doi:10.1111/j.1558-5646.2010.01094.x

RECONSTRUCTING ORIGINS OF LOSS OF SELF-INCOMPATIBILITY AND SELFING IN NORTH AMERICAN *ARABIDOPSIS LYRATA*: A POPULATION GENETIC CONTEXT

John Paul Foxe,[1,2] Marc Stift,[2,3,4,5] Andrew Tedder,[3] Annabelle Haudry,[3] Stephen I. Wright,[6,7] and Barbara K. Mable[3]

[1]*Department of Biology, York University, 4700 Keele St. Toronto, Ontario M3J 1P3, Canada*

[3]*Division of Ecology and Evolutionary Biology, University of Glasgow, Glasgow G12 8QQ, Scotland*

[5]*E-mail: marcstift@gmail.com*

[6]*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, Ontario M5S 3B2, Canada*

[7]*Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Canada*

Received January 28, 2010
Accepted July 12, 2010

Theoretical and empirical comparisons of molecular diversity in selfing and outcrossing plants have primarily focused on long-term consequences of differences in mating system (between species). However, improving our understanding of the causes of mating system evolution requires ecological and genetic studies of the early stages of mating system transition. Here, we examine nuclear and chloroplast DNA sequences and microsatellite variation in a large sample of populations of *Arabidopsis lyrata* from the Great Lakes region of Eastern North American that show intra- and interpopulation variation in the degree of self-incompatibility and realized outcrossing rates. Populations show strong geographic clustering irrespective of mating system, suggesting that selfing either evolved multiple times or has spread to multiple genetic backgrounds. Diversity is reduced in selfing populations, but not to the extent of the severe loss of variation expected if selfing evolved due to selection for reproductive assurance in connection with strong founder events. The spread of self-compatibility in this region may have been favored as colonization bottlenecks following glaciation or migration from Europe reduced standing levels of inbreeding depression. However, our results do not suggest a single transition to selfing in this system, as has been suggested for some other species in the *Brassicaceae*.

KEY WORDS: **Arabidopsis, bottlenecks, breakdown of self-incompatibility, demography, effective population size, inbreeding depression, mating system evolution, population genetics.**

Inbreeding has often been posited as an evolutionary dead end because of the accumulation of slightly deleterious mutations and reduced adaptability (Stebbins 1957; Takebayashi and Morrell 2001). Paradoxically, the transition to selfing is cited as one of the most common major evolutionary transitions across the plant kingdom and is well documented in a variety of species across many genera (Stebbins 1950, 1970; Grant 1981; Barrett et al. 1996; Barrett 2002; Igic et al. 2008). Selfers have two main advantages over outcrossers: an inherent transmission advantage

[2]These authors contributed equally to this work.

[4]Present address: Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, R. Padre Armando Quintas, 4485-661, Vairão, Portugal.

(Fisher 1941) and the ability to reproduce without mates (Darwin 1876; Kalisz et al. 2004; Charlesworth 2006). Although outcrossers only transmit 50% of their genome to their offspring, strict selfers transmit their whole genome and can at the same time act as pollen donors for seed produced by other individuals (Fisher 1941). Moreover, unlike outcrossers, selfers can reproduce when pollinators or potential mates are limited (reproductive assurance, first proposed by Darwin (1876)). Thus, a selfing lifestyle can result in increased colonization ability, as a new population may be founded from a single plant (Baker 1955; Stebbins 1956, 1957; Pannell and Barrett 1998). Particularly in the initial stages after a transition to selfing, increased homozygosity may lead to inbreeding depression (the reduction in fitness of selfed versus outcrossed individuals) due to expression of recessive deleterious load. Hence, theory predicts that selfing is only likely to evolve when the advantages of selfing outweigh the costs associated with inbreeding depression (Charlesworth and Charlesworth 1987).

A selfing strategy is also expected to come at the cost of reduced genetic variation (Charlesworth et al. 1993; Nordborg 2000; Charlesworth and Wright 2001; Glémin et al. 2006; Wright et al. 2008). First, selfing increases levels of homozygosity, thereby reducing the effective population size (N_e) and levels of diversity up to twofold under complete selfing (Pollak 1987). This increased homozygosity also leads to a reduction in effective recombination rate, resulting in increased linkage disequilibrium (LD) across loci (Nordborg 2000). This facilitates genetic hitchhiking through both positive (selective sweeps (Maynard Smith and Haigh 1974)) and negative selection (background selection (Charlesworth et al. 1993)). Genetic hitchhiking exacerbates the reduction in N_e and diversity. Finally, increased population turnover and colonization bottlenecks in selfing plants may contribute to further reductions in diversity (Ingvarsson 2002).

Empirically, the population genetic consequences associated with a transition to selfing have been well documented at the species level in both plant and animal systems (Charlesworth and Yang 1998; Baudry et al. 2001; Chiang et al. 2003; Cutter and Payseur 2003; Glémin et al. 2006). Selfing species are typically characterized by greater than twofold reductions in diversity, consistent with roles for genetic hitchhiking and/or increased colonization bottlenecks (Wright et al. 2008). These patterns of reduced genetic diversity in selfers have been found across a number of plant genera that include both outcrossing and selfing species, including *Leavenworthia* (Charlesworth and Yang 1998; Liu et al. 1998; Filatov and Charlesworth 1999; Liu et al. 1999), *Arabidopsis* (Savolainen et al. 2000; Wright et al. 2003; Ross-Ibarra et al. 2008), *Lycopersicon* (Baudry et al. 2001), and *Miscanthus* (Chiang et al. 2003).

The model plant system *Arabidopsis thaliana* is thought to have evolved self-fertilization approximately 1 million years ago through inactivation of the self-incompatibility locus, referred to as the *S*-locus (Tang et al. 2007). Evidence for the role of the *S*-locus stems from transformation studies, which identified five accessions in which full self-incompatibility could be restored by transformation with a functional *S*-locus, and implies that all other genes required for SI are still intact in these accessions (Tang et al. 2007; Boggs et al. 2009). Recent results suggest that a mutation in the male component of self-incompatibility (SCR) has resulted in loss of SI, apparently across a wide range of accessions (Tsuchimatsu et al. 2010). In addition, a modifier locus has been identified, unlinked to the *S*-locus (Liu et al. 2007), which suggests that *S*-locus inactivation may not be the sole mechanism by which SI broke down in *A. thaliana* and different mechanisms of loss could have operated in different accessions (Boggs et al. 2009).

Systems with more recent transitions from outcrossing to selfing may provide a more direct picture of the causes and short-term consequences of mating system evolution (Foxe et al. 2009; Guo et al. 2009; Ness et al. 2010). For example, if the evolution of selfing involves the long-term spread of modifiers through previously outcrossing populations, recently derived selfing populations are expected to retain reasonably high levels of ancestral polymorphism, as recently observed in *Eichhornia paniculata* (Ness et al. 2010). In contrast, if a highly selfing lineage evolves rapidly from a small number of founders, we would expect a severe loss of genetic variation, as seen in *Capsella rubella* (Foxe et al. 2009). Here, we have set out to investigate the loss of self-incompatibility in North American populations of the normally outcrossing species *Arabidopsis lyrata* (*Brassicaceae*).

It has been suggested that *A. lyrata* colonized North America from ancestral European populations (Clauss and Mitchell-Olds 2006; Ross-Ibarra et al. 2008), which are highly self-incompatible and exclusively outcrossing. The North American populations are unique because some are still predominantly outcrossing, despite the occurrence of self-compatible individuals at low frequency, whereas others are almost entirely self-compatible and have undergone a transition to high rates of selfing (Mable et al. 2005; Mable and Adam 2007). This transition to selfing in *A. lyrata* appears to be very recent, as selfing populations belong to a chloroplast lineage that also contains outcrossing populations (Hoebe et al. 2009). Moreover, selfing populations are not characterized by smaller flowers (Hoebe 2009), which contrasts with other systems in which the transition to selfing has led to notable floral evolution toward smaller flowers (Hurka and Neuffer 1997; Charlesworth and Vekemans 2005; Tang et al. 2007; Foxe et al. 2009; Guo et al. 2009).

Previous work on North American populations of *A. lyrata* in the Great Lakes region (where loss of self-incompatibility has occurred) has been based on chloroplast sequences and microsatellite genotypes. The former, as a marker with a single coalescent history and limited variation across populations, allowed

(limited) phylogeographic inferences (Hoebe et al. 2009, Tedder et al. 2010). The latter allowed basic inferences about population structure and diversity (Mable et al. 2005; Mable and Adam 2007; Hoebe et al. 2009), but suffers from potential limitations due to homoplasy and uncertainties in the mutation model, particularly for the individual markers that we had been using (Muller et al. 2008). However, nuclear gene sequences are more powerful to explicitly test population genetic predictions about the reductions in diversity in selfing populations. They also allow for the detection of recombination, and for testing whether selfing populations show more evidence of departures from demographic equilibrium than outcrossing populations. We have extended the population sampling presented in previous work (Mable et al. 2005; Mable and Adam 2007; Hoebe et al. 2009), primarily to establish if there are more populations that have undergone a transition to selfing.

In this study, we integrate polymorphism information from nuclear genes, chloroplast markers, and nuclear microsatellites, to obtain a detailed picture of the demographic history and population structure of *A. lyrata* in the Great Lakes region of North America. Our ultimate goal is to use this framework to elucidate the origins of the selfing populations. Specifically, we aim to: (1) investigate the demographic and population genetic consequences of losses of SI and transition to selfing, by describing population structure and testing the effects of individual selfing phenotype and selfing rate on genetic diversity; and (2) elucidate the extent to which severe population bottlenecks may have played a role in the transition to inbreeding.

Methods

SAMPLING

The samples for this study were collected from 22 locations throughout the Great Lakes region of eastern North America (Fig. 1, Table S1). From each location, we collected batches of seeds from 25–30 independent plants (growing at least 5 m apart). More detailed sampling description is available as a supplement (Supporting Information). In one location (Tobermory Cliffs, on the Bruce Peninsula that extends into Georgian Bay), we collected seeds from two spatially separated areas: in the first, plants were previously demonstrated to be highly selfing (dubbed TC: Mable et al. 2005); in the second, plants were observed to have lower and more variable seed set in the field and were thus suspected of being more highly outcrossing (TCA). In another location (Tobermory Singing Sands, also on the Bruce Peninsula, but on the Lake Huron side), we collected seeds from two spatially separated areas: in the first, plants grew on the characteristic sand dune habitat and were previously demonstrated to be highly outcrossing (dubbed TSS: Mable et al. 2005); in the second, plants grew on alvar (limestone pavement) and were observed to have high seed set in the field, suggestive of selfing (TSSA). We grew up eight plants from each of 10 seed batches from each of the 24 population samples (22 locations, cf. Table S1), forming a collection of 192 plants, grown in a common greenhouse environment at the Scottish Crop Research Institute (Invergowrie, UK) under a constant regime of 16 h light: 8 h dark, 22°C days and 18°C night.

SELFING PHENOTYPE DETERMINATION

For all individual plants that flowered, we manually self-pollinated six flowers. The resulting siliques were scored as either negative (no seeds), small (a short silique smaller than 9 mm with no more than three seeds), or positive (a silique of 9 mm or longer with more than three seeds). Small siliques were considered as negative for the purposes of classifying selfing phenotypes (as in Mable et al. 2005) but were recorded to enable assessment of whether the degree of leakiness in the SI system varied by population or geographic region. Based on this, we defined the selfing phenotype of each plant based on the siliques produced after manual self-pollination as: (1) self-incompatible (SI): zero or one (out of six) positive siliques; (2) self-compatible (SC): five or six (out of six) positive siliques; and (3) partially self-compatible (PC): two, three, or four (out of six) positive siliques. To exclude pollen sterility or total sterility causing a false SI phenotype, plants classified as self-incompatible were crossed with plants from the same population to test for fertility. All self-incompatible plants were cross-fertile with at least one other plant tested.

MATING SYSTEM DETERMINATION (ESTABLISHING POPULATION LEVEL OUTCROSSING RATES)

For 12 populations (IND, LPT, LSP, MAN, PIC, PIN, PTP, PUK, RON, TC, TSS, and WAS), multilocus outcrossing rates had been determined in previous studies (Mable and Adam 2007). For nine of the newly sampled populations (BEI, HDC, KTT, OWB, PCR, PIR, PRIA, SBD, and TSSA) we used the same microsatellite markers and procedures to calculate outcrossing rates as outlined in Mable and Adam (2007) and Hoebe et al. (2009). In brief, we genotyped progeny arrays (6–10 offspring per mother) from 17 to 27 mothers per population, and estimated multilocus outcrossing rates using MLTR version 2.3 (Ritland 2002), which implements the mixed-mating model described by Ritland and Jain (1981). Too few seed families from IOM and TCA germinated to allow reliable evaluation of outcrossing rates but a rough estimate from TCA was obtained from five maternal families and those for IOM were obtained from Yvonne Willi (pers. comm.), who used a similar set of microsatellite markers on the same seed samples. Similarly, too few seed families from NCM germinated for reliable outcrossing estimates. Therefore, mating system for NCM was assumed based upon selfing phenotypes, which showed a predominance of SI individuals.

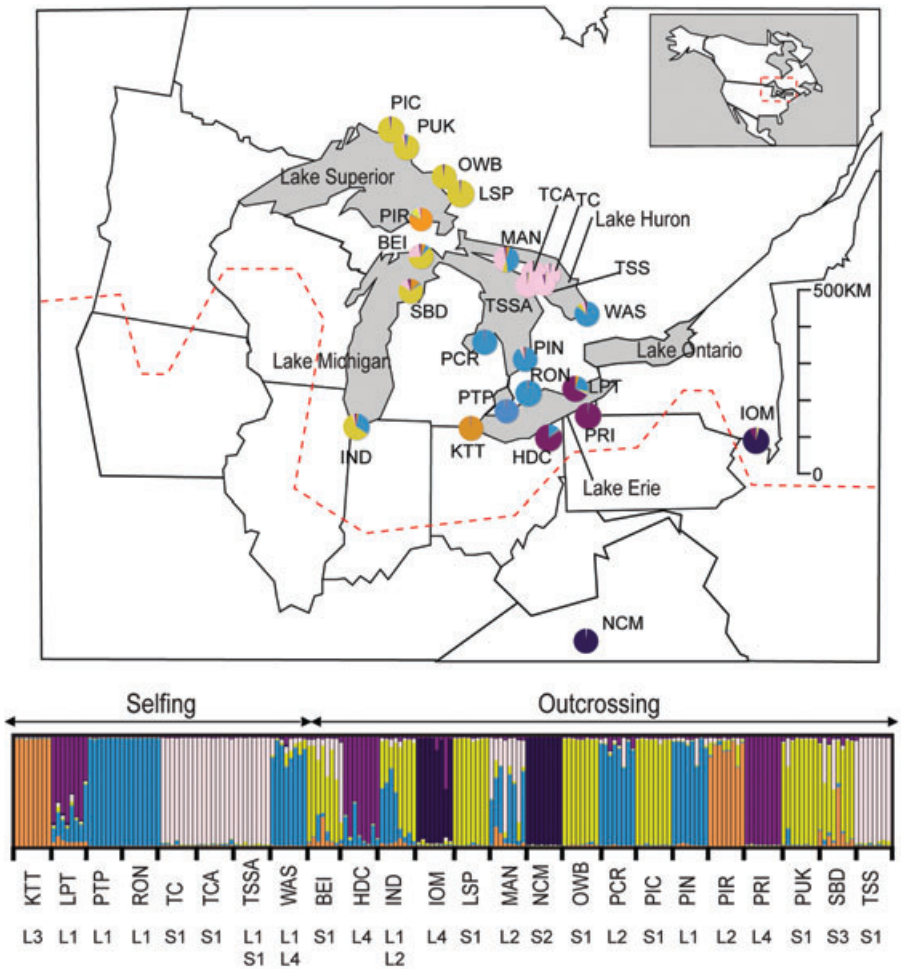

Figure 1. **Posterior probabilities of Bayesian clustering analysis (InStruct) using the combined nuclear gene sequence and microsatellite datasets, based on a prior of six clusters ($k = 6$). Bar plots show individual posterior probabilities, pie charts on the map show mean posterior probabilities for each population (averaged over eight individuals). Chloroplast *trn*F(GAA) region haplotypes found in each population are listed beneath the population labels below the bar plots. The dotted line on the map indicates the approximate southern limit of the ice sheet during the last glacial maximum.**

MICROSATELLITE GENOTYPING

Nine microsatellite loci previously used by Mable and Adam (2007) were screened for variation across all 192 individuals: ADH-1, AthZFPG, ATTS0392, F20D22, ICE12, ICE9 (Clauss et al. 2002), LYR104, LYR133, and LYR417 (obtained from V. Castric and X. Vekemans, pers. comm.). Products were amplified by multiplex PCR, using the default reagent concentrations recommended by the kit instruction manual (QIAGEN Multiplex PCR Kit, QIAGEN Ltd, Crawley-West Sussex, UK, exact primer concentrations can be requested from the authors). Thermocycling was performed on PTC-200 (MJ research) machines using the following programme: initial denaturation at 95°C for 15 min followed by 34 cycles of 94° for 30 s, 55°C for 90 s, 72°C for 90 s, (ramp to 72°C at 0.7°C/s), and a final 72°C extension for 10 min. Multiplex products (1:160 dilutions) were genotyped using an ABI 3730 sequencer (by The Sequencing Service, University of Dundee, Dundee, UK). Genotypes were analyzed using GENEMAPPER 4.0 (Applied Biosystems, Warrington, UK) and corrected manually.

PCR AND SEQUENCING OF NUCLEAR GENES

For each of the 192 individuals across 24 populations, products were produced from PCR primer pairs that were previously designed and confirmed to amplify large exons from 18 nuclear genes (putative functions listed in Table S1 in Ross-Ibarra et al. 2008) following methods described by Wright et al. (2006) and Ross-Ibarra et al. (2008). PCR reactions were performed in 25 μL reaction volumes (15 mM PCR (10X) buffer, 2 mM $MgSO_4$, 10 mM dNTPs, 10 μM forward primer, 10 μM reverse primer, 1U Tsg polymerase and 50–100 ng DNA) on an Eppendorf Mastercycler with the following program: 2 min at 94°C, 20 sec at 94°C, 20 sec at 55°C, 40 sec at 72°C, for 35 cycles, with a final extension time of 4 min at 72°C.

Sequencing reactions were carried out by Lark Technologies, Texas. Chromatograms were analyzed using Sequencher 4.6 (Gene Codes Corporation, Ann Arbor, MI), using the "call secondary peaks" option to aid in the identification of heterozygous sites. All chromatograms were checked manually for heterozygous nucleotide positions, using the sequence from both strands to confirm putative heterozygous sites. Due to a significant amount of sequencing failure across all 18 loci for individuals from LPT, this population was removed from subsequent nuclear data analyses. All sequences have been submitted to Genbank, with accession numbers HM168020-HM171110.

DNA SEQUENCING OF CHLOROPLAST DNA

The noncoding cpDNA region *trn*L(UAA)3′exon-TrnF(GAA) was amplified with the primers E 5′-GGTTCAAGTCC CTCTATCCC-3′ and *F* 5′-ATTTGAACTGGTGACACGAG-3′ (Taberlet et al. 1991) and sequenced. This includes a region with pseudogene copies of the *trn* gene (Koch and Kiefer 2005; Ansell et al. 2007; Tedder et al. 2010). Using the same primers in a smaller population sample, Hoebe et al. (2009) identified a short haplotype (515bp, dubbed S1) and two long haplotypes (741bp, dubbed L1 and L2). After purification with QiaQuick gel extraction kits (Qiagen Ltd, Crawley-West Sussex, UK), all PCR products were sequenced directly on an ABI 3730 sequencer by The Sequencing Service, University of Dundee. Sequences were visually checked using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, MI) and aligned to the previously identified L1, L2, and S1 haplotypes (Hoebe et al. 2009).

NUCLEAR GENE SEQUENCE ANALYSIS

We reconstructed individual haplotypes of unphased diploid sequences using the software PHASE (Stephens et al. 2001), as implemented in DnaSP Version 5.0 (Librado and Rozas 2009). For the sequence data, synonymous and nonsynonymous sites were identified by aligning each fragment to the corresponding fragment in the *A. thaliana* genome sequence, identified using BLAST (Altschul et al. 1990), and using the protein annotation from *A. thaliana*. Standard population genetic descriptives, including numbers of synonymous and nonsynonymous sites, estimates of synonymous (π_{syn}) and nonsynymous (π_{rep}) diversity and Tajima's D, were calculated using a modified version of Polymorphurama, a Perl script written by D. Bachtrog and P. Andolfatto (available from http://ib.berkeley.edu/labs/bachtrog/data/polyMORPHOrama/polyMORPHOrama.html). Significance of within-population mean Tajima's D was determined by conducting 10,000 coalescent simulations as implemented in the HKA software (available from http://genfaculty.rutgers.edu/hey/software#HKA, Kliman et al. 2000). The within population recombination parameter ρ (where $\rho = 4N_e r$, N_e being the effective population size and r recombination rate) was calculated for each locus with more than three segregating sites and for each population by using the maxdip program (available from http://genapps.uchicago.edu/maxdip/index.html) for diploid unphased data. Maxdip applies a composite likelihood approach fit to the observed pairwise SNP frequencies (Hudson 2001) and assumes an infinite-sites constant-population-size neutral model.

MICROSATELLITE ANALYSIS

For the microsatellite data, we used MSA (Dieringer and Schlotterer 2003) to calculate observed and expected heterozygosity (H_o and H_e) for each locus.

THE EFFECTS OF SELFING PHENOTYPE AND MATING SYSTEM ON GENETIC DIVERSITY AND HETEROZYGOSITY

Observed heterozygosity at nuclear loci was calculated using Perl scripts written by A.H. For each individual and each locus, the heterozygosity status was determined by comparing the two gene copies carried by the individual. If the two haplotypes were different, the status was described as heterozygous; if they were identical, the status was described as homozygous. Individual H_o was estimated using the average H_o over all the loci for each individual.

Linear regressions as implemented in JMP 8.0.2 (SAS Institute Inc., Cary, NC) were used to test whether summary statistics describing genetic diversity and heterozygosity at the population level (π_{syn}, ρ, and H_o for nuclear gene data), H_o and H_e (for microsatellites) varied in relation to outcrossing rate (T_m) and/or the proportion of SC individuals in each population. One-way ANOVA, also implemented in JMP 8.0.2 was used to test the effect of selfing phenotype on individual heterozygosity.

EXPLAINING THE REDUCED GENETIC DIVERSITY OF SELFING POPULATIONS: TESTING FOR BOTTLENECK EFFECTS BEYOND SELFING ALONE

Selfing reduces effective population size (Nordborg 2000). Therefore, genetic diversity is expected to decrease with selfing rate, up to a twofold reduction for completely selfing populations. Other demographic effects and genetic hitchhiking could further decrease genetic diversity. We tested if genetic diversity (π_{syn} and π_{rep} for gene sequences) in populations classified as selfing (based on multilocus outcrossing rates) was lower than expected based on the selfing rate ($S = 1 - T_m$) alone. To do this, we "corrected" the estimated genetic diversity using the formula $\theta_{corrected} = \theta_{obs}(1 + F)$, where θ is the genetic diversity measure being corrected, and $F = S/2 - S$, where F is the inbreeding coefficient and S is the selfing rate for the population (Nordborg 2000). Similarly, we corrected ρ using the formula $R_{corrected} = R_{obs}/(1 - S)$ where R is the recombination rate and S is the selfing rate for the population

(Nordborg 2000). Then, we used linear regressions to test if population outcrossing rates (T_m) predicted corrected π_{syn}, π_{rep}, microsatellite H_o, H_e and ρ. If so, and corrected genetic diversity decreased with selfing rate, this would provide evidence that genetic diversity in selfing populations is reduced due to additional factors beyond selfing alone (e.g., a bottleneck, selection).

BAYESIAN INFERENCE OF POPULATION STRUCTURE

We used InStruct (version 1.0, Gao et al. 2007) to infer population structure using a combination of phased nuclear gene sequence and microsatellite data. For comparative purposes, we also used STRUCTURE (version 2.3.2, Pritchard et al. 2000) to infer population structure. Both programs perform Bayesian clustering and work by assigning individuals to a given number of clusters in such a way that deviations from Hardy–Weinberg equilibrium are minimized. Unlike STRUCTURE, InStruct can accommodate nonrandom mating due to selfing. Based on exploratory runs, we restricted the number of clusters (k) to range from $k = 1$ to $k = 12$, and ran both programs for 2,000,000 generations with a burnin of 200,000 generations, with five independent chains (runs) for each k. We ran InStruct in the mode that allows for admixture and individual selfing rates. We used Perl scripts written by Joseph Hughes (available from http://linnaeus.zoology.gla.ac.uk/~jhughes/Bioinformatics.html) on InStruct output files, and Structure Harvester (version 0.3 by D.A. Earl: http://taylor0.biology.ucla.edu/struct_harvest/) on STRUCTURE output files to excise the probability matrices for each level of k and perform matrix alignment using the software CLUMPP version 1.1.1 (Jakobsson and Rosenberg 2007). DISTRUCT version 1.1 (Rosenberg 2004) was used to create bar plots of the aligned matrices.

INFERRING THE ORIGIN OF SELFING POPULATIONS

For each locus, we counted the total number of microsatellite alleles and gene sequence haplotypes over all populations by using Perl scripts written by A.H. Then, we grouped populations according to their classification as selfing ($T_m < 0.5$) and outcrossing ($T_m > 0.5$). For both groups, we then counted the total number of microsatellite alleles and nuclear gene sequence haplotypes, and those that were unique to either group. Finally, we counted those that were shared between the groups. Then, we tested if the selfing and outcrossing populations had a significantly different number of unique variants relative to the total number of variants observed in each of these groups. We did this separately for microsatellite alleles (across all loci) and unique nuclear gene sequence haplotypes (across all genes) using G-test for goodness of fit to assess significance.

We then looked at the haplotypes that were unique to the selfing population group (i.e., the haplotypes not shared between the selfing and outcrossing group of populations) in a separate analysis. Specifically, we explored whether haplotypes unique to selfing populations were derived from haplotypes occurring in particular outcrossing populations, or unrelated to any haplotypes in outcrossing populations. We considered haplotypes to be derived from another haplotype if they had up to two base pair differences, and unrelated if they differed by more than two base pairs from any other haplotype in our sampling. The presence of such "unrelated" haplotypes within selfing populations would indicate an origin not included in our sampling, or mutations that have accumulated subsequent to divergence from a shared ancestor. Note that, if a selfing population originated from another selfing population, this would be reflected by two or more selfing populations sharing haplotypes that are derived from the same ancestral population. As in the previous paragraph, this would be indistinguishable from a scenario of independent colonizations from the same outcrossing population.

Results

SELFING PHENOTYPES AND OUTCROSSING RATES

Population-level multilocus outcrossing rate estimates (T_m) ranged from 0.09 to 0.99 (Table 1; Table S2). We categorized populations as selfing if T_m was < 0.5, and outcrossing if $T_m > 0.5$. Eight populations were classified as selfing ($T_m < 0.5$: KTT, LPT, PTP, RON, TC, TCA, TSSA, and WAS). All of the remaining populations were classified as outcrossing ($T_m > 0.5$: BEI, HDC, IND, IOM, LSP, MAN, OWB, PCR, PIC, PIN, PIR, PRI, PUK, SBD, TSS). NCM was classified as outcrossing based on a preponderance of SI individuals, but outcrossing rates were not estimated due to insufficient seeds per mother (Table 1). Small siliques were found in all predominantly outcrossing populations but not in the populations that were predominantly self-compatible, emphasizing that their occurrence represents leakiness of the self-incompatibility system rather than a complete loss of it, as suggested previously (Mable et al. 2005). Overall, the proportion of self-compatible individuals in each population was a strong predictor of outcrossing rates (linear regression, beta $= -0.68$, $R^2 = 0.887$, $F_{[1,21]} = 165.7$, $P < 0.0005$). Nevertheless, it is worth noting that some self-compatible (SC) or partially self-compatible (PC) plants were found in the majority of the outcrossing populations (Table S2). SC individuals were found in five of the 15 outcrossing populations (HDC, OWB, IND, MAN and PRI) and PC individuals were found in nine (BEI, HDC, IOM, NCM, OWB, PCR, PIR, PUK, and SBD). Likewise, some of the populations classified as selfing included some SI and PC individuals, with fully SI individuals found in two of the eight selfing populations (TSSA and TC).

Table 1. Population-level outcrossing rates (T_m), proportion of self-compatible (SC) individuals (as an indication of selfing phenotype), summary statistics and observed heterozygosity for 18 nuclear gene sequences, and observed and expected heterozygosities (H_o and H_e) across nine microsatellite loci, with populations ordered by increasing outcrossing rates.

			Based on nuclear gene sequences of 18 loci										Based on nine microsatellite loci		
Population	T_m[1]	Prop. SC plants[2]	Syn. sites[3]	Repl. sites[4]	π_{syn}[5]	Corr. π_{syn}[6]	π_{rep}[7]	Corr. π_{rep}[6]	Taj D[8]	ρ[9]	Corr. ρ[10]	H_o	H_o	H_e	Corr. H_e[6]
PTP	0.09	1	1880.8	6225.2	0.005	0.0087	0.001	0.0018	0.421	0	0	0.022	0.069	0.077	0.14
LPT[11]	0.13	1	–	–	–	–	–	–	–	–	–	–	0.069	0.13	0.23
TC	0.18	0.88	1868.03	6174.97	0.011	0.0193	0.002	0.0034	0.505*	0.005	0.025	0.117	0.181	0.281	0.48
WAS	0.25	1	1862.12	6174.88	0.005	0.0073	0.001	0.0016	0.774**	0	0	0.104	0.083	0.136	0.22
RON	0.28	1	1870.3	6190.8	0.004	0.0056	0.001	0.0016	−0.831***	0	0	0.045	0.028	0.028	0.044
KTT	0.31	1	1876.4	6211.6	0.008	0.012	0.001	0.0015	0.631**	0.008	0.024	0.057	0	0.044	0.067
TSSA	0.41	0.5	1880.52	6225.48	0.006	0.0082	0.002	0.0028	−0.278	0	0	0.084	0.097	0.187	0.27
TCA	0.48	1	1880.66	6225.34	0.013	0.0182	0.002	0.0027	0.509**	0	0	0.063	0.042	0.121	0.16
OWB	0.64	0.29	1835.9	6090	0.008	0.0094	0.001	0.0012	0.754**	0.004	0.006	0.27	0.35	0.3	0.36
HDC	0.65	0.14	1879.8	6226.3	0.003	0.0037	0.001	0.0012	0.480	0.009	0.014	0.17	0.069	0.14	0.17
PIC	0.77	0	1823.2	6015.8	0.013	0.0142	0.002	0.0023	0.513*	0	0	0.21	0.36	0.32	0.36
MAN	0.83	0.13	1790.4	5850.6	0.016	0.0177	0.002	0.0022	0.700**	0.005	0.005	0.23	0.22	0.27	0.29
PIN	0.84	0	1878.4	6218.7	0.014	0.0149	0.002	0.0022	0.191	0.002	0.002	0.22	0.18	0.26	0.28
PIR	0.88	0	1780.5	5929.5	0.012	0.013	0.002	0.0021	0.170	0.001	0.001	0.21	0.26	0.26	0.28
PRI	0.89	0.13	1854.2	6143.8	0.008	0.0086	0.001	0.0011	0.527*	0	0	0.076	0.14	0.15	0.16
TSS	0.91	0	1880.1	6225.9	0.01	0.0101	0.002	0.0021	0.578**	0.005	0.005	0.301	0.361	0.456	0.48
IOM	0.94	0	1873.9	6199	0.01	0.0106	0.002	0.0021	0.419*	0.003	0.003	0.23	0.28	0.4	0.41
LSP	0.94	0	1781.3	5829.7	0.005	0.0057	0.001	0.001	−0.335	0.01	0.011	0.16	0.14	0.11	0.11

Continued.

Table 1. Continued.

Population	T_m[1]	Prop. SC plants[2]	Based on nuclear gene sequences of 18 loci										Based on nine microsatellite loci		
			Syn. sites[3]	Repl. sites[4]	π_{syn}[5]	Corr. π_{syn}[6]	π_{rep}[7]	Corr. π_{rep}[6]	Taj D[8]	ρ[9]	Corr. ρ[10]	H_o	H_o	H_e	Corr. H_e[6]
SBD	0.94	0	1813.3	6010.7	0.015	0.0157	0.003	0.0031	0.757***	0.023	0.024	0.27	0.42	0.54	0.56
PUK	0.96	0	1856.8	6150.2	0.016	0.0161	0.002	0.002	0.822***	0.009	0.009	0.26	0.34	0.34	0.34
BEI	0.98	0.14	1853.8	6135.2	0.02	0.0204	0.003	0.003	0.683**	0.016	0.016	0.33	0.32	0.39	0.39
PCR	0.98	0	1871.6	6195.4	0.012	0.0119	0.001	0.001	−0.009	0.002	0.002	0.23	0.21	0.3	0.3
IND	0.99	0.14	1834.6	6067.4	0.011	0.0111	0.002	0.002	0.421*	0.005	0.005	0.3	0.43	0.47	0.47
NCM[12]	–	0	1741.9	5668.1	0.003	–	0.001	–	0.875**	0.003	–	0.12	0.14	0.17	–

*$P<0.05$, **$P<0.01$, ***$P<0.001$.

[1]Based on multilocus estimates obtained from microsatellite variation in progeny arrays, using MLTR version 2.3 (Ritland 2002).

[2]Based on manual self-pollinations (see Table S2 for more details).

[3]Total number of synonymous sites across loci.

[4]Total number of replacement (nonsynonymous) sites across loci.

[5]π_{syn}: synonymous nucleotide diversity: the average number of pairwise differences between two sequences across loci.

[6]For each of the diversity measures corrections for the reduction in effective population size due to selfing rate ($S=1-T_m$) were performed using the formula: $\theta_{corrected}=\theta_{obs}(1+F)$, where θ represents the genetic diversity measure, and $F=S/2-S$, where F is the inbreeding coefficient and S is the selfing rate for the population.

[7]π_{rep}: replacement (nonsynonymous) nucleotide diversity: the average number of pairwise differences between two sequences across loci.

[8]Taj D: average Tajima's D across loci.

[9]ρ: population recombination parameter (see text for details), median across loci.

[10]The recombination parameter ρ was corrected for the reduction in effective population size due to selfing rate ($S=1-T_m$) using the formula $R_{corrected}=R_{obs}/(1-S)$ where R is the recombination parameter and S is the selfing rate for the population.

[11]Nuclear gene sequencing failed for this population due to PCR problems.

[12]Outcrossing rates could not be obtained for this population due to insufficient seeds in the batches for progeny arrays.

Table 2. Linear regressions of outcrossing rate (T_m) and the proportion of self-compatible (SC) individuals per population on synonymous diversity (π_{syn}), corrected synonymous diversity, the recombination parameter (ρ), the corrected recombination parameter and observed heterozygosity (H_o) across 18 nuclear gene sequences; and observed and expected microsatellite heterozygosity (H_o and H_e) across nine microsatellite loci.

	T_m[1]					Proportion of SC individuals[2]				
	r^2	F ratio	Beta	Degrees freedom	P value	r^2	F ratio	Beta	Degrees of freedom	P value
π_{syn}[3]	0.3	8.65	0.008	21	0.008	0.21	5.4	−0.004	22	0.03
Corrected π_{syn}[4]	0.039	0.81	0.003	21	0.38	0.14	0.28	−0.001	21	0.61
π_{rep}[5]	0.171	4.12	0.008	21	0.0558	0.112	2.77	−0.0005	22	0.11
Corrected π_{rep}[4]	0.0069	0.139	−0.00019	21	0.71	0.0004	0.08	0.0001	21	0.78
ρ[6]	0.149	3.49	−0.005	21	0.0761	0.129	2.96	−0.005	22	0.1006
Corrected ρ[7]	0.0008	0.0156	0.0003	21	0.90	0.0003	0.0053	0.0003	21	0.94
Nuclear H_o	0.64	35.23	0.250	21	<0.0001	0.58	1.13	−0.169	22	<0.0001
Microsatellite H_o	0.49	19.1	0.306	22	0.0003	0.51	20.5	−0.223	23	0.0002
Microsatellite H_e	0.44	15.6	0.313	22	0.0008	0.44	15.5	−0.223	23	0.0008

[1]Estimated based on multilocus microsatellite variation in progeny arrays, using MLTR version 2.3 (Ritland 2002).

[2]Proportion of self-compatible (SC) individuals within the population (cf. Table S2).

[3]Synonymous diversity across all 18 nuclear loci in each population as measured by π_{syn}, where π is the average number of pairwise differences between two sequences.

[4]For each of π_{syn} and π_{rep} corrections for the reduction in effective population size due to selfing rate ($S=1-T_m$) were performed using the formula: $\theta_{corrected}=\theta_{obs}(1+F)$, where θ represents the genetic diversity measure, and $F=S/2-S$, where F is the inbreeding coefficient and S is the selfing rate for the population.

[5]Replacement diversity across all 18 nuclear loci in each population as measured by π_{rep}, where π is the average number of pairwise differences between two sequences.

[6]The population recombination parameter ρ; here we use the median ρ across all 18 nuclear loci in each population on sequences with more than three segregating sites.

[7]The recombination parameter ρ was corrected for the reduction in effective population size due to selfing rate ($S=1-T_m$) using the formula $R_{corrected}=R_{obs}/(1-S)$ where R is the recombination parameter and S is the selfing rate for the population.

THE EFFECT OF SELFING PHENOTYPE AND MATING SYSTEM (OUTCROSSING RATES) ON GENETIC DIVERSITY AND HETEROZYGOSITY

Significant linear regressions indicated that both outcrossing rate and the proportion of SC individuals in a population were good predictors of average multilocus synonymous nucleotide diversity (π_{syn}) and expected microsatellite heterozygosity (H_e) (Table 2): both π_{syn} and H_e diversities increased with increasing outcrossing rate (T_m) (Fig. 2) and decreased with increasing proportion of SC individuals. Although a similar effect was observed for π_{rep}, it was not statistically significant. Average population multilocus H_o was found to increase with increasing outcrossing rate and decrease with increasing proportion of SC individuals (Table 2). In contrast, outcrossing rate and the proportion of SC individuals did not explain the population recombination parameter ρ (Table 2).

Observed heterozygosities (H_o) calculated for each individual for the nuclear and the microsatellite datasets were highly correlated (Spearman rank correlation: $r^2 = 0.58$, $P < 0.0001$). For both datasets, the selfing phenotype (PC, SC, SI) had a significant effect on individual heterozygosity (One-way ANOVA, $F_{[2,167]} = 19.0$, $P < 0.0001$). SC individuals (mean $H_o = 0.10$ for the nuclear gene sequences, mean $H_o = 0.11$ for the microsatellites) were found to be less heterozygous than SI (mean $H_o = 0.21$ for the nuclear gene sequences, mean $H_o = 0.26$ for the microsatellites) and PC individuals (mean $H_o = 0.22$ for the nuclear gene sequences, mean $H_o = 0.25$ for the microsatellites). When the effect of selfing phenotype (SI, PC, SC) on heterozygosity was tested only considering outcrossing populations, the selfing phenotype did not have a significant effect for either the nuclear gene sequence data (overall mean $H_o = 0.22$), or for the microsatellite data (overall mean $H_o = 0.27$).

EXPLAINING THE REDUCED GENETIC DIVERSITY OF SELFING POPULATIONS: TESTING FOR BOTTLENECK EFFECTS BEYOND SELFING ALONE

After correcting π_{syn} for the differences in effective population size expected due to selfing alone, neither outcrossing rate (T_m) nor proportion of SC individuals per population explained levels of diversity ($r^2 = 0.039$, $P > 0.05$, $r^2 = 0.14$, $P > 0.05$), indicating that the neutral effects of selfing alone may explain the reduction in diversity in selfing populations (Table 2).

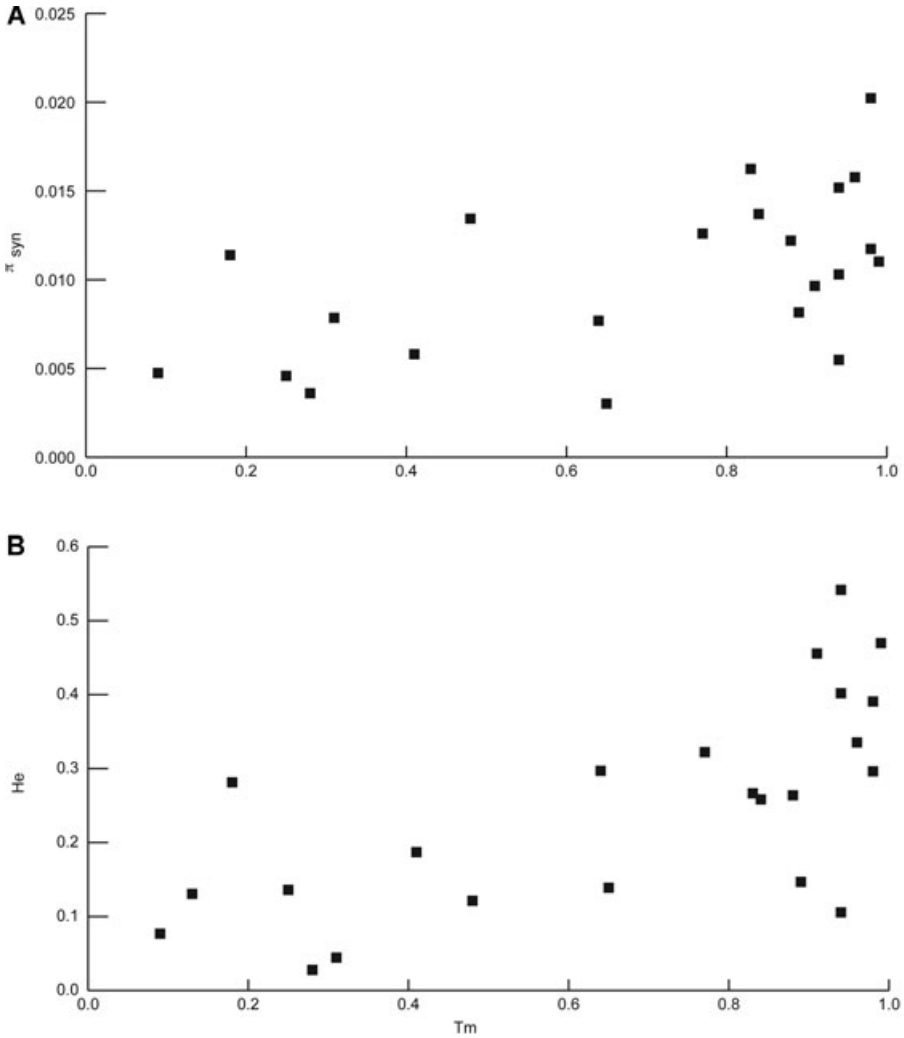

Figure 2. Comparisons between microsatellite-based multilocus outcrossing rates (T_m) and genetic diversity: (A) Synonymous diversity (π_{syn}) based on nuclear sequence data from 18 unlinked loci, where π is the average number of pairwise differences between two sequences; (B) Expected heterozygosity (H_e) estimated from microsatellite data.

Many populations showed average Tajima's D values that had significant departures from a standard neutral model; in particular, the average Tajima's D was significantly positive in 15 of our 22 populations, and significantly negative in one (Table 1). It is worth noting, however, that neither the selfing nor outcrossing populations consistently display more significant departures from neutrality.

CHLOROPLAST HAPLOTYPE DISTRIBUTION

Expanding on the results and following the same naming reported by Hoebe et al. (2009), the previously identified 741 bp L1 and L2 haplotypes and two additional 741 bp haplotypes (dubbed L3 and L4) were found among the 24 populations sampled here. In addition, we found the previously identified 515 bp S1 haplotype (Hoebe et al. 2009) and two additional (498 bp) haplotypes (dubbed S2 and S3). The shorter haplotypes S2 and S3 differed by 1 bp from one another. Based on the eight individuals per population sampled, most populations (21) were fixed for a single chloroplast haplotype, whereas three populations contained a mixture of cpDNA haplotypes (Fig. 1). Throughout the Great Lakes region, haplotypes L1, L2, L4, and S1 predominate (Fig. 1). Self-compatible individuals were found with L1, L2, L3, L4, and S1 chloroplast haplotypes and partially compatible individuals were found with haplotypes L1, L2, S1, S2, and S3. Predominantly selfing populations were associated with L1 (LPT, PTP, RON, TSSA), S1 (TC, TCA, TSSA), L3 (KTT), and L4 (WAS). Of those, only L3 was unique to selfing populations. Self-incompatible individuals, and outcrossing populations in general were found with all haplotypes except L3.

BAYESIAN CLUSTERING ANALYSES

Bayesian clustering using InStruct with estimation of individual selfing levels gave similar results as STRUCTURE (compare Figs. S1 and S2) and identified six main clusters when combining

nuclear gene sequence and microsatellite data (Fig. 1; Fig. S3). Four clusters contained both selfing and outcrossing populations, whereas the two remaining clusters contained only outcrossing populations. Clustering based on nuclear gene or microsatellite data alone in general agreed well with the results from the combined dataset, and also identified six clusters (data not shown). Four of the clusters included both selfing and outcrossing populations and there were no clusters that consisted only of selfing populations. There were only a few populations with evidence of admixture, of which the MAN and IND population had the strongest signal.

INFERENCES ON THE ORIGIN OF SELFING POPULATIONS

Only one microsatellite allele was found to be unique to the group of selfing populations ($T_m < 0.5$). The remaining 30 microsatellite alleles that occurred in the selfing group were shared with the 52 alleles occurring in the outcrossing group of populations (Table 3). There was a highly significant under-representation of unique microsatellite alleles in the selfing versus outcrossing populations (Table 3). A similar pattern emerged for the gene sequence haplotypes. Although in absolute terms there were more variants unique to the selfing group (55 of 145), the majority of variants (90) still were shared with the 449 haplotypes occurring in the group of outcrossing populations (Table 3). The group of selfing populations thus appeared to have a subset of the microsatellite alleles and gene sequences haplotypes found in the outcrossing group. None of these patterns were driven by particular microsatellite loci or nuclear genes (Table S3).

The haplotypes unique to the group of selfing populations were mostly uninformative with regard to revealing a potential origin because they were either not closely related (i.e., three or more bp difference) to any haplotype found in outcrossing populations (Table S4), or closely related (i.e., only one or two bp differences) to haplotypes that occurred in multiple outcrossing populations (Table S4). Finally, each selfing population had its own unique haplotypes (not shared with other selfing populations in our sampling) and these were never closely related to haplotypes unique to other selfing populations (Table S4). Haplotypes that were not closely related (i.e., three or more bp differences) to any haplotype in outcrossing populations or other selfing populations, occurred in all selfing populations except PTP (Table S4).

Discussion

BREAKDOWN OF SELF-INCOMPATIBILITY AND PATTERNS OF GENETIC VARIATION

Previous investigations into the levels of genetic diversity in mixed mating populations of North American *A. lyrata* revealed expected patterns, where genetic diversity in selfing populations has been reduced in comparison with outcrossing populations (Mable et al. 2005; Mable and Adam 2007; Hoebe et al. 2009). These studies were based on microsatellite variation alone whereas in this study we combine nuclear gene sequence data with microsatellites and found a striking concordance between them. We found an increase of synonymous nucleotide diversity (π_{syn}) for nuclear sequence data and H_e for microsatellite data (Table 2, Fig. 2) with increasing outcrossing rate, which corroborates previous conclusions (Mable and Adam 2007; Hoebe et al. 2009) and confirms the theoretical prediction that a shift to selfing comes at a cost to genetic diversity (Charlesworth et al. 1993; Nordborg 2000; Charlesworth and Wright 2001; Glémin et al. 2006; Wright et al. 2008). For both datasets, individual H_o was significantly lower for SC individuals versus SI individuals. This effect appeared to be solely due to the transition to inbreeding (so an effect of mating system rather than loss of SI), for when the heterozygosity of SC, PC, and SI individuals was compared excluding the inbreeding

Table 3. Total and unique number of variants (microsatellite alleles across nine loci, nuclear gene haplotypes across 18 genes) for the group of inbreeding populations and for the group of outcrossing populations; overall total number of different variants; number of variants shared across inbreeding and outcrossing populations. For the unique alleles, expected numbers under the null-hypothesis (equal proportion of unique alleles in inbreeding vs. outcrossing populations) are given in brackets. A goodness-of-fit test (G-test) was used to evaluate if the observed numbers of unique alleles significantly differed from null-expectations (*$P<0.05\cdot10^{-2}$, **$P<0.05\cdot10^{-6}$).

	Group of inbreeding populations		Group of outcrossing populations		G-test statistic	Total number of different variants	Total shared between inbreeding and outcrossing
	Total	Unique	Total	Unique			
Microsatellites (expected)	31	1 (8.59)	52	22 (14.4)	14.3*	53	30
Nuclear genes (expected)	145	55 (101.1)	449	359 (312.9)	31.7**	504	90

populations ($T_m < 0.5$), selfing phenotype (SC, PC, SI) had no effect on heterozygosity. Maintenance of high heterozygosity in SC and PC individuals in outcrossing populations emphasizes that loss of SI does not always lead to shifts to inbreeding (i.e., there is a two-step process). Even though we only sampled eight individuals per population, self-compatible individuals were found in five of the 15 predominantly outcrossing populations, and partially compatible individuals were found in nine, suggesting that self-compatibility is widespread. The comparable levels of heterozygosity of SC and SI individuals in outcrossing populations, and the lack of any clustering according to selfing phenotype (Figs. S1 and S2), also suggests that self-compatible plants in outcrossing populations, despite their ability to self-fertilize, still predominantly outcross. This is not entirely unexpected because neither the loss of SI nor the shift to selfing in *A. lyrata* is associated with a reduction in flower size (Hoebe 2009), which would promote exclusive selfing. These results are compatible with a scenario of one or more mutations facilitating the loss of SI having occurred early in the colonization history of North American *A. lyrata*, and that segregation of these mutations causes segregation of SC phenotypes in all populations, but shifts toward selfing only in a subset.

NO ROLE FOR BOTTLENECKS IN THE BREAKDOWN OF SELF-INCOMPATIBILITY OF NORTH AMERICAN *A. LYRATA*?

Population bottlenecks may be expected to be common in highly selfing populations, particularly if strong founder events were important in their origins (Foxe et al. 2009; Guo et al. 2009). If this is the case, we would expect a more severe reduction in diversity in selfing populations than expected under neutrality (i.e., a greater than twofold reduction in diversity under complete selfing). Similarly, we would expect a greater skew in the allele frequency spectrum in selfing populations, generating more positive Tajima's D values. When levels of π_{syn} were corrected for differences in N_e due to selfing alone, no significant correlation between π_{syn} and multilocus estimation of the outcrossing rate (T_m) was found. Thus, similar to recent results from *E. paniculata* (Ness et al. 2010), our data provided no evidence for reductions in diversity beyond neutral expectations, providing no signature of elevated demographic effects or hitchhiking in selfing populations. Furthermore, although 15 populations had a significantly positive Tajima's D value (Table 1), which is expected if recent bottlenecks have played a role, there was no difference in Tajima's D values between selfing and outcrossing populations.

Although population bottlenecks alone may result in a positive Tajima's D, admixture resulting from gene flow can also elevate such estimates (Wright and Gaut 2005). It is possible that positive Tajima's D values found in the outcrossing populations are the result of elevated incoming gene flow when compared to the selfing populations; although our clustering analyses do not suggest that this is generally true, larger within-population samples might reveal greater gene flow among outcrossing than selfing populations.

DEMOGRAPHIC HISTORY AND THE BREAKDOWN OF SELF-INCOMPATIBILITY OF NORTH AMERICAN *A. LYRATA*

The results from Bayesian clustering analyses for the combined nuclear sequence and microsatellite data suggest the existence of six clusters across the populations sampled from eastern North America (Fig. 1). There was good agreement between InStruct and STRUCTURE (Figs. S1 and S2). This is surprising given that the STRUCTURE assumption of random mating (Pritchard et al. 2000) is clearly violated in this system with varying levels of inbreeding, and may be reassuring for studies that have used STRUCTURE in systems in which selfing plays a role (e.g., Foxe et al. 2009; Hoebe et al. 2009). It is worth noting that the clusters appear to be fairly isolated; a signal of admixture was only particularly strong in IND and MAN. These populations are also characterized by a mixture of chloroplast haplotypes (IND, L1 and L2: cf. Fig. 1; MAN, L1 and L2: cf. Hoebe et al. 2009; Tedder et al. 2010).

The overwhelming pattern here is that populations are not clustered by mating system or selfing phenotype (Fig. 1), which would have suggested that selfing evolved only once in the region. Instead, they are predominantly clustered by geographic location. This geographic clustering of the populations by both nuclear markers and chloroplast haplotypes (Fig. 1) likely reflects the recolonization history of North America after the end of the Wisconsin glaciation (ca. 10,000 years ago), before which the entire Great Lakes region was covered in ice (Lewis et al. 2008). Hoebe et al. (2009) concluded that a mating system transition may have occurred more than once in North American *A. lyrata,* as the loss of self-incompatibility and a transition to high levels of selfing occurred in multiple chloroplast lineages (L1, S1; Hoebe et al. 2009). The more extensive geographic sampling here confirmed this and identified a new population that has undergone a complete transition to predominant selfing (KTT), which was characterized by a fourth origin based on clustering analysis and a unique chloroplast haplotype (L3).

The variation among different selfing populations within a geographic area is once again in stark contrast with the uniformity of different populations of *C. rubella* (Foxe et al. 2009; Guo et al. 2009). This may reflect that North American *A. lyrata* is at an earlier stage in the transition to selfing, or that the underlying forces driving the transition are inherently different. None of the selfing populations had a strong relationship with any of the outcrossing populations we sampled in terms of sharing of haplotypes (data not shown), contrary to what would be expected

if selfing populations had an origin in specific outcrossing populations in our sampling. Selfing populations as a group appear to harbor a subset of the genetic variation of outcrossing populations, with most haplotypes shared with outcrossing populations and a significant under-representation of unique variants (Table 3).

OUTCROSSING RATE AND THE PROPORTION OF SELF-COMPATIBLE INDIVIDUALS DO NOT ACCOUNT FOR LEVELS OF RECOMBINATION

The transition from outcrossing to selfing can result in a number of consequences at the genotypic level (reviewed in Wright et al. 2008). Selfing is associated with a decrease in levels of polymorphism and an increase in levels of linkage disequilibrium. Increased levels of homozygosity in selfing populations should lead to less-efficient recombination (r) and a reduction in N_e, thus reflected in a reduced estimate of ρ ($\rho = 4N_e r$). However, in our dataset outcrossing rate and the proportion of SC individuals did not explain the population recombination parameter ρ (Table 2). One possible explanation for this is that the power to estimate ρ in selfing populations may be diminished due to low levels of diversity. Alternatively, recent transitions to selfing may retain ancestral recombination events, especially at local physical distances (Tang et al. 2007).

WHAT FAVORS SELF-COMPATIBILITY IN THE GREAT LAKES REGION?

The question remains as to why selfing broke down and has persisted in North American *A. lyrata*, and has not occurred in the European subspecies. Selfing is only expected to become established in a previously outcrossing population when the advantages associated with selfing outweigh the costs in terms of inbreeding depression and reduced adaptability (Charlesworth and Charlesworth 1987). On a coarse scale, the geographic distribution of selfing populations is not consistent with expectations of reproductive assurance being favored in peripheral populations at the front of colonization wave (Baker's law, Baker 1955; also see Pannell and Barrett 1998), because both selfing and outcrossing populations have colonized new areas within the past 10,000 years. Selfing populations tend to be distributed toward the southern part of the habitat that opened up first following the last glacial maximum (and so are not at the periphery of the current distribution), and there does not appear to be a difference in population size or obvious habitat differences between the two types of populations (Mable and Adam 2007).

An alternative explanation is that North American populations have generally experienced reduced inbreeding depression through purging, hence lowering the cost associated with a transition to inbreeding. Population bottlenecks, particularly accompanied by long-term reductions in effective population size, can lead to significant purging and/or fixation of deleterious alleles (Bataillon and Kirkpatrick 2000). Lower levels of genetic diversity found in North American versus European *A. lyrata* have been suggested to reflect a long-term population bottleneck associated with the colonization of North America from European populations (Ross-Ibarra et al. 2008). Therefore, reduced inbreeding depression as a consequence of this bottleneck may have partly facilitated the evolution of selfing. As the Great Lakes populations of *A. lyrata* are thought to have spread north from their glacial refugia following the last Ice Age, this may have further contributed to a substantial purging of deleterious alleles in founder populations. Consistent with this possibility, it has recently been shown that population range expansion can lead to a significant decrease in genetic load in *Mercurialis annua* (Pujol et al. 2009). Comparative studies of inbreeding depression, both within North American and European populations, could enable a test of this hypothesis in *A. lyrata*. Previous studies have demonstrated high levels of inbreeding depression in European populations (Kärkkäinen et al. 1999), which does not seem to be as apparent in *A. lyrata* (Hoebe 2009).

CONCLUSIONS

In summary, we have shown that, compared to predominantly outcrossing populations, genetic diversity was reduced in selfing populations of North American *A. lyrata*. We found a strong concordance between chloroplast markers, nuclear gene sequence, and microsatellite data. The general reduction in diversity appeared to be the consequence of the transition to a selfing mating system, and not of the loss of SI alone. We found no evidence of severe bottlenecks associated with the transition to selfing beyond the bottleneck expected due to the transition itself. Although we assume that the transition to selfing in this system is recent (at least much more recent than the transition to selfing in other systems), and that there have been multiple independent origins of selfing, we did not find a clear relation of any of the selfing populations to a particular outcrossing "parent" population. The genetic basis for the loss of SI has not yet been elucidated in *A. lyrata* but it is important to identify potentially unique origins of self-compatibility prior to designing and interpreting crossing studies to investigate mechanisms of loss. For example, initial investigations of loss of SI in *A. thaliana* suggested a selective sweep across all populations (Shimizu et al. 2004); later studies that have sampled populations more broadly have found that the story is more complicated, with the possibility of multiple independent losses, perhaps involving different mechanisms (Bechsgaard et al. 2006; Sherman-Broyles et al. 2007; Tang et al. 2007). The North American *A. lyrata* populations offer an exciting system to unravel the causes and consequences of the loss of SI and an evolutionary transition from outcrossing to selfing, as it involves a much more recent change than that which has given rise to the highly selfing *A. thaliana* species.

ACKNOWLEDGMENTS

We thank Y. Willi and D. Remington for generously providing seeds and outcrossing rates; A. Adam, P. Hoebe, and H.-G. Zha for assistance with laboratory work and plant maintenance; P. Hoebe, R. Ness, and T. Slotte for useful discussions; J. Hughes for assistance with Perl scripting; H. Gao for providing assistance with the InStruct source code; three anonymous referees for their constructive and detailed comments that greatly improved the manuscript. We are grateful for funding from the Natural Environment Research Council (NE/D013461/1) and European Research Area Network in Plant Genomics/Biotechnology and Biosciences Research Council joint funding (ERAPGFP-06.058A) to BKM. We thank Parks Canada, Ontario Parks, Michigan State Parks Authority, U.S. National Park Service, Ohio Department of Natural Resources, and the Ohio Nature Conservancy for access to protected park areas and advice on plant locations.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Ansell, S. W., H. Schneider, N. Pedersen, M. Grundmann, S. J. Russell, and J. C. Vogel. 2007. Recombination diversifies chloroplast trnF pseudogenes in *Arabidopsis lyrata*. J. Evol. Biol. 20:2400–2411.

Baker, H. G. 1955. Self-compatibility and establishment after long-distance dispersal. Evolution 9:347–349.

Barrett, S. C. H. 2002. The evolution of plant sexual diversity. Nat. Rev. Genet. 3:274–284.

Barrett, S. C. H., L. D. Harder, and A. Worley. 1996. The comparative biology of pollination and mating in flowering plants. Phil. Trans. R. Soc. Ser. B. 351:1271–1280.

Bataillon, T., and M. Kirkpatrick. 2000. Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. Genet. Res. 75:75–81.

Baudry, E., C. Kerdelhue, H. Innan, and W. Stephan. 2001. Species and recombination effects on DNA variability in the tomato genus. Genetics 158:1725–1735.

Bechsgaard, J. S., V. Castric, D. Charlesworth, X. Vekemans, and M. H. Schierup. 2006. The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. Mol. Biol. Evol. 23:1741–1750.

Boggs, N. A., J. B. Nasrallah, and M. E. Nasrallah. 2009. Independent S-locus mutations caused self-fertility in *Arabidopsis thaliana*. PLoS Genetics 5:e1000426.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effects of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Charlesworth, D. 2006. Evolution of plant breeding systems. Curr. Biol. 16:R726–R735.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18:237–268.

Charlesworth, D., and X. Vekemans. 2005. How and when did *Arabidopsis thaliana* become highly self-fertilising. Bioessays 27:472–476.

Charlesworth, D., and S. I. Wright. 2001. Breeding systems and genome evolution. Curr. Opin. Genet. Dev. 11:685–690.

Charlesworth, D., and Z. Yang. 1998. Allozyme diversity in *Leavenworthia* populations with different inbreeding levels. Heredity 81:453–461.

Chiang, Y. H., B. A. Schaal, C. H. Chou, S. Huang, and T. Y. Chiang. 2003. Contrasting selection modes at the *Adh1* locus in outcrossing *Miscanthus sinensis* vs. inbreeding *Miscanthus condensatus* (Poaceae). Am. J. Bot. 90:561–570.

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

Clauss, M. J., and T. Mitchell-Olds. 2006. Population genetic structure of *Arabidopsis lyrata* in Europe. Mol. Ecol. 15:2753–2766.

Cutter, A. D., and B. A. Payseur. 2003. Selection at linked sites in the partial selfer *Caenorhabditis elegans*. Mol. Biol. Evol. 20:665–673.

Darwin, C. R. 1876. The effects of cross and self-fertilization in the vegetable kingdom. John Murray, London.

Dieringer, D., and C. Schlotterer. 2003. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Res. 13:2242–2251.

Filatov, D. A., and D. Charlesworth. 1999. DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus. Genetics 153:1423–1434.

Fisher, R. A. 1941. Average excess and average effect of a gene substitution. Ann. Eugen. 11:53–63.

Foxe, J. P., T. Slotte, E. A. Stahl, B. Neuffer, H. Hurka, and S. I. Wright. 2009. Rapid morphological evolution and speciation associated with the evolution of selfing in *Capsella*. Proc. Natl. Acad. Sci. USA 106:5241–5245.

Gao, H., S. Williamson, and C. D. Bustamante. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635–1651.

Glémin, S., E. Bazin, and D. Charlesworth. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc. R. Soc. B. 273:3011–3019.

Grant, V. 1981. Plant Speciation. Columbia Univ. Press, New York.

Guo, Y.-L., J. S. Bechsgaard, T. Slotte, B. Neuffer, M. Lascoux, D. Weigel, and M. H. Schierup. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. Proc. Natl. Acad. Sci. USA 106:5246–5251.

Hoebe, P. N. 2009. Evolutionary dynamics of mating systems in populations of North American *Arabidopsis lyrata*. Dissertation, 164pp. University of Glasgow.

Hoebe, P. N., M. Stift, A. Tedder, and B. K. Mable. 2009. Multiple losses of self-incompatibility in North-American *Arabidopsis lyrata*: phylogeographic context and population genetic consequences. Mol. Ecol. 18:4924–4939.

Hudson, R. R. 2001. Two-locus sampling distributions and their application. Genetics 159:1805–1817.

Hurka, H., and B. Neuffer. 1997. Evolutionary processes in the genus *Capsella* (*Brassicaceae*). Plant Syst. Evol. 206:295–316.

Igic, B., R. Lande, and J. R. Kohn. 2008. Loss of self-incompatibility and its evolutionary consequences. Int. J. Plant Sci. 169:93–104.

Ingvarsson, P. K. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. Evolution 56:2368–2373.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Kalisz, S., D. W. Volger, and K. M. Hanley. 2004. Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. Nature 430:884–887.

Kärkkäinen, K., H. Kuittinen, R. Van Treuren, C. Vogl, S. Oikarinen, and O. Savolainen. 1999. Genetic basis of inbreeding depression in *Arabis petraea*. Evolution 53:1354–1365

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley, and J. Hey. 2000. The Population Genetics of the Origin and Divergence of the *Drosophila simulans* Complex Species. Genetics 156:1913–1931.

Koch, M., and M. Kiefer. 2005. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species—*Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. Am. J. Bot. 92:761–767.

Lewis, C. F. M., P. F. Karrow, S. M. Blasco, F. M. G. McCarthy, J. W. King, T. C. Moore, Jr., and D. K. Rea. 2008. Evolution of lakes in the Huron basin: deglaciation to present. Aquat. Ecosys. Health Manage. 11:127–136.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

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

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

Liu, P., S. Sherman-Broyles, M. E. Nasrallah, and J. B. Nasrallah. 2007. A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. Curr. Biol. 17:734–740.

Mable, B. K., and A. Adam. 2007. Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. Mol. Ecol. 16:3565–3580.

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

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–25.

Muller, M. H., J. Leppala, and O. Savolainen. 2008. Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. Heredity 100:47–58.

Ness, R. W., S. I. Wright, and S. C. H. Barrett. 2010. Mating-System Variation, Demographic History and Patterns of Nucleotide Diversity in the Tristylous Plant *Eichhornia Paniculata*. Genetics 184:381–392.

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

Pannell, J. R., and S. C. H. Barrett. 1998. Baker's law revisited: reproductive assurance in a metapopulation. Evolution 52:657–668.

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

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pujol, B., S. R. Zhou, J. Sanchez Vilas, and J. R. Pannell. 2009. Reduced inbreeding depression after species range expansion. Proc. Natl. Acad. Sci. USA 106:15379–15383.

Ritland, K. 2002. Extensions of models for the estimation of mating systems using n independent loci. Heredity 88:221–228.

Ritland, K., and S. K. Jain. 1981. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. Heredity 47:35–52.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.

Ross-Ibarra, J., S. I. Wright, J. P. Foxe, A. Kawabe, L. DeRose-Wilson, G. Gos, D. Charlesworth, and B. S. Gaut. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. PLoS One 3:e2411.

Savolainen, O., C. H. Langley, B. P. Lazzaro, and H. Freville. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol. Biol. Evol. 17:645–655.

Sherman-Broyles, S., N. Boggs, A. Farkas, P. Liu, J. Vrebalov, M. E. Nasrallah, and J. B. Nasrallah. 2007. S locus genes and the evolution of self-fertility in *Arabidopsisthaliana*. Plant Cell 19:94–106.

Shimizu, K. K., J. M. Cork, A. L. Caicedo, C. A. Mays, R. C. Moore, K. M. Olsen, S. Ruzsa, G. Coop, C. D. Bustamante, P. Awadalla, et al. 2004. Darwinian selection on a selfing locus. Science. 306:2081–2084.

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia Univ. Press, New York.

———. 1956. Taxonomy and evolution of genera with special reference to family *Gramineae*. Evolution 10:235–245.

———. 1957. Self-fertilization and population variability in higher plants. Am. Nat. 41:337–354.

———. 1970. Adaptative radiation of reproductive characteristics in angiosperms. I. Pollination mechanisms. Ann. Rev. Ecol. Syst. 1:307–326.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant. Mol. Biol. 17:1105–1109.

Takebayashi, N., and P. L. Morrell. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. Am. J. Bot. 88:1143–1150.

Tang, C., C. Toomajian, S. Sherman-Broyles, V. Plagnol, Y. L. Guo, T. T. Hu, R. M. Clark, J. B. Nasrallah, D. Weigel, and M. Nordborg. 2007. The evolution of selfing in *Arabidopsis thaliana*. Science 317:1070–1072.

Tedder, A., P. N. Hoebe, S. W. Ansell, and B. K. Mable. 2010. Using chloroplast trnF pseudogenes for phylogeography in *Arabidopsis lyrata*. Diversity 2:653–678.

Tsuchimatsu T., K. Suwabe, R. Shimizu-Inatsugi, S. Isokawa, P. Pavlidis, T. Städler, G. Suzuki, S. Takayama, M. Watanabe, and K. K. Shimizu. 2010. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. Nature 464:1342–1346.

Wright, S. I., J. P. Foxe, L. DeRose-Wilson, A. Kawabe, M. Looseley, B. S. Gaut, and D. Charlesworth. 2006. Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. Genetics 174:1421–1430.

Wright, S. I., and B. S. Gaut. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol. Biol. Evol. 22:506–519.

Wright S. I., B. Lauga, and D. Charlesworth. 2003. Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Mol. Ecol. 12:1247–1263.

Wright, S. I., R. W. Ness, J. P. Foxe, and S. C. H. Barrett. 2008. Genomic consequences of outcrossing and selfing in plants. Int. J. Plant Sci. 169:105–118.

Associate Editor: J. Pannell

Supporting Information

The following supporting information is available for this article:

Figure S1. Posterior probabilities of Bayesian clustering analysis (using STRUCTURE, 2,000,000 generations with a burnin of 200,000), using a combination of the nuclear haplotype and microsatellite data.

Figure S2. Posterior probabilities of Bayesian clustering analysis (using InStruct, 2,000,000 generations with a burnin of 200,000), using a combination of the nuclear haplotype and microsatellite data.

Figure S3. Distribution of the −ln probability (−ln(P), represented by closed circles) and its variance (Var[ln(P)], represented by open squares) for Bayesian clustering analysis (2,000,000 generations with a burnin of 200,000, five chains for each setting of the number of predefined clusters (k), k ranging from 1 to 12) with (A) InStruct and (B) STRUCTURE.

Table S1. Explanation of population abbreviations, lakefront (where relevant), and geographic specifications (state/province, country and coordinates) for each of the 24 populations used in this study (more details regarding sampling method are given in the Supporting information).

Table S2. Population outcrossing rates (T_m) based on progeny arrays genotyped for nine microsatellite loci calculated using MLTR version 2.3 (Ritland 2002), along with single-locus estimates (T_s), the difference between T_m and T_s) (standardized by T_m) as an indication of biparental inbreeding, as well as the number of families (N families) and individuals (N individuals) on which the estimates are based; the proportion of individuals in each population that were SI (0 or 1 full sized siliques/6), SC (5 or 6 full sized siliques/6) or PC (2–4 full sized siliques/6) based on controlled self-pollinations of eight individuals per population.

Table S3. Number of microsatellite alleles (for each of nine loci) and of nuclear gene haplotypes (for each of 18 genes): total and unique within the group of inbreeding populations (KTT, LPT, PTP, RON, TC, TCA, TSSA, WAS) and outcrossing populations (BEI, HDC, IND, IOM, LSP, MAN, NCM, OWB, PCR, PIC, PIN, PIR, PRI, PUK, SBD, TSS); total across all populations and shared between inbreeding and outcrossing populations. Overall totals are given in bold for microsatellite loci and gene sequences.

Table S4. The number of haplotypes across 18 genes unique to the group of inbreeding populations (T_m<0.5) per population per gene.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is 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 corresponding author for the article.