

The relative importance of factors determining genetic drift: mating system, spatial genetic structure, habitat and census size in *Arabidopsis lyrata*

Yvonne Willi^{1,2} and Kirsti Määttänen¹

¹Institute of Integrative Biology, Plant Pathology, ETH Zürich, 8092 Zürich, Switzerland; ²Institute of Biology, Evolutionary Botany, University of Neuchâtel, 2009 Neuchâtel, Switzerland

Summary

Author for correspondence:

Yvonne Willi

Tel: +41 32 718 23 61

Email: yvonne.willi@unine.ch

Received: 5 September 2010

Accepted: 24 October 2010

New Phytologist (2010)

doi: 10.1111/j.1469-8137.2010.03569.x

Key words: genetic diversity, inbreeding, outcrossing, plant mating system, self-fertilization, small population size, small-scale spatial genetic structure.

• The mating system, dispersal and census size are predicted to determine the magnitude of genetic drift, but little is known about their relative importance in nature.

• We estimated the contributions of several population-level features to genetic drift in 18 populations of *Arabidopsis lyrata*. The factors were outcrossing rate, within-population spatial genetic structure, census size and substrate type. The expected heterozygosity (H_E) at 10 microsatellite loci was taken to reflect the effective population size (N_e) and the strength of genetic drift.

• The mating system explained most of the variation in H_E (60%), followed by substrate (10%), genetic structure (9%) and census size (6%). The most outcrossing population had a +0.32 higher predicted H_E than the most selfing population; the estimated N_e of selfing populations was less than half that of outcrossing populations. Rocky outcrops supported populations with a +0.14 higher H_E than did sandy substrates. The most structured population had a +0.24 higher H_E than the least structured population, and the largest population had a +0.18 higher H_E than the smallest population.

• This study illustrates the importance of outcrossing, genetic structure and the physical environment – together with census size – in maintaining H_E , and suggests that multiple population-level characteristics influence N_e and the action of genetic drift.

Introduction

Genetic drift is a major determinant of population genetic diversity. Drift is a stochastic force that shapes within- and among-population genetic variation at neutral and weakly adaptive portions of the genome (Wright, 1931). The magnitude of genetic drift is inversely proportional to the effective population size N_e (Wright, 1931; Kimura, 1955). N_e refers to an ideal population consisting of monoecious diploid individuals with discrete, nonoverlapping generations that reproduce by random sampling of gametes (Wright, 1931). Hence, N_e depends not only on the number of reproducing individuals – one of its better studied aspects – but also on factors such as the mating system and the extent of dispersal (reviewed in Caballero, 1994). Our main goal in this study was to assess the simultaneous importance of three factors predicted to have an impact

on genetic drift via their effect on N_e – the mating system, spatial genetic structure and census size – in relatively isolated populations of a plant species over a large range of its distribution.

The mating system of animals and plants can vary from fixed for obligate outcrossing to mixed mating to entirely selfing (Schemske & Lande, 1985; Goodwillie *et al.*, 2005; Jarne & Auld, 2006). Selfing can be considered as an extreme form of limited gamete dispersal, causing increased levels of homozygosity and linkage disequilibrium (Charlesworth & Wright, 2001; Glémin *et al.*, 2006; Wright *et al.*, 2008). In a completely selfing population, all individuals should be fully homozygous for all loci and N_e is reduced twofold (Pollak, 1987). N_e may be further reduced in selfing populations because of enhanced levels of linkage when there are selective sweeps involving advantageous mutations or selection against deleterious mutations (Charlesworth & Wright,

2001). From studies of selfing plants, such as *Arabidopsis thaliana*, we know that all individuals at a particular locality are often completely fixed for the same alleles at marker loci, and even variable marker loci are homozygous in most individuals (Kuittinen *et al.*, 1997; Bomblies *et al.*, 2010).

A further factor enhancing genetic drift is limited gene flow, which means that reproduction may be random within demes or subgroups, but is nonrandom over the whole (meta-)population. Limited gene flow causes spatial genetic structure, with the predicted consequence of increased homozygosity within subgroups and linkage disequilibrium across subgroups (reviewed in Felsenstein, 1976). Spatial genetic structure over a continuous area reduces local N_e , but may increase effective size over the whole population compared with a nonstructured population with an equal number of individuals. Whitlock & Barton's (1997) model revealed that subdivision is expected to increase genetic variation only if the size and contribution of subgroups to the dynamics of the whole population are similar. So far, empirical results on the impact of within-population spatial genetic structure on genetic diversity across demes are rare. However, it is known that the mating system may interfere with spatial genetic structure in its effect on N_e . Selfing plant species show more spatial genetic structure than outcrossing ones (reviewed in Vekemans & Hardy, 2004).

An additional factor causing genetic drift is small census size. In a nonstructured, panmictic population of monoecious, diploid individuals, drift is predicted to scale with census size (N) by $1/2N$ (Wright, 1931). This is the best studied source of genetic drift. Many studies of natural populations report a positive relationship between census size and neutral genetic marker diversity, reflecting the increased action of drift under small N (reviewed in Ellstrand & Elam, 1993; Frankham, 1996; Willi *et al.*, 2007). Moreover, in the laboratory, enforcing inbreeding via small census size causes declining quantitative genetic variation for nonfitness traits (Van Buskirk & Willi, 2006; Willi *et al.*, 2006).

In this study, we estimate the impact of the mating system, the extent of within-population spatial genetic structure, census size and substrate type on genetic drift across 18 populations of the rockcress *Arabidopsis lyrata*. *Arabidopsis lyrata* has a (mostly) functioning sporophytic self-incompatibility system, which means that sporophytic (diploid) information on the pollen coat determines whether the (diploid) stigma accepts the pollen to fertilize ovules (Takayama & Isogai, 2005). However, there is variation within and among populations in the occurrence of self-incompatibility (Mable *et al.*, 2005; Mable & Adam, 2007). Mable & Adam (2007) described populations around the Great Lakes of North America that were nearly fixed for either self-compatibility or predominant selfing. For this reason, we worked with North American populations, two of which were identical to populations studied in

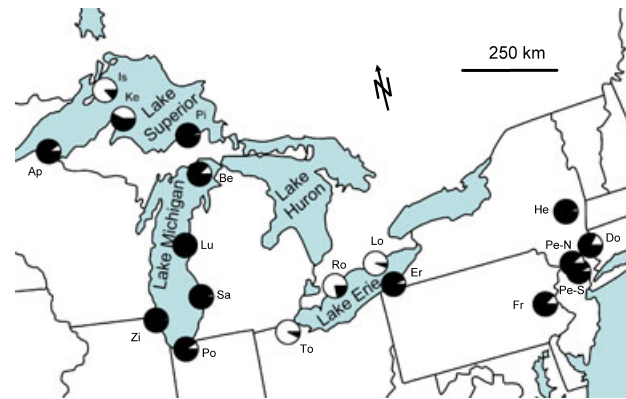


Fig. 1 Geographic distribution of 18 populations of *Arabidopsis lyrata* in North America, varying in the mating system from predominant outcrossing to mixed mating and predominant selfing (reproduced with permission from Willi & Määtänen (2010), copyright European Society for Evolutionary Biology, published by Blackwell Publishing). The black shading of the circles represents the population multilocus outcrossing rate t_m , the white shading the selfing rate ($1-t_m$). The first two letters of the site name are given beside the circles (see Table 1).

Mable & Adam (2007) (Fig. 1). A genetic analysis of the 18 populations studied here indicated that they comprised two ancestral clusters, eastern and western, and that both clusters contained selfing populations (Willi & Määtänen, 2010). Populations of *A. lyrata* in this region vary in surface area from a hundred square meters to a hundred hectares, and the species occurs in two distinct habitat types: sand and rock. Hence, *A. lyrata* shows natural variation in several population features suspected to influence N_e and genetic drift. This study evaluates their relative importance by assessing their relationship to the expected heterozygosity H_E , which depends solely on N_e and mutation rate in the absence of substantial gene flow (Crow & Kimura, 1970; Ohta & Kimura, 1973). Indeed, gene flow among populations of *A. lyrata* is relatively low. Gaudel *et al.* (2007) reported significant genetic differentiation over distances of a couple of hundred meters in Scandinavian populations, and we found that our two closest outcrossing populations, only 2.5 km apart, had a significant pairwise F_{ST} of 0.13 (Willi & Määtänen, 2010).

Materials and Methods

Plant material and census size

In summer 2007, we collected plant and seed material from 15 populations of *Arabidopsis lyrata* L. O'Kane and Al-Shehbaz and one population of *Arabidopsis thaliana* L. Heynh. in the Great Lakes region of the USA (Fig. 1; Table 1). At each site, green tissue and at least one ripe fruit (silique) from 30 plants were collected at 5-m intervals along three parallel transects separated by 5 m, over a grid area of

Table 1 The 18 *Arabidopsis lyrata* and one *Arabidopsis thaliana* populations, sorted according to their multilocus outcrossing rate t_m

Site name	Lat. N/ Long. W	Sub.	Area (m ²)	Density (m ⁻²)	Census size	N fam./ N F1	$t_m \pm SD$	N F0	m	$H_E \pm SD$
Ludington, MI	44.01/86.48	Sand	1 633 500	6.8	11 107 800	30/181	1.030 $\pm$ 0.055	30	-0.00494	0.366 $\pm$ 0.226
Zion, IL	42.42/87.80	Sand	1 727 250	3.6	6 218 100	30/178	0.974 $\pm$ 0.045	30	-0.04639*	0.499 $\pm$ 0.234
Saugatuck, MI	42.70/86.20	Sand	31 956	4.9	157 650	30/181	0.966 $\pm$ 0.021	31	-0.04437**	0.484 $\pm$ 0.243
Pictured Rocks, MI	46.67/86.02	Sand	750	0.9	700	35/201	0.961 $\pm$ 0.035	35	-0.03117*	0.255 $\pm$ 0.264
Helderberg, NY	42.66/74.02	Rock	400	0.4	164	26/150	0.957 $\pm$ 0.032	26	-0.06401**	0.464 $\pm$ 0.230
Peekskill South, NY	41.30/73.98	Rock	250	1.7	417	28/167	0.934 $\pm$ 0.023	30	-0.01186	0.435 $\pm$ 0.293
Erie, A. <i>lyrata</i> , PA	42.17/80.07	Sand	8340	34.0	283 560	30/177	0.934 $\pm$ 0.058	30	-0.01460	0.253 $\pm$ 0.245
Portage, IN	41.61/87.19	Sand	42 181	3.5	146 227	32/178	0.907 $\pm$ 0.040	32	-0.03163*	0.440 $\pm$ 0.287
Apostle Islands, WI	46.73/90.81	Sand	1366	4.7	6375	30/182	0.903 $\pm$ 0.034	30	-0.06325**	0.429 $\pm$ 0.260
Peekskill North, NY	41.32/73.99	Rock	90	2.0	182	26/156	0.881 $\pm$ 0.036	26	-0.01815	0.405 $\pm$ 0.322
Beaver Island, MI	45.75/85.50	Sand	8129	5.1	41 187	31/185	0.878 $\pm$ 0.037	31	-0.03913(*)	0.351 $\pm$ 0.283
Friedensville, PA	40.57/75.40	Sand ¹	6333	1.2	7600	31/182	0.870 $\pm$ 0.090	30	-0.01428	0.214 $\pm$ 0.273
Dover Plains, NY	41.73/73.56	Sand ¹	10 362	35.5	367 506	31/181	0.827 $\pm$ 0.118	31	0.01732	0.283 $\pm$ 0.215
Keweenaw, MI	47.38/87.96	Sand	750	1.5	1150	30/185	0.574 $\pm$ 0.065	30	-0.11513**	0.251 $\pm$ 0.217
Rondeau, ON	42.26/81.85	Sand	466 200	3.9	1 802 640	30/162	0.215 $\pm$ 0.192	(30)	0.05673	0.080 $\pm$ 0.110
Isle Royale, MI	48.16/88.45	Rock	2550	0.4	1037	33/183	0.134 $\pm$ 0.045	34	-0.05057	0.136 $\pm$ 0.219
Toledo, OH	41.62/83.79	Sand	2250	7.6	17 100	22/106	0.075 $\pm$ 0.124	(22)	-0.02298	0.100 $\pm$ 0.183
Long Point, ON	42.58/80.39	Sand	586	5.7	3315	15/56	0.056 $\pm$ 0.040	(15)	-0.04983	0.053 $\pm$ 0.145
Erie, A. <i>thaliana</i> , PA	42.01/80.38	Sand	5618	10.6	59 546	29/165	0.050 $\pm$ 0.290	24	-0.11891*	0.051 $\pm$ 0.085

¹At Friedensville, the substrate was gravel, whereas, at Dover Plains, the substrate was a mixture of mostly sand and some eroding limestone. The table lists site names with the State/Province in North America, geographic location, substrate type, area of continuous occurrence, local density, census size (area $\times$ local density), number of seed families and offspring assessed per population for estimating t_m (N fam./N F1), t_m with standard deviation (SD), number of plants sampled in the field for population genetic analysis (N F0, inferred genotypes in parentheses), coefficient of spatial genetic structure m (P values of randomization test: (*), $P < 0.1$; *, $P < 0.05$; **, $P < 0.01$) and expected heterozygosity H_E averaged over 10 microsatellite loci with standard deviation.

10 m $\times$ 45 m. If no plant with ripe fruits was found within 2.5 m of a grid point, we located a replacement plant along a 5-m extension of one of the transects. Transect sampling could not be applied in three populations of *A. lyrata* occurring on rocky outcrops, because plants grew in patches. There, plants were sampled such that distances within patches of occurrence were maximized and the combined surface area of the patches sampled was *c.* 450 m². We noted the location of each plant by measuring its distance and angle from the nearest grid point, so that we could precisely reconstruct the spatial sampling scheme. Seeds from three additional populations at Long Point and Rondeau, ON, Canada, and Toledo, OH, USA, were kindly provided by B. Mable. Her material was collected in a comparable manner, although over a somewhat larger area.

The census size at each site was estimated as the surface area occupied by the species multiplied by a measure of mean local density. This rough estimate of census size combined information on the area of the available habitat and local habitat suitability, two measures that made sense in both sandy and rocky habitats. The surface area was estimated by mapping all patches with *Arabidopsis* and summing their area. Patch boundaries were recognized by an absence of aggregations of plants (< 20) for at least 100 m. In 14 of the 18 populations, the distance to the next site was actually many hundreds of meters or kilometers, based on our surveys and local species' distribution

databases. Density estimates came from the three sampling transects. At each of the 30 grid points, all bolted plants were counted over a surface area of 0.25 m². For the smallest three populations, we produced a more precise estimate of census size by counting all bolted plants. These data were collected in June/July 2007 for most populations and in June 2009 for Long Point, Rondeau and Toledo. We revisited Zion, Portage, Saugatuck and Ludington in 2009 for more precise mapping. Populations were generally past peak flowering at the time at which we estimated the density.

Mating system, genetic structure and genetic diversity

Microsatellite genotyping was used to reveal the mating system and extent of spatial genetic structure within populations, and to assess genetic diversity. The analysis of the mating system involved both field-collected plant material and offspring raised from seeds of these plants. The other two analyses involved only field-collected maternal plants, except for three populations (Long Point, Rondeau and Toledo), for which we had no maternal plant material, but reconstructed the most likely maternal parent based on offspring genotypes.

Microsatellite genotyping We genotyped maternal plants, for which tissue had been dried in silica gel immediately after

sampling, together with six offspring per maternal plant that were reared in the glasshouse (Willi & Määttänen, 2010). Plant tissue was disrupted and DNA was isolated according to the DNeasy 96 Plant Kit protocol (QIAGEN, Hombrechtikon, Switzerland). Genotyping for population genetic analysis involved 10 microsatellite loci, 8–10 of which were used to genotype also the progeny array. The loci were ADH1, AthDET1, AthZFP6, ATTSO392, ELF3, F20D22, ICE12, ICE13, ICE14 and LYR417 (Clauss *et al.*, 2002; Mable & Adam, 2007; Muller *et al.*, 2008). Reactions were run and fragments were analyzed according to Willi & Määttänen (2010).

Pre-analysis of microsatellite data Maternal genotypes were used to check for pairwise linkage and the presence of null alleles. Linkage among the 10 loci was evaluated using the program Genepop v4.0.10 (Raymond & Rousset, 1995). Tests for genotypic linkage disequilibrium revealed no significant linkage patterns (two of 45 pairwise comparisons of genotypic linkage disequilibrium between loci with $0.05 > P > 0.01$, all others $P > 0.05$; Markov chain parameters: dememorization 10 000; batches 100; iterations per batch 5000). We checked all population–locus combinations for the presence of significant null alleles with the program INEst, which simultaneously estimates null alleles and inbreeding coefficients (Chybicki & Burczyk, 2009). We excluded monomorphic loci (see Supporting Information Table S1) and implemented the individual inbreeding model of INEst with 10 000 iterations. Across 123 polymorphic *A. lyrata* population–loci combinations, 17 population–loci showed significant evidence of null alleles (Table S1).

Mating system We used genotypes of the sporophytes and offspring to estimate the proportions of offspring produced by outcrossing and selfing for each population separately. Loci that showed significant evidence of null alleles or were not polymorphic within a particular population were excluded from the analysis. Multilocus outcrossing rates t_m were determined by inference of parentage based on the most likely parent with the program MLTR v3.2 (Ritland, 2002). To estimate standard deviations, the bootstrap number of resampling entire families was set at 1000. This analysis also revealed the inferred maternal genotypes for the three populations of Long Point, Rondeau and Toledo.

Within-population spatial genetic structure We measured the degree of genetic structure over a spatial range of 3–47 m, the distance range sampled by our grid/transect approach, by estimating the rate at which relatedness declined with distance. Moran's I was calculated for all pairs of maternal plants and regressed against the log_e-transformed distance between plants (cm) with the program SPAGeDi v1.3 (Hardy & Vekemans, 2002; jackknifed mean estimates;

P values based on 10 000 permutations of locations). The slope of the regression line, here called the coefficient of spatial genetic structure m , was interpreted as the strength of genetic structure over the spatial scale sampled (Vekemans & Hardy, 2004). Moran's I is a product-moment correlation coefficient that is not influenced by the selfing rate (Hardy & Vekemans, 1999). In one population, Erie, PA, we sampled and genotyped an extra 30 plants at a second 10 m × 45 m grid, 100 m away from the first grid, to determine the repeatability of our estimate of spatial genetic structure.

It could be argued that high gene diversity facilitates the detection of genetic structure, and therefore we tested whether m , the relationship between Moran's I and the log-transformed geographic distance, was influenced by the amount of gene diversity at each site. The analysis asked whether Moran's I for all pairs of plants within all populations and for all loci was influenced by geographic distance, H_E at the level of locus/population and the interaction between distance and H_E . Population and locus were crossed random subjects. A distance-by- H_E interaction would indicate that the degree of spatial genetic structure in a population was dependent on H_E . The analysis was performed with the lme4 package in R (Baayen *et al.*, 2008).

Genetic diversity We estimated H_E for each locus after adjusting for sample size (SPAGeDi v1.3). In addition, we assessed the inbreeding coefficient F_{IS} (INEST) and allelic richness and number of private alleles adjusted for unequal sample sizes by rarefaction (HP-RARE; Kalinowski, 2005) (Table S1). For diploids, H_E corresponds to Nei's gene diversity H_s (Nei, 1973). Over the course of acquiring the genotypic data, it became clear that rock populations, despite their small size, had high genetic diversity, and therefore substrate type was included as a factor in the analyses. In addition, a parallel study of population divergence discovered that the 18 *A. lyrata* populations comprised two clusters, western and eastern, with the split through Lake Erie and evidence for a contact zone on Lake Superior (Willi & Määttänen, 2010). Therefore, we checked whether H_E differed between the two clusters, and tested whether population means of H_E were dependent on the outcrossing rate, spatial genetic structure, substrate type and census size using type III sum of squares (proc GLM with centered covariates; SAS Institute, 2002). We evaluated the independence of the covariates by Pearson or Spearman correlations, and tested all variables for spatial autocorrelation using Mantel tests (program zt; Bonnet & Van de Peer, 2002; 10 000 permutations). The extent to which H_E was reduced in selfing populations was calculated and analyzed according to Foxe *et al.* (2010).

We checked whether a lack of spatial genetic structure was the result of a recent decline in N_e , possibly caused by local disturbance. The principle of the analysis is that, after a decline, H_E is higher than the equilibrium H_E expected

from the observed number of alleles (Bottleneck v1.2.02; Cornuet & Luikart, 1996). We assumed a two-phase mutation model with 70% stepwise mutations and a variance of 30 (10 000 iterations). We excluded loci that were monomorphic (or with a second allele occurring only once) or had significant null allele frequencies (Table S1). For the outcrossing populations ($t_m > 0.8$) and the one mixed mating population ($t_m = 0.6$), the assumption of Hardy–Weinberg equilibrium (HWE) was fulfilled (tested with Genepop; Markov chain parameters: dememorization 10 000; batches 100; iterations per batch 5000), with P values > 0.1 , except Erie with $P = 0.0869$. By contrast, the four populations with the highest mean selfing rate had significant deviations from HWE because of an excess of homozygotes (Rondeau: $P = 0.0022$; Isle Royale: $P < 0.0001$; Toledo: $P = 0.0008$; Long Point: $P = 0.0100$).

Results

The mating system, spatial genetic structure, habitat and census size

Thirteen of the 18 *A. lyrata* populations had a multilocus outcrossing rate $t_m > 0.8$, and therefore can be considered to be largely outcrossing (Table 1). Three populations were predominantly selfing ($t_m < 0.2$), one was on the verge of predominant selfing ($t_m < 0.215$) and another had a value typical of mixed mating ($t_m = 0.574$). The selfing and mixed mating populations occurred around Lake Erie and on Lake Superior, and both lakes also had outcrossing populations on their shores (Fig. 1). The single population of *A. thaliana* had an outcrossing rate of nearly zero, confirming the selfing reproductive mode of this species (Table 1).

Spatial genetic structure was present in some populations of *A. lyrata* and in *A. thaliana* (Table 1). One population, Erie, was assessed for spatial genetic structure in two places separated by 100 m, and coefficients of spatial genetic structure were similar (m , the relationship between Moran's I and the log-transformed geographic distance $\pm$ standard error: -0.019 ± 0.008 and -0.015 ± 0.020). For three of the rock populations, we had good sample sizes at distances below 3 m and could therefore explore the consequences of sampling the other populations only over the range of 3–47 m (Fig. 2). Helderberg, but not the two Peekskill populations, showed evidence for small-scale spatial structure, and this was similar when considering only the range of 3–47 m (Table 1). In addition, spatial genetic structure was not influenced by gene diversity: we detected no interaction between log-transformed geographic distance and H_E at a locus in their effect on Moran's I ($P > 0.8$). Population census size varied over nearly five orders of magnitude, from *c.* 180 plants to a few million (Table 1).

The features of *A. lyrata* populations were largely unrelated to one another, except for census size and substrate type

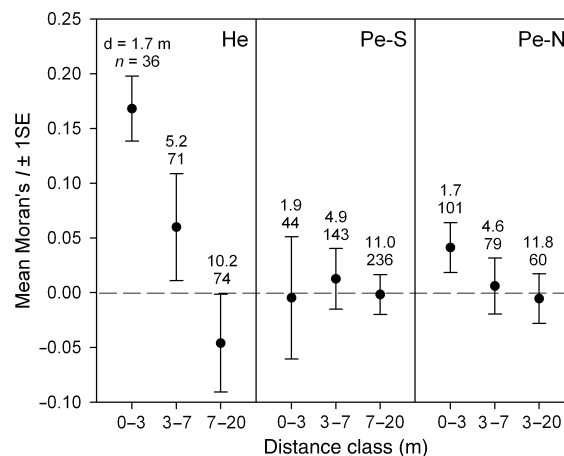


Fig. 2 Mean Moran's I of three distance classes (0–3 m, 3–7 m and 7–20 m; jackknifed estimates) for three rock populations of *Arabidopsis lyrata*: Helderberg (He), Peekskill-South (Pe-S) and Peekskill-North (Pe-N). For each distance class, the mean spatial distance (d) and sample size (n) are reported. Small-scale spatial genetic structure was present in Helderberg, but not in the two Peekskill populations.

(Table S2). Rock populations were significantly smaller than sand populations ($N = 18$, $t = 3.33$, $P = 0.0042$; Fig. 3a) because they covered smaller surface areas and tended to have lower densities, the latter as a result of plants growing only in crevices (Table 1). The estimates of multilocus outcrossing rate and spatial genetic structure were not significantly correlated with each other ($r = -0.05$, $P > 0.8$). H_E and the four population features showed no significant evidence of spatial autocorrelation (coefficients: -0.09 for H_E , -0.10 for t_m , 0.09 for spatial genetic structure, 0.06 for census size and 0.27 for substrate; all $P > 0.01$; Bonferroni-adjusted $\alpha = 0.01$). The trend was that nearby populations were somewhat similar in substrate type.

Recent demographic history estimated from genetic markers was unrelated to the variation in spatial genetic structure. Four of the 18 populations, Zion, Saugatuck, Helderberg and Erie, were judged to have declined recently because they showed gene diversity excess across loci, although these would not be significant under Bonferroni correction (one-tailed Wilcoxon tests; $P = 0.019$, 0.019 , 0.020 and 0.047 ; Bonferroni-corrected $\alpha = 0.0125$). There was no association between the coefficient of spatial genetic structure and evidence for bottleneck (no = 0, yes = 1; Mann–Whitney test, $N = 18$, $Z = -1.01$, $P > 0.1$). Indeed, the difference in means was opposite from that expected if the absence of genetic structure was caused by recent bottlenecks.

The relative importance of mating system, genetic structure, habitat and census size on genetic variation

H_E was not significantly different between the eastern and western clusters of populations ($P > 0.5$), and therefore

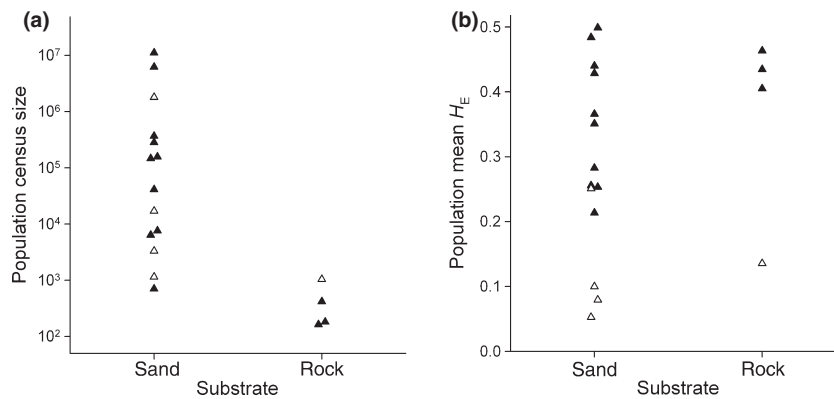


Fig. 3 Census size (a) and mean heterozygosity H_E across 10 microsatellite loci (b) of 18 populations of *Arabidopsis lyrata* on two types of substrate (sand and rock). Closed symbols, outcrossing populations with $t_m > 0.8$; open symbols, populations with a selfing or mixed mating reproductive mode. Rock populations are significantly smaller in census size but, given their size, harbor high levels of gene diversity.

cluster was not included in further analyses. Table 2 summarizes test statistics for the relationship between H_E and the mating system, spatial genetic structure, substrate type and census size. H_E differed significantly between substrate types, and was positively associated with multilocus outcrossing rate and census size, and negatively associated with the coefficient of spatial genetic structure. Rock populations had higher levels of H_E (Fig. 3b). More outcrossing, more structure and larger census size implied greater heterozygosity (Fig. 4). The proportion of variation in type III sum of squares of H_E explained by the four independent variables indicated that the mating system was more important than the other three (t_m , 60%; substrate type, 10%; m , 9%; census size, 6%). Predicted differences in H_E between populations with the most extreme parameter values were calculated on the basis of estimates from the fitted model. The predicted difference in H_E between the most selfing and the most outcrossing population was +0.32, between sand and rock populations +0.14, between the most spatially structured population and the least structured –0.24, and between the smallest and the largest population +0.18. Results from only the outcrossing populations revealed somewhat stronger effect

sizes (Table 2), possibly because selfing populations exhibited little variation in H_E .

In secondary analyses, we investigated how the mating system influenced several additional population genetic parameters (Table S3). Selfing and mixed mating populations had a significantly lower fraction of polymorphic loci than outcrossing populations (mean $\pm$ SE for selfing/mixed mating vs outcrossing populations: 0.42 ± 0.10 vs 0.81 ± 0.03 ; exact Wilcoxon two-sample tests, $P = 0.002$; Bonferroni-adjusted $\alpha = 0.01$), together with reduced mean allelic richness R_a (1.46 ± 0.13 vs 2.86 ± 0.16 ; $P = 0.000$) and higher inbreeding coefficients F_{IS} (0.560 ± 0.079 vs 0.063 ± 0.026 ; $P = 0.000$). The two types of population did not differ significantly in the number of private alleles per locus ($P = 0.025$). There was a linear relationship between H_E and t_m , even after correcting for spatial genetic structure, substrate type and census size, and after adjusting for inbreeding caused by selfing [$H_{E \text{ adjusted}} = H_E (1 + \{(1-t_m)/[2-(1-t_m)]\})$]. The estimated slope was $0.214 (\pm 0.055 \text{ SE}, N = 18, t = 3.90, P = 0.0013)$, and the mean $H_{E \text{ adjusted}}$ values for the four selfing and the 13 outcrossing populations were 0.209 and 0.386, respectively. This shows that selfing reduced N_e by more than two-fold relative to outcrossing.

Table 2 Linear models testing the effect of population multilocus outcrossing rate (t_m), coefficient of spatial genetic structure (m ; negative values indicate high structure), substrate type and census size on expected heterozygosity H_E of populations of *Arabidopsis lyrata*

Source of variation	Dependent variable							
	H_E , all populations				H_E , outcrossing populations			
	Estimate	SE	t value	P	Estimate	SE	t value	P
Intercept	0.273	0.019	14.54	< 0.000	0.329	0.019	16.93	< 0.000
t_m	0.328	0.046	7.15	< 0.000				
Coeff. of spatial struct., m	–1.409	0.514	–2.74	0.017	–2.956	0.705	–4.19	0.002
Substrate: rock rel. to sand	0.144	0.050	2.88	0.013	0.198	0.052	3.78	0.004
$\log_{10}(\text{census size})$	0.037	0.016	2.33	0.037	0.050	0.014	3.45	0.007

All populations: $n = 18$, $R^2 = 0.85$; outcrossing populations only: $n = 13$, $R^2 = 0.76$. P values < 0.05 are indicated in bold.

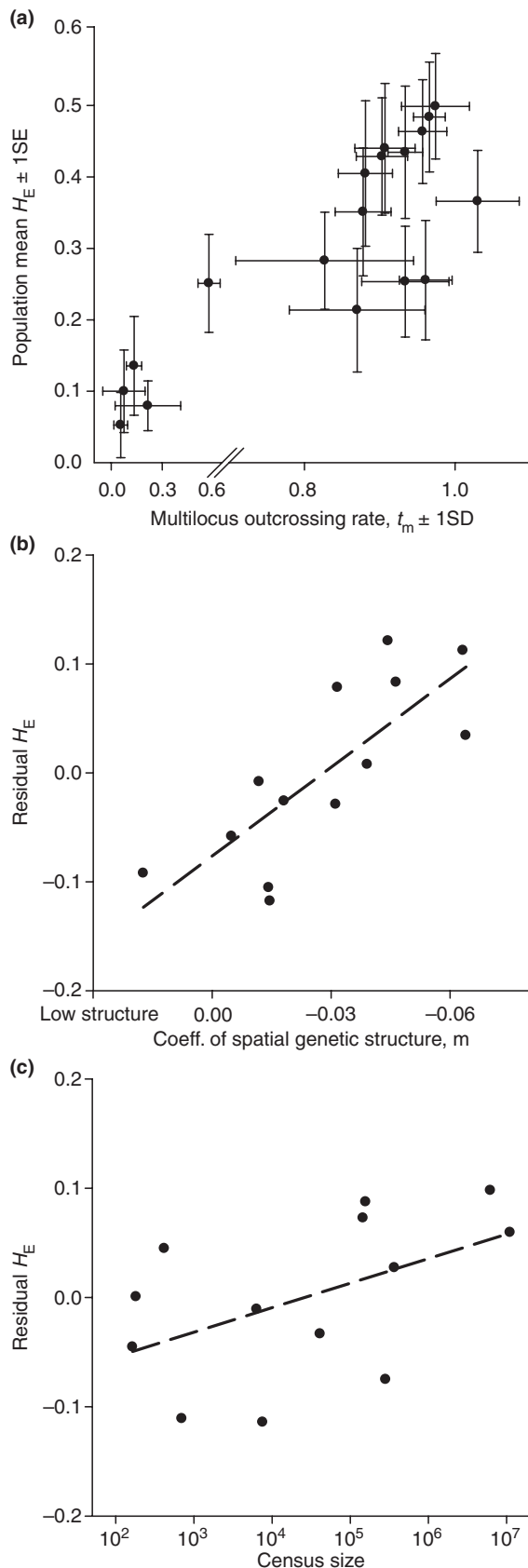


Fig. 4 Relationships between mean heterozygosity (H_E) and the population-level multilocus outcrossing rate t_m (a), the coefficient of spatial genetic structure m (b) and census size (c) across replicate populations of *Arabidopsis lyrata*. Spatial genetic structure was the slope of the regression of Moran's I among pairs of plants against spatial distance. High levels of gene diversity were associated with high mean outcrossing rates, strong spatial genetic structure within the population (more negative values of m) and large census size. See Table 2 for test statistics. Symbols are population means ($n = 18$ /all populations in a; $n = 13$ /outcrossing populations with $t_m > 0.8$ in b and c) with $\pm 1SD$ or $\pm 1SE$ in (a). In (b) and (c), the population mean H_E are residuals from a model without spatial genetic structure and census size, respectively (including only outcrossing populations).

Discussion

Genetic variation within populations of *A. lyrata* is affected strongly by the mating system and, to an appreciable extent, by spatial genetic structure, habitat type and census size. Remarkably, the four variables together explained more than 80% of the variation in H_E . When only outcrossing populations were considered, the last three factors still explained 76% of the variation in H_E . These results are important because they illustrate the relative magnitudes of several processes affecting genetic drift.

The mating system was the most important factor in determining gene diversity. The range of mating systems exhibited by *A. lyrata* in the Great Lakes region of North America has been shown previously to vary from selfing to outcrossing (Mable *et al.*, 2005; Mable & Adam, 2007; Hoebe *et al.*, 2009). The reproductive modes of mixed mating and selfing were found here and in previous studies to result in small numbers of polymorphic loci, low levels of gene diversity and high inbreeding coefficients, as predicted by theory. However, our results deviate from the expectation that pure selfing decreases N_e by one-half compared with an outcrossing population (Pollak, 1987). The N_e of selfing populations in our study was *c.* two to three times lower than that expected if selfing reduced N_e by half. Assuming a stepwise mutation model and a mutation rate of $\mu = 10^{-4}$, our H_E values corrected for all population differences, including inbreeding caused by selfing, would occur at $N_e = 745$ for selfing populations and $N_e = 2060$ for outcrossing populations (Ohta & Kimura, 1973). Under the infinite alleles model, the corresponding values of N_e are 660 and 1570 (Crow & Kimura, 1970). Therefore, other forces enhancing genetic drift must be at play in selfing populations. On the basis of theoretical work, these could include selective sweeps within populations or selection against deleterious mutations under strong linkage, fluctuations in size or recurrent extinctions and recolonizations of local demes (Caballero, 1994; Charlesworth & Wright, 2001; Ingvarsson, 2002). Alternatively, in the occasional event of pollen movement over large distances, selfing populations may be less likely to accommodate gene flow.

A recent study of *A. lyrata* did not find greater than the expected decline in H_E in selfing populations (Foxy *et al.*, 2010), but this may be explained by confounding effects of habitat, within-population structure and census size.

We found that spatial genetic structure caused by limited gene flow was positively associated with population genetic variation (Fig. 4b). Theory shows that subdivision can have nearly any effect on among-deme H_E . For example, in a linear stepping-stone model, if the migration rate among local demes exceeds a threshold value, the whole population behaves as if it were panmictic, and the rate of loss of heterozygosity is $1/[2N_T]$, where N_T is the effective size of the whole population (Maruyama, 1970; reviewed in Felsenstein, 1976). If the migration rate is below this threshold, the rate of loss of heterozygosity depends neither on the local size nor on the size of the whole population, but on the number and spatial layout of demes, together with the migration rate (Felsenstein, 1976). Under such conditions, subdivision can enhance the total effective size. By contrast, Whitlock & Barton's (1997) model suggested that subdivision increases N_e only under restricted conditions of low variance in size and reproductive output of local demes. Our discovery that genetic diversity and spatial genetic structure are connected in *A. lyrata* is important, even without a well-understood mechanism. Spatial genetic structure within populations is common, particularly in plants (Heywood, 1991), and may be an overlooked but quantitatively important driver of the maintenance of genetic variation.

Two alternative interpretations of the relationship between H_E and spatial structure seem less likely. One interpretation is that both spatial genetic structure and high heterozygosity are by-products of reduced disturbance. Under this scenario, disturbance destroys spatial genetic structure and simultaneously reduces N_e by creating local bottlenecks and extinctions. If this explanation were true, we would expect to find genetic evidence of recent bottlenecks in populations with low spatial genetic structure. This was not observed. We found no evidence that gene diversity in the less-structured populations was higher than expected based on allele numbers, which implies that N_e has not declined recently in these populations (Cornuet & Luikart, 1996). Another interpretation of Fig. 4(b) is that an interaction between spatial genetic structure and the sporophytic self-incompatibility system is responsible for the patterns of genetic diversity. If the *S*-locus is under frequency-dependent selection (Wright, 1939) and therefore neutral markers are under associative balancing selection, we would find increased N_e and genetic diversity under spatial structuring. However, Cartwright's (2009) simulations showed that allelic diversity at a neutral marker unlinked to the *S*-locus is hardly influenced by whether the species is self-incompatible or outcrossing with the random occurrence of biparental inbreeding.

Census size was also related to gene diversity and genetic drift. However, given that census size varied over five orders of magnitude, the predicted difference of -0.18 in H_E between the largest and smallest populations was slight. The most important reason may be that all large populations occurred on sand and the largest ones on lake shores. Such habitats are generally prone to disturbance, so that the populations may never reach equilibrium H_E . Nevertheless, disturbance and population fluctuations must be limited in occurrence over space and time, even in the largest populations, because of the positive relationship found between census size and H_E .

Unexpectedly, we found that gene diversity was considerably higher on rock substrates than on sand. Two potentially confounding factors, census size and spatial genetic structure, can be excluded on the basis of our results. Rock populations had significantly lower census sizes than sand populations, and were not more spatially structured. Another explanation for the high diversity on rock is that the three rock populations in the east are remnants of once larger and more connected populations. Perhaps high diversity has been maintained by relatively stable environmental conditions and longer generation times on rocky outcrops than on lakeshore sand dunes, slowing down the decay of gene diversity. By contrast, some of the sand populations, although not situated on dynamic outer dunes, may have experienced more disturbance from wildfire and the movement of sand. If this is true, then genetic drift would have been enhanced by fluctuations in census size caused by phases of withdrawal and population extension (Caballero, 1994). Disturbance also tends to shorten the generation time, and this could accelerate the process of genetic drift in sand populations. Although this explanation is speculative, it is interesting to note that the substrate type has a strong impact on population genetics.

Several studies have reported that the North American subspecies of *A. lyrata* (*A. lyrata lyrata*) shows less genetic variation than the Eurasian subspecies *petraea* (Wright *et al.*, 2003, 2006; Muller *et al.*, 2008). Our results suggest that the North American samples in these studies may not be typical for the subspecies *lyrata* or may come from sandy sites with less genetic diversity to begin with. One of the two sites in Wright *et al.* (2006) was a selfing population, Rondeau, and the North Carolina site in Muller *et al.* (2008) had especially low gene diversity for outcrossing *A. lyrata lyrata*. The European populations included in Muller *et al.* do not differ in H_E from the outcrossing ones studied here. Our study also shows that the small-scale spatial genetic structure within the first few meters, reported by Clauss & Mitchell-Olds (2006) and Lundemo *et al.* (2010) in European populations, may not be the general rule. One of the three populations shown in Fig. 2 had small-scale spatial structure, but the other two did not.

In conclusion, we have discovered that factors other than census size play an important role in determining the levels

of neutral genetic diversity and genetic drift. Among the factors that have been previously overlooked are the within-population spatial genetic structure and habitat. In particular, spatial genetic structure appears to reduce genetic drift on a population level, whereas limited pollen dispersal via self-fertilization affects drift in the opposite direction. Therefore, selfing may have a debilitating impact on the micro-evolutionary fate of populations by decreasing adaptive potential (Morran *et al.*, 2009), whereas factors such as spatial genetic structure, stable rocky substrate and large census size are likely to enhance the evolutionary potential of populations (Willi *et al.*, 2006; Hoffmann & Willi, 2008).

Acknowledgements

We thank Robert Duenner and Megan McDonald for technical assistance; Barbara Mable for providing seeds from three populations; Ed Masteller and Michelle Crowder for help with finding sites; and Josh Van Buskirk, Barbara Mable and Marc Stift for comments on the manuscript. Collection permits were granted by the Palisades Interstate Park Commission, the Nature Conservancy of Eastern New York, the New York State Office of Parks, the Commonwealth of Pennsylvania, the United States National Park Service, the Illinois Department of Natural Resources, the Michigan Department of Natural Resources and John Haataja. The research was supported by the Swiss National Science Foundation (31003A-116270 and PP00P3-123396/1), the Genetic Diversity Centre of ETH Zürich and the Fondation Pierre Mercier pour la Science.

References

- Baayen RH, Davidson DJ, Bates DM. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412.
- Bombliès K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *Plos Genetics* 6: e1000890.
- Bonnet E, Van de Peer Y. 2002. *zt: a software tool for simple and partial Mantel tests* [WWW document]. URL <http://ideas.repec.org/a/jss/jstsof/07i10.html> [accessed on 7 July 2010].
- Caballero A. 1994. Developments in the prediction of effective population size. *Heredity* 73: 657–679.
- Cartwright RA. 2009. Antagonism between local dispersal and self-incompatibility systems in a continuous plant population. *Molecular Ecology* 18: 2327–2336.
- Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics and Development* 11: 685–690.
- Chybicki IJ, Burczyk J. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* 100: 106–113.
- Clauss MJ, Cobban H, Mitchell-Olds T. 2002. Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Molecular Ecology* 11: 591–601.
- Clauss MJ, Mitchell-Olds T. 2006. Population genetic structure of *Arabidopsis lyrata* in Europe. *Molecular Ecology* 15: 2753–2766.
- Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001–2014.
- Crow JF, Kimura M. 1970. *An introduction to population genetics theory*. Minneapolis, MN, USA: Burgess.
- Ellstrand NC, Elam DR. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* 24: 217–242.
- Felsenstein J. 1976. The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* 10: 253–280.
- Foxe JP, Stift M, Tedder A, Haudry A, Wright SI, Mable BK. 2010. Reconstructing origins of loss of self-incompatibility and selfing in North-American *Arabidopsis lyrata*: a population genetic context. *Evolution* doi: 10.1111/j.1558-5646.2010.01094.x.
- Frankham R. 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology* 10: 1500–1508.
- Gaudeul M, Stenoien HK, Ågren J. 2007. Landscape structure, clonal propagation, and genetic diversity in Scandinavian populations of *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany* 94: 1146–1155.
- Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society B: Biological Sciences* 273: 3011–3019.
- Goodwillie C, Kalisz S, Eckert CG. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology, Evolution and Systematics* 36: 47–79.
- Hardy OJ, Vekemans X. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* 83: 145–154.
- Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.
- Heywood JS. 1991. Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics* 22: 335–355.
- Hoebé PN, Stift M, Tedder A, Mable BK. 2009. Multiple losses of self-incompatibility in North-American *Arabidopsis lyrata*: phylogeographic context and population genetic consequences. *Molecular Ecology* 18: 4924–4939.
- Hoffmann AA, Willi Y. 2008. Detecting genetic responses to environmental change. *Nature Reviews Genetics* 9: 421–432.
- Ingarvarsson P. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution* 56: 2368–2373.
- Jarne P, Auld JR. 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution* 60: 1816–1824.
- Kalinowski ST. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology Notes* 5: 187–189.
- Kimura M. 1955. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences, USA* 41: 144–150.
- Kuittinen H, Mattila A, Savolainen O. 1997. Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity* 79: 144–152.
- Lundemo S, Stenoien HK, Savolainen O. 2010. Investigating the effects of topography and clonality on genetic structuring within a large Norwegian population of *Arabidopsis lyrata*. *Annals of Botany* 106: 243–254.
- Mable BK, Adam A. 2007. Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. *Molecular Ecology* 16: 3565–3580.

- Mable BK, Robertson AV, Dart S, Di Berardo C, Witham L. 2005. Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* 59: 1437–1448.
- Maruyama T. 1970. Rate of decrease of genetic variability in a subdivided population. *Biometrika* 57: 299–311.
- Morrison LT, Parmenter MD, Phillips PC. 2009. Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* 462: 350–352.
- Muller MH, Leppälä J, Savolainen O. 2008. Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity* 100: 47–58.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA* 70: 3321–3323.
- Ohta T, Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* 22: 201–204.
- Pollak E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117: 353–360.
- Raymond M, Rousset F. 1995. GENEPOP version 1.2.: population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.
- Ritland K. 2002. Extensions of models for the estimation of mating systems using n independent loci. *Heredity* 88: 221–228.
- SAS Institute. 2002. *SAS OnlineDoc®*, version 9.1. Cary, NC, USA: SAS Institute.
- Schemske DW, Lande R. 1985. The evolution of self-fertilization and inbreeding depression in plants. 2. Empirical observations. *Evolution* 39: 41–52.
- Takayama S, Isogai A. 2005. Self-incompatibility in plants. *Annual Review of Plant Biology* 56: 467–489.
- Van Buskirk J, Willi Y. 2006. The change in quantitative genetic variation with inbreeding. *Evolution* 60: 2428–2434.
- Vekemans X, Hardy OJ. 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13: 921–935.
- Whitlock MC, Barton NH. 1997. The effective size of a subdivided population. *Genetics* 146: 427–441.
- Willi Y, Määtänen K. 2010. Evolutionary dynamics of mating system shifts in *Arabidopsis lyrata*. *Journal of Evolutionary Biology* 10: 2123–2131.
- Willi Y, Van Buskirk J, Hoffmann AA. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution and Systematics* 37: 433–458.
- Willi Y, Van Buskirk J, Schmid B, Fischer M. 2007. Genetic isolation of fragmented populations is exacerbated by drift and selection. *Journal of Evolutionary Biology* 20: 534–542.
- Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159.
- Wright S. 1939. The distribution of self-sterility alleles in populations. *Genetics* 24: 538–552.
- Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Looseley M, Gaut BS, Charlesworth D. 2006. Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* 174: 1421–1430.
- Wright SI, Lauga B, Charlesworth D. 2003. Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology* 12: 1247–1263.
- Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences* 169: 105–118.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Genetic diversity estimates for the 18 *Arabidopsis lyrata* populations and a single *A. thaliana* population at 10 microsatellite loci

Table S2 Matrix of correlation coefficients (bottom left) and significances (top right; Bonferroni-adjusted $\alpha = 0.0083$) of the population characteristics of the 18 *Arabidopsis lyrata* populations

Table S3 Population genetic parameters of *Arabidopsis lyrata* (A and B) and *A. thaliana* (C) populations, sorted according to their multilocus outcrossing rate (t_m)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.