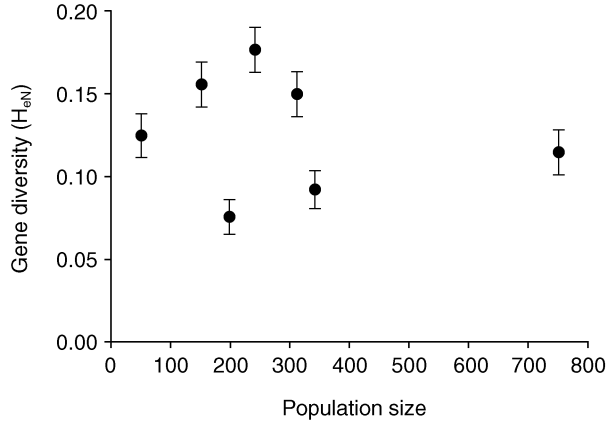

Fig. 5 Pairwise genetic (Bayesian F_{ST}-values) and geographical distances along the river.

distance (r_M = 0.000, P = 0.580, 1000 permutations). These results were similar when straight-line distances and other F_{ST} measures were used, and are not reported here.

Assignment analysis

In the assignment analysis 183 of the total of 202 (90.6%) individuals were assigned to the subpopulation from which they were sampled (Table 4), suggesting that each original sample site formed a distinct subpopulation. This result is in agreement with the significant differences in genotypic frequencies between populations, and high bootstrap values (94–100%) of the UPGMA dendrogram (Fig. 4).

Assignment success was lowest in JAK (63%) and NUR (77%) subpopulations. In NUR subpopulation eight of 30 individuals were assigned to the JAK, one to KIT and two to KUU subpopulations; seven of the individuals from JAK subpopulations were assigned to the NUR subpopulation. All these individuals had low posterior probabilities that they were not immigrants (0–0.153), a moderate probability that they were themselves immigrants (0.282–0.333) and much higher probabilities (0.565–0.667) that one of their parents was from the other subpopulation.

The assignment success was very high in the other subpopulations. In the PUR, TIK, KIT and PAA subpopulations all individuals were assigned correctly. One of the KUU individuals was assigned to the JAK subpopulation. It had a very low probability (0.090) that it was not an immigrant, and a moderate probability that it was itself an immigrant (0.303) and a reasonably high probability (0.607) that one of its parents was from the JAK subpopulation.

Discussion

Within-subpopulation diversity

In *S. tatarica* mean estimates for gene diversity of the subpopulations were 0.127 (Nei's estimate) and 0.165 (Bayesian estimate). Recently the AFLP technique has been successively used in many genetic studies with a wide range of different taxa (Travis *et al*. 1996; Gaudeul *et al*. 2000; Keiper & McConchie 2000; Mariette *et al*. 2001). *S. tatarica* seems to exhibit an intermediate level of intrapopulational genetic diversity similar to other endangered or vulnerable species obtained with AFLP-markers. For example, in an umbrella fern *Stichorus flabellatus* samples from protected and unprotected areas in Sydney region suggested a range of Nei's genetic diversity (H_{eN}) of 0.0395–0.1175 (Keiper & McConchie 2000). In a critically endangered *Astragalus* species AFLP diversity values [estimated genetic diversities ranged from 0.0373 to 0.1337 (Travis *et al*. 1996)] were comparable to those of *S. flabellatus*. The mean Nei's expected heterozygosity among 14 populations of endangered *Eryngium alpinum* species was 0.198 (Gaudeul *et al*. 2000).

The total population nucleotide diversity (π_T) in *S. tatarica* was 0.0295, which is at a similar level than nucleotide diversity estimated from AFLP data in other species (Innan *et al*. 1999; Miyashita *et al*. 1999). For example in *Arabidopis thaliana* the average nucleotide diversity is 0.0106 (Miyashita *et al*. 1999), and in soybeans *Glycine soja* and *G. max* the nucleotide diversities are 0.023 and 0.008, respectively (Innan *et al*. 1999).

Table 4 Results of the assignment analysis of 202 *Silene tatarica* individuals. The underlined figures represent possible migrants

Given subpopulation	Inferred subpopulation							
	PUR	NUR	TIK	JAK	KIT	KUU	PAA	
PUR	**30**	0	0	0	0	0	0	30
NUR	0	**19**	0	**8**	**1**	**2**	0	29
TIK	0	0	**30**	0	0	0	0	30
JAK	0	**7**	0	**23**	0	0	0	30
KIT	0	0	0	0	**29**	0	0	29
KUU	0	0	0	**1**	0	**23**	0	24
PAA	0	0	0	0	0	0	**30**	30
								202

In contrast to theoretical expectations (e.g. Nei 1987) and empirical results (Nei & Graur 1984; Frankham 1996; Travis *et al*. 1996; Muller-Scharer & Fischer 2001), we did not find any significant association between census subpopulation size and the amount of genetic diversity within the investigated subpopulations (Fig. 3). Even though migration between the subpopulations may have decreased the expected correlation, such an association has been observed in species with a similar level of gene flow between subpopulations than in *S. tatarica* (Muller-Scharer & Fischer 2001). Loss of association between population size and genetic diversity has been confirmed in other species with a nonequilibrium population structure (Schmidt & Jensen 2000). The observed pattern suggests that the subpopulations are not in mutation drift equilibrium in *S. tatarica*, and that the genetic diversity of subpopulations may reflect historical development of population sizes, for example recent population bottlenecks and expansions after them.

Relative high amounts of linkage disequilibrium within subpopulations in *S. tatarica* also suggested recent population bottlenecks or population admixture (cf. Ardlie *et al*. 2002; McVean 2002). Our field observations indicate that new subpopulations in *S. tatarica* are founded by a few individuals (A. Jäkäläniemi, unpublished), and this founder effect is the probable reason for observed linkage disequilibrium within subpopulations. Linkage disequilibrium also seems to be common in subpopulations of other species with a metapopulation structure (Stewart *et al*. 1999).

Among subpopulation differentiation and gene flow

Although most genotypic diversity was found within subpopulations (63%; Table 3), the genetic differentiation among subpopulations was relatively high. The traditional pairwise F_{ST} values between subpopulations differed only slightly when different breeding systems were assumed, suggesting that the structure was insensitive to the mating system (given that the mating pattern is similar in all subpopulations). The overall Bayesian estimate of F_{ST} (0.287) was typical of animal-pollinated, mixed selfing and outcrossing plant species (mean $F_{ST} = 0.22$; see Frankham *et al*. 2002).

Genetic surveys of metapopulations of related species have suggested less differentiation among subpopulations than in *S. tatarica*. In a circumpolar tundra species, *S. acaulis*, the estimate of F_{ST} was 0.007, suggesting little differentiation between subpopulations ≈1 km apart (Gehring & Delph 1999). In an archipelago metapopulation of *S. dioica*, the F_{ST} values were 0.057 among young subpopulations and 0.030 among old ones (Giles & Goudet 1997). In *S. latifolia* the F_{ST} values among both old ($F_{ST} = 0.197$) and recently established ($F_{ST} = 0.126$) ruderate populations appeared to be slightly lower (McCauley 1997) than in *S. tatarica*.

According to the assignment analysis there were 19 possible immigrants (even though they do not necessarily represent independent migration events) among the 220 plants, i.e. the proportion of migrants was 0.086. Assuming reciprocal and similar migration between all seven subpopulations, there are 42 (7×6) possible migration routes between subpopulations. This would give us an estimate of the average migration rate among the seven subpopulations, $m = 0.0021$. Assuming that the harmonic mean estimated census number of fertile plants (161.7) is the effective population size (N) of the subpopulations, this would suggest an estimate of 0.333 migrants (Nm) per generation between a pair of subpopulations.

The mean pairwise Bayesian F_{ST} estimate for *S. tatarica* was 0.310 (± 0.08). If the island model is assumed, this would give an estimate of $Nm = 0.556$. The mean frequency of private alleles among the subpopulations was 0.217 and the private allele method (Slatkin 1985b; see also Barton & Slatkin 1986) would suggest an estimate of $Nm = 0.147$. The estimates of the coalescent theory-based simulation method of Ciofi *et al*. (1999) suggested similar rates of migration into each subpopulation (mean $Nm = 0.348$; range 0.133–0.677; see Table 1).

The different estimates suggested a low and comparable amount of gene flow between subpopulations. When extinctions and recolonizations occur, the estimate of Nm using the private allele method in fact estimates $N(m + e)$ where e is the extinction rate (Slatkin 1985b), and this estimate is supposed to be slightly larger than the other estimates. Contrary to expectations, it was slightly lower. However, the private allele method is not appropriate for dominant markers.

The observed migration events were probably not pollen-mediated because, typically, pollen gene flow seems to be effective only within several tens of metres and diminish in a logarithmic scale with increasing distance (e.g. Richards *et al*. 1999). Thus, the long-term (over 10 km) dispersal events were due probably to zoochoric or hydrochoric dispersal of seeds. The highest number of migrants was observed between NUR and JAK subpopulations, these subpopulations being the only ones which are easily accessible to man from the gravel road parallel to the river (see Fig. 2). Day-trippers and fisherman who often visit these sites are probably spreading seeds of *S. tatarica* via their boots and equipment. This is supported by the fact that immigration was at a similar level both down- and upstream. Other possible seed vectors are reindeer staying on sandy banks of the River Oulankajoki in order to avoid blood-sucking insects (Helle & Aspi 1984), or ducks and other waterfowl. Even though both reindeer and human movement has been restricted strictly across the border zone between Finland and Russia, migration between

these areas was also bidirectional. There was a probable long-term (*c.* 12.5 km along the river) immigration event downstream from the JAK to KUU subpopulations, and also upstream from the KUU to NUR subpopulations (28.8 km). Water may have been a vector for the downstream migration, while birds would have been required as vectors for the long-term migration against the current.

In *S. tatarica*, long-distance seed dispersal by zoo- and hydrochory seems to be common enough to prevent spatial genetic structuring, even though it does not prevent subpopulation differentiation. Long-distance seed dispersal has been documented only rarely (see Cain *et al*. 2000; Nathan & Muller-Landau 2000; for recent reviews) because of methodological problems. In *Hibiscus moscheutos*, a genetic analysis has emphasized the relative importance of hydrochory into the gene flow compared to the gene flow by pollen (Kudoh & Whigham 1997). In a floodplain specialist, *Boltonia decurrens*, colonization events seem to involve long-distance dispersal of seeds drawn from a small number of source populations (J. DeWoody *et al.* unpublished).

Population structure

Our analysis of population structure shows clearly that the subpopulations did not form a single panmictic unit with free gene flow throughout the whole population (Fig. 1a). On the other hand, in the '2mod' analysis none of the simulations supported the pure drift model, suggesting that the subpopulations would not represent only the fragments of a formerly continuous population without any present-day migration between the subpopulations (Fig. 1b). However, the drift model assumes that all the subpopulations become separated at the same time. If the fragmentation process is more gradual, it will bias the result of the simulation towards the drift gene-flow model. However, our assignment analysis did not support total isolation between subpopulations either. Thus, the fragmentation model was not rejected only because of the violated assumptions of the simulation model.

The stepping-stone model (Fig. 1c) did not seem to be plausible, because there was no correlation between geographical and genetic distances and because genetic diversity was not higher near the centre of the habitat than near the ends (even though more private alleles were found in subpopulations of the middle region), as expected in the linear population models (Hutchison & Templeton 1999; Wilkins & Wakeley 2002). On the other hand, the results of the assignment analysis suggested that migration was highest between subpopulations near to each other (JAK and NUR) in a stepping-stone fashion, and that long-term migration was rare. It is possible that correlation between genetic and geographical distances may fail in a stepping-stone population because of nonequilibrium migration (Hutchison & Templeton 1999). For example, the region may have been colonized too recently from a common pool of migrants. Even though migration outside our study metapopulation may not be totally ruled out in the case of *S. tatarica*, given the geographical distribution of the species it is not very probable.

We found a relatively high level of linkage disequilibrium at total population level. Linkage disquilibrium at the total population level using the AFLP technique has been estimated rarely. However, Miyashita *et al*. (1999) estimated linkage disequilibrium in worldwide samples of *Arabidopsis thaliana*, and found that the proportion of significant linkage disequlibria was close to random expectation. Conversely, Sharbel *et al*. (2000) found significant levels of multilocus linkage disequilibrium among *A. thaliana* accessions when using a larger number of accessions than Miyashita *et al*. (1999). Linkage disequilibrium at the total population level in *S. tatarica* indicates possible metapopulation-type dynamics with recent colonizations and extinctions (Ohta 1982), even though population substructuring and limited migration may also increase the amount of linkage disequilibrium (e.g. Ardlie *et al*. 2002). Because there were no consistent differences in the amount of genetic variation between geographical areas (cf. Gaggiotti & Smouse 1996), and because the migration pattern did not suggest only one or a few source populations, the source-sink metapopulation structure in *S. tatarica* seems to be less probable than the 'classical' metapopulation structure.

In conclusion, the overall pattern of distribution of genetic variation and linkage disequilibrium within and between subpopulations did not support an equilibrium model of population structure in *S. tatarica*, but a lack of regional equilibrium with drift being more influential than gene flow. The results seem to support a 'classical' metapopulation model, even though it is not possible reject the other models completely. However, all the models presented in Fig. 1 are idealized models, and the true model describing *S. tatarica* population genetics is likely to be different to any of the analysed models.

The theoretical metapopulation analysis of Pannell & Charlesworth (2000) suggests that when there is rapid population turnover (i.e. extinction rate is higher than migration rate), we may use an estimate of F_{ST} to predict the likely mode of colonization in a metapopulation. If F_{ST} is less than about 0.5, colonists originate most probably from all the occupied patches (migrant pool model; Slatkin 1977); whereas for $F_{ST} > 0.5$ the source of colonists is more probably a single subpopulation (propagule-pool model; Slatkin 1977). Preliminary results from yearly extinction rates for subpopulations of *S. tatarica* have been estimated for years 1999–2000 and 2000–01 (A. Jäkäläniemi, unpublished) and they were high (mean $3.4 * 10^{-2}$), even though there was a high variation between the study periods. When extrapolated over the estimated average lifespan

(5–7 years) of *S. tatarica*, the estimate of extinction rate per generation ($1.7 * 10^{-1}$–$2.4 * 10^{-1}$) appears to be clearly higher than the migration rate estimated on the basis of the assignment tests ($m = 2.1 * 10^{-3}$). Our estimates of F_{ST} on the basis of AFLP fragments were 0.287–0.425, and 0.580 on the basis of nucleotide diversities. These values suggest that the model of colonization is probably intermediate between the two extreme models proposed originally by Slatkin (1977). However, a more comprehensive study with F_{ST} estimates for recently established and older populations will be required to solve this issue (see, e.g. Hanski 1999). To investigate whether there is also genetic structuring within subpopulations we need to study the distribution of genetic diversity in a smaller microspatial scale.

Acknowledgements

We are indebted to H. Innan for providing us with the computer program to estimate nucleotide diversities. We thank P. Pamilo, P. Ingvarsson and two unknown referees for useful comments on an earlier version of the manuscript. This study is supported by the University of Oulu and Academy of Finland (Population Genetics Graduate School).

References

Ardlie KG, Kruglyak L, Seielstadt M (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews: Genetics*, **3**, 299–309.

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.

Barton NH, Slatkin M (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity*, **56**, 409–415.

Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*, **13**, 202–206.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.

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

Ciofi C, Beaumont MA, Swingland IR, Bruford MW (1999) Genetic divergence and units for conservation in the Komodo dragon *Varanus komodoensis*. *Proceedings of the Royal Society of London, Series B*, **266**, 2269–2274.

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.

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

Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, **10**, 1500–1508.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.

Gaggiotti OE, Smouse PE (1996) Stochastic migration and maintenance of genetic variation in sink populations. *American Naturalist*, **147**, 919–945.

Gaudeul M, Taberlet P, Till-Bottraud I (2000) Genetic diversity in an endangered alpine plant, *Eryngium alpinum*, L. (Apiacae), inferred from amplified fragment length polymorphism markers. *Molecular Ecology*, **9**, 1625–1637.

Gehring JL, Delph LJ (1999) Fine scale genetic structure and clinal variation in *Silene acaulis* despite high gene flow. *Heredity*, **82**, 628–637.

Giles BE, Goudet J (1997) Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *American Naturalist*, **149**, 507–526.

Giles BE, Lundqvist E, Goudet J (1998) Restricted gene flow and subpopulation differentiation in *Silene dioica*. *Heredity*, **80**, 715–723.

Hanski I (1999) *Metapopulation Ecology*. Oxford University Press, Oxford.

Helle T, Aspi J (1984) Do sandy patches help reindeer against insects? *Reports of the Kevo Subarctic Research Station*, **19**, 57–62.

Holsinger KE, Lewis PO, Dipak KD (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, **11**, 1157–1164.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Ingvarsson PK, Giles BE (1999) Kin-structured colonization and small-scale genetic differentiation in *Silene dioica*. *Evolution*, **53**, 605–611.

Innan H, Terauchi R, Kahl G, Tajima F (1999) A method for estimating nucleotide diversity from AFLP data. *Genetics*, **151**, 1157–1164.

Keiper FJ, McConchie R (2000) An analysis of genetic variation in natural populations of *Sticherus flabellatus* [R. Br. (St John)] using amplified fragment length polymorphism (AFLP) markers. *Molecular Ecology*, **9**, 572–581.

Kimura M (1953) 'Stepping-stone' model of population. *Annual Report of the National Institute of Genetics, Japan*, **3**, 62–63.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.

Koutaniemi L (1984) The role of ground frost, snow cove, ice break-up and flooding in the fluvial processes of the Oulankajoki River, NE Finland. *Fennia*, **162**, 127–161.

Krauss SK (2000) Accurate genetic diversity estimates from amplified fragment length polymorphism. *Molecular Ecology*, **9**, 1241–1245.

Kudoh H, Whigham DF (1997) Microgeographic genetic structure and gene flow in *Hibiscus mocheutos* (Malvaceae) populations. *American Journal of Botany*, **84**, 1285–1293.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Manly BJF (1985) *The Statistics of Natural Selection on Animal Populations*. Chapman & Hall, London.

Mariette S, Chagne Lezier C, Pastuskzka P *et al.* (2001) Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. *Heredity*, **86**, 469–479.

McCauley DE (1997) The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity*, **88**, 257–263.

McVean GAT (2002) A genealogical interpretation of linkage disequilibrium. *Genetics*, **162**, 987–991.

Miller MP (1997) *Tools for Population Genetic Analyses (TFPGA). A Windows® Program for the Analysis of Allozyme and Molecular Population Genetic Data Version 1.3*. Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ.

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

Muller-Scharer H, Fischer M (2001) Genetic structure of the annual weed *Senecio vulgaris* in relation to habitat type and population size. *Molecular Ecology*, **10**, 17–28.

Nathan R, Muller-Landau HC (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution*, **15**, 278–285.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

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

Nei M, Graur D (1984) Extent of protein polymorphism and the neutral theory. *Evolutionary Biology*, **17**, 73–118.

Nei M, Miller JC (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics*, **125**, 873–879.

Ohta T (1982) Linkage disequilibrium with the island model. *Genetics*, **101**, 139–155.

Pannell JR, Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and colonization. *Evolution*, **53**, 664–676.

Pannell JR, Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society London, Series B*, **355**, 1851–1864.

Pritchard JK, Stephans M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Richards CM, Church S, McCauley DE (1999) The influence of population size and isolation on gene flow by pollen in *Silene alba*. *Evolution*, **53**, 63–73.

Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology*, **5**, 69–76.

Schmidt K, Jensen K (2000) Genetic structure and AFLP variation of remnant populations in the rare plant *Pedicularis palustris* (Schropulariacae) and its relation to population size and reproductive components. *American Journal of Botany*, **87**, 678–689.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: a Software for Population Genetics Data Analysis*, Version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.

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

Slatkin M (1977) Gene flow and genetic drift in species subject to frequent local extinctions. *Theoretical Population Biology*, **12**, 253–262.

Slatkin M (1985a) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**, 393–430.

Slatkin M (1985b) Rare alleles as indicators of gene flow. *Evolution*, **39**, 53–65.

Stewart WA, Dallas JF, Piertney SB *et al.* (1999) Metapopulation genetic structure in the water vole, *Arvicola terrestris*, in NE Scotland. *Biological Journal of the Linnean Society*, **68**, 159–171.

Tilman P, Kareiva P (1997) *Spatial Ecology: the Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, NJ.

Travis SE, Manchinski J, Keim P (1996) An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. *Molecular Ecology*, **5**, 735–745.

Ulvinen T (1997) Tataarikohokki — ryssglim. In: *Uhanalaiset Kasvimme*, (eds Ryttäri T, Kettunen T), p. 246. Suomen Ympäristökeskus and Kirjayhtymä Oy, Helsinki.

Vos P, Hogers R, Bleeker M *et al.* (1995) A new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Weir BS (1990) *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Wilkins JF, Wakeley J (2002) The coalescent in a continuous, finite, linear population. *Genetics*, **161**, 873–888.

Wright S (1969) *Evolution and Genetics of Populations*, Vol. 2. *The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Yeh F, Yang RC, Boyle T (1995) *POPGENE, the User-Friendly Shareware for Population Genetic Analysis*. Molecular Biology and Biotechnology Center, Universty of Alberta, Edmonton.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

The present study is part of N. Tero's PhD thesis on *Silene tatarica* metapopulation genetics funded by the University of Oulu and Academy of Finland (Population Genetics Graduate School). Her research was performed under the supervision of Jouni Aspi (associate professor in genetics at the University of Oulu) and Pirkko Siikamäki (director of the Oulanka Biological Station). Anne Jäkäläniemi is a PhD student working in Oulanka Biological Station on *Silene tatarica* metapopulation ecology. Juha Tuomi is Professor in plant ecology at the University of Oulu.
