Molecular Biology and Evolution

Research article

Demographic Signatures Accompanying the Evolution of Selfing in *Leavenworthia alabamica*

Jeremiah W. Busch,*†,[1] Simon Joly,†,[2,3] and Daniel J. Schoen[2]

[1]School of Biological Sciences and The Center for Reproductive Biology, Washington State University

[2]Biology Department, McGill University, Montréal, QC, Canada

[3]Institut de recherche en biologie végétale, Université de Montréal and Montreal Botanical Garden, Montréal, QC, Canada

†These authors contributed equally to this work.

***Corresponding author:** E-mail: jwbusch@wsu.edu.

Associate editor: Naoki Takebayashi

Abstract

The evolution of selfing from outcrossing is a common transition, yet little is known about the mutations and selective factors that promote this shift. In the mustard family, single-locus self-incompatibility (SI) enforces outcrossing. In this study, we test whether mutations causing self-compatibility (SC) are linked to the self-incompatibility locus (S-locus) in *Leavenworthia alabamica*, a species where two selfing races (a2 and a4) co-occur with outcrossing populations. We also infer the ecological circumstances associated with origins of selfing using molecular sequence data. Genealogical reconstruction of the *Lal2* locus, the putative ortholog of the *SRK* locus, showed that both selfing races are fixed for one of two different S-linked *Lal2* sequences, whereas outcrossing populations harbor many S-alleles. Hybrid crosses demonstrated that S-linked mutations cause SC in each selfing race. These results strongly suggest two origins of selfing in this species, a result supported by population admixture analysis of 16 microsatellite loci and by a population tree built from eight nuclear loci. One selfing race (a4) shows signs of a severe population bottleneck, suggesting that reproductive assurance might have caused the evolution of selfing in this case. In contrast, the population size of race a2 cannot be distinguished from that of outcrossing populations after correcting for differences in selfing rates. Coalescent-based analyses suggest a relatively old origin of selfing in the a4 race (~150 ka ago), whereas selfing evolved recently in the a2 race (~12–48 ka ago). These results imply that S-locus mutations have triggered two recent shifts to selfing in *L. alabamica*, but that these transitions are not always associated with a severe population bottleneck, suggesting that factors other than reproductive assurance may play a role in its evolution.

Key words: automatic selection, reproductive assurance, self-incompatibility, population bottlenecks, *SCR*, *SRK*.

Introduction

Mating systems have a major influence on population genetic structure, population viability, and the evolutionary process (Stebbins 1957; Clegg 1980; Glemin et al. 2006; Wright et al. 2008), and many studies have attempted to compare the genetic and ecological circumstances associated with the transition from outcrossing to selfing (Lloyd 1965; Wyatt 1988). The majority of flowering plant species possess mechanisms that prevent selfing, such as self-incompatibility (SI), dioecy, and the temporal or spatial separation of male and female sexual functions within the flower or plant (Barrett 2002). These mechanisms for enforcing outcrossing are thought to be adaptive because they prevent inbreeding depression (Charlesworth and Charlesworth 1987). In many angiosperms, however, outcrossing mechanisms have often been lost, resulting in a shift in the mating system to partial or near complete self-fertilization (Igic et al. 2008). The loss of mechanisms enforcing outcrossing is one of the most common evolutionary trends in the flowering plants, having occurred in nearly every major lineage (Stebbins 1974), yet selfing taxa are generally believed to be evolutionary dead-ends (Stebbins 1957), an idea supported by estimates of extinction rates (Goldberg et al. 2010). The existence of many selfing lineages, coupled with the high estimated extinction rates of these lineages, suggests that transitions from outcrossing to selfing must occur at high rates and that extant selfing taxa should often be recently derived.

The spread and fixation of genes for selfing in response to natural selection is thought to occur when the benefits of selfing exceed the detrimental effects of increased homozygosity and inbreeding depression (Holsinger 2000). Although a number of mechanisms have been proposed to explain this shift (Jain 1976), two competing hypotheses are thought to be important (Uyenoyama et al. 1993). One advantage of selfing, as noted by Darwin (1876) and Baker (1955), arises from the ability to produce seeds when pollinators are scarce and has been referred to as the "reproductive assurance hypothesis." The other advantage, the so-called "automatic selection hypothesis," arises because a mutation causing selfing enjoys up to a 50% transmission advantage (Fisher 1941) over outcrossing alleles. This advantage occurs because a gene promoting selfing in a population can be transmitted, on average, in two doses through selfing as well as in an additional dose via male gametes through outcrossing. On average, a gene causing outcrossing is only transmitted in two doses, as it

contributes only one gamete as a pollen and seed parent. These hypotheses make different predictions regarding the population genetic signatures accompanying the evolution of selfing. Under the reproductive assurance hypothesis, selfing is expected to be accompanied by a strong population bottleneck. On the other hand, when selfing evolves via automatic selection, the effective population size is expected to be reduced by 50% at most (Schoen et al. 1996; Nordborg 2000). Distinguishing among these hypotheses addresses the more general question of whether major evolutionary transitions in plants are driven by ecological factors or intrinsic genetic ones.

With the application of molecular approaches, we are gaining deeper insight into how and why selfing evolves. Evidence of recent independent derivations of selfing has been uncovered in *Arabidopsis thaliana* (Shimizu et al. 2008), where multiple mutations have been shown to cause selfing (Boggs et al. 2009). Outside of this model species, the mutations that underlie the shift in mating system have rarely been characterized (Nasrallah et al. 2007). Consequently, we have few clear data that pertain to the number of times selfing has evolved in a species. Indeed, although several studies suggest multiple independent origins of selfing (e.g., Foxe et al. 2010; Ness et al. 2010), it is rarely clear exactly how the selfing phenotype arose (Guo et al. 2009). This poses a particular challenge in the case of more ancient selfing taxa, as there may be multiple mutations underlying the selfing syndrome, many of which may have arisen after the initial breakdown of outcrossing (Kondo et al. 2002; Igic et al. 2008). Studies of the number of times selfing has evolved are informative as to the transition rate from outcrossing to selfing, which in turn is relevant to models describing the distribution of mating systems and in testing models of species-level selection of the mating system (Igic et al. 2008; Schoen and Busch 2008).

Population genetic studies have been informative on the demographic circumstances involved in the evolution of selfing. The transition to selfing is sometimes associated with a strong population bottleneck, as inferred in *Capsella rubella* (Foxe et al. 2009; Guo et al. 2009), *A. thaliana* (Tang et al. 2007), and species of *Solanum* and *Leavenworthia* (Liu et al. 1998; Baudry et al. 2001; Roselius et al. 2005). In contrast, selfing populations of *A. lyrata* (Foxe et al. 2010) and *Eichhornia paniculata* (Ness et al. 2010) have apparently not experienced strong population bottlenecks. But it is important to note that these inferences could be misleading, as a strong population bottleneck can occur after the evolution of selfing, and thereby give the false impression that selfing evolved under reproductive assurance. Alternatively, incoming gene flow to a selfing population could mask evidence of a bottleneck. Recent coalescent methods for analyzing and interpreting patterns of nucleotide polymorphism help to remove some of the uncertainty by disentangling the effects of migration and divergence (Hey and Nielsen 2004; Becquet and Przeworski 2007) or by reconstructing historical population sizes through time (Heled and Drummond 2008).

To evaluate hypotheses for the evolution of selfing, it is useful to study species where selfing has evolved recently, as the first mutation(s) causing selfing are more likely to be identified and the genetic signatures associated with the shift to selfing are less likely to be confounded by later events. Species where the molecular genetic details underlying the control of SI have been worked out are particularly useful in this regard (Bechsgaard et al. 2006). In such systems, the evolutionary dynamics are relatively straightforward; for example, because S-alleles are maintained by strong negative frequency-dependent selection (Wright 1939; Billiard et al. 2007) and S-linked markers can be used to test hypotheses on the genetic basis of selfing. In these cases, crosses between SC and SI plants may reveal whether S-alleles naturally found in selfing populations harbor or are linked to the mutation causing SC. If the mutation causing SC is linked to the S-locus, the S-allele of the SC parent will cosegregate with the SC phenotype in hybrid offspring. Even if the mutation causing SC is not contained within the S-linked marker per se, studies of this marker provide the means to determine if the SC mutation is S-linked, as there is strong recombination suppression between genes in the S-locus region (Hagenblad et al. 2006).

In *L. alabamica*, both outcrossing and selfing populations occur over a very small geographic range, and these have been well characterized in terms of their morphology, ecology, and mating system (Lloyd 1965; Busch 2005a). Recently, the putative S-locus receptor kinase (*SRK*) ortholog, *Lal2*, was characterized in this species and shown to cosegregate with S-locus specificities in diallel crosses (Busch et al. 2008). Populations of *L. alabamica* have been classified into different races based on the presence of SI and floral morphology (Lloyd 1965). Outcrossing populations belong to "race a1" in which plants are primarily SI and have large flowers, whereas the selfing populations studied fall into "race a2" or "race a4." The a4 race is typified by plants with small flowers and introrse anthers that promote spontaneous selfing, whereas a2 plants have flowers that are closer in size to those of the outcrossing race. Given the co-occurrence of populations with SI and SC plants in *L. alabamica* and the fact that the species is probably less than 300,000 years old (Joly and Schoen, submitted), shifts to selfing probably occurred recently (Koelling et al. 2010). The molecular level characterization of SI in *L. alabamica*, together with the recent derivation of selfing, provides an excellent opportunity to identify the mutation(s) causing SC and to help infer the population-level selective factors associated with this shift in the mating system.

Methods

Plant Material

Seeds were collected from natural populations in the spring of 2002 and 2007 (fig. 1). In 2002, 20 siliques were collected from different maternal plants in the small-flowered selfing populations of race a4 (Tuscumbia, Russellville, and Morgan) and a large-flowered selfing population (Landersville) from the a2 race. Seeds were collected from SI populations

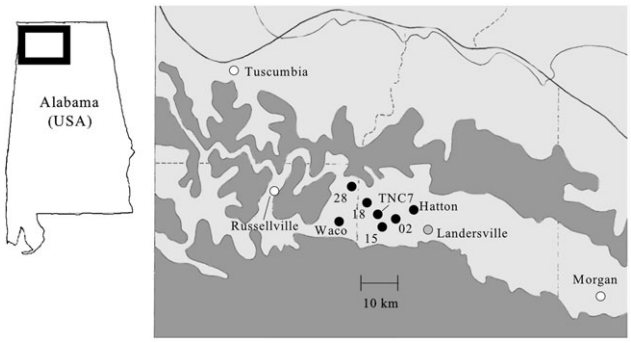

FIG. 1. Outcrossing and selfing populations of *Leavenworthia alabamica*. Black circles (race a1) denote outcrossing populations where most plants are SI. Plants in Landersville (grey circle; race a2) are SC yet retain floral adaptations for outcrossing. Plants in the remaining populations (white circles; race a4) are SC and possess adaptations for self-pollination.

of the a1 race (TNC7, Waco, 15, 18, 28, and 02) during the spring of 2007. The geographic locations of these populations are reported elsewhere (Busch 2005b; Busch et al. 2010). Material for genotyping and sequencing was extracted from either seedlings or seeds near germination using Qiagen DNeasy Plant Mini Kits (Qiagen, Valencia, CA).

S-locus Genealogy and Population Diversity

A portion of the highly variable S-domain (exon1) of the *Lal2* gene was sequenced for several individuals in selfing populations: Morgan ($n = 11$ individuals), Tuscumbia ($n = 12$), Russelville ($n = 11$), and Landersville ($n = 15$), following Busch et al. (2008). An additional 15 individuals from Landersville were genotyped by single strand conformation polymorphism (SSCP) to screen for different S-alleles (see Busch et al. 2010). The *Lal2* sequences from SC individuals were aligned with previously published *Lal2* sequences obtained from outcrossing populations (Busch et al. 2008; Joly and Schoen, accepted) using MUSCLE (Edgar 2004). A maximum likelihood genealogy of *Lal2* was reconstructed with the software phyml (Guidon and Gascuel 2003) using a HKY + Γ_4 model, estimating all parameters during the search.

The Genetic Basis of Selfing

Two separate crosses were conducted to determine whether the mutations associated with SC in the Tuscumbia (race a4) and Landersville (race a2) populations are linked to the S-locus. To help determine the genetic basis of SC in the Tuscumbia population, an SI plant from the Waco (race a1) population, heterozygous at the *Lal2* gene, was first forcibly self-pollinated in bud. Seeds were collected from this cross, and an offspring plant homozygous for the *Lal2-2* allele was then crossed to a plant homozygous for the *Lal2-Tu* allele (from the Tuscumbia population). The resulting F_1 was self-compatible and autonomously produced F_2 offspring. To help determine the basis of SC in the Landersville (race a2) population, a single plant was crossed with a self-incompatible Waco (race a1) plant, and F_2 progeny were raised in the McGill University greenhouses as described above. Progeny from both sets of crosses were scored for autonomous fruit production and pollen tube growth following forced self-pollination (Busch et al. 2010) and were genotyped at the *Lal2* locus.

Genotyping of the Waco × Tuscumbia F_2 progeny was done using DNA extracted from F_1 and F_2 rosettes as described previously. Allele-specific primers were designed to genotype 10 F_1 and 64 F_2 plants in the cross involving plants from the Tuscumbia population (W16B-2Fnew: GAGCAATACAGAACCTTTTGTGG and W16B-2Rnew: TCAATCCCCGTGCATATTGG) and *Lal2-Tu* (a4_Lal2F: CGAGCTTATATGATCCATCAAG and a4_Lal2R: CTCCT ATTAACTCCGTCAGCC). Polymerase chain reactions (PCRs) were conducted using 1.5 mM $MgCl_2$ and were run for 4 min 30 s at 94.0 °C, followed by 30 cycles with a 30 s denaturation at 94.0 °C, 45 s annealing at 48.0 °C, and 1 min extension at 72.0 °C. Because these PCRs produced bands of different size, amplicons for each individual plant were combined and visualized on 1.0% agarose gels in 1× Tris-borate-EDTA buffer. For the progeny from the Waco × Landersville cross, S-linked *Lal2* genotypes were inferred using SSCP analysis for 25 F_2 plants (following Busch et al. 2010). Cosegregation of the SC phenotype and autonomous fruit production with each of the *Lal2* sequences sampled from selfing populations was then tested.

Plants in each of the two crosses described above were used in a full diallel to examine patterns of cross-incompatibility, as this approach can help to reveal whether the mutation causing SC occurs in the pollen or pistil gene within the S-locus. Plants were assigned random numbers to avoid any potential sources of bias in scoring cross-incompatibility. Following emasculation, flowers were cross-pollinated and marked with jeweler's tags. Stigmas were harvested 1 day following cross-pollination and fixed in 10% acetic acid–ethanol overnight at 4 °C. Pistils were cleared in 37 °C NaOH for 1 h and stained using a 0.1% aniline blue solution. Fluorescence microscopy with a 4′-6′ diamidino-2-phenylindole filter and a digital camera was used to visualize pollen tube growth.

Population Genetic Structure and Selfing Rate Estimates

Between 10 and 20 plants from outcrossing populations and all the selfing populations were genotyped at 15 or 16 microsatellite loci using established protocols (Molecular Ecology Resources Primer Development Consortium et al. 2009). PCR was carried out using a Bio-Rad C1000 thermal cycler (Bio-Rad Laboratories, Inc., Hercules, CA) using locus-specific running conditions (Busch et al. 2010). Primers were labeled with the florescent dyes 6FAM, NED, PET, or VIC (Applied Biosystems, Foster City, CA). Fragment length analysis was performed with an Applied Biosystems 3730 capillary sequencer and interpreted using Peak Scanner software (Applied Biosystems, Foster City, CA). Allele sizes were scored manually, and ambiguous results were discarded.

The genetic structure of populations was inferred using InStruct (Gao et al. 2007), which uses a model that is similar to STRUCTURE (Pritchard et al. 2000) but does not enforce

random mating, making it suitable for the study of populations with inbreeding. We used microsatellite data for this analysis but did not include sequence data (see below), as there was incomplete genotyping of individuals across the eight loci that were sequenced. We used the population structure and population selfing rates model, which infers the maximum likelihood estimate of a cluster's selfing rate by estimating the number of generations before an outcrossing event for individuals. This estimate of the selfing rate is based solely on heterozygote deficiency, and so is likely to be upwardly biased, as other factors (e.g., population subdivision) reduce heterozygosity. We ran simulations for two to ten clusters. Each simulation consisted of 200,000 burn-in chains followed by 200,000 iterations and was replicated ten times. All other options were left at default settings. The most likely number of clusters was determined by the delta-*K* method of Evanno et al. (2005).

Relationships of the Selfing Populations and Estimation of Population Sizes

To help resolve relationships among the outcrossing and selfing populations, we reconstructed a population tree from nuclear gene sequences using *BEAST (Heled and Drummond 2010), an approach based on the multispecies coalescent. This method allows population size changes along branches of the phylogeny and incorporates sequences from multiple genes (not necessarily from the same individuals). Sequences were obtained for eight nuclear genes using the SSCP approach described in detail in Busch et al. (2010). Briefly, amplicons (<450 bp) were obtained and run on SSCP gels to separate the alleles. Bands representing single alleles were cut out from the gel, re-amplified, and sequenced. Unambiguous sequences were retained for the analyses.

Only nonrecombining portions of genes were used for the *BEAST analysis. The complete alignments obtained using the MUSCLE algorithm were tested for the presence of recombination using the PHI test (Bruen et al. 2006), as implemented in SplitsTree4 (Huson and Bryant 2006). For genes that tested positive for recombination ($P < 0.05$; we did not correct the significance level for multiple tests as we preferred to be conservative in eliminating regions that have undergone recombination), the alignments were further analyzed using the software DualBrothers (Minin et al. 2005), in Geneious version 1.4.8 (Drummond et al. 2010), to detect the recombination breakpoints. Portions of the genes containing the majority of single nucleotide polymorphisms (SNPs) and that appeared to lack recombination (as inferred using DualBrothers) were selected and retested for recombination with the PHI test to ensure that there was no evident recombination.

For the *BEAST analysis, substitutions models, clock models, and trees were unlinked among genes. A strict molecular clock was enforced for all genes and the rates were estimated relative to the *Adh1* gene. The Hasegawa, Kishino and Yano (HKY) substitution model was used for all genes, estimating base frequencies during the search. Several priors and operators were modified (the .xml file with the prior and operator modifications in our analyses is available as supplementary Material). Briefly, the following priors were used: root.height = gamma(0.15, 0.25), pop.mean = gamma(0.3, 0.1), clock.rate = gamma(2, 1), and kappa = gamma(2, 3). Two independent chains were run for 1×10^8 steps and sampled every 5,000 steps. The first 10% of samples were removed as burn-in. Convergence always occurred much before this point. Convergence among chains was determined visually using Tracer and statistically using Gelman's *R* (Gelman et al. 2009) as implemented in the package coda (Plummer et al. 2010) in R (R Development Core Team 2009). Sequences from the sister species *L. crassa* (Joly and Schoen accepted) were included in the analyses to root the tree.

*BEAST does not account for migration rates among populations, and thus, it is possible that estimates of divergence times and population sizes obtained with this method are biased. To investigate whether this is the case, we also estimated divergence times and population sizes using the MIMAR software (Becquet and Przeworski 2007), which uses an "isolation with migration model." We used the same gene regions as for the *BEAST analyses, which did not show evidence of recombination. We ran two independent Markov chain Monte Carlo of 5 million generations with samples taken every 200 steps, removing the first 50,000 generations as burn-in (the chains always reached stationary phase well before that point). Flat priors were used for θ (actual: 0.0001 − 0.05 and ancestral: 0.0001 − 0.1) and divergence times ($10 - 1 \times 10^6$ years) and for the symmetric migration rate parameter $M(e^{-10} - e^2)$. The mutation rates of loci (relative to *Adh1*) were the mean of the posterior distributions obtained from the *BEAST analyses (*Adh2*: 1.475, *Adh3*: 0.639, *Lal8*: 1.983, *MS* 2.372, *MtN21a*: 1.29, *MtN21b*: 3.262, and *PRK*: 0.81). Two hundred genealogies were simulated at each step. Tuning parameters were set to 2×10^{-3}, 2×10^{-3}, 2×10^5, 2×10^{-3}, 0.5, and 0.5. Convergence was assessed using the same tools as for the population tree.

Reconstruction of Ancestral Population Sizes

We used extended Bayesian skyline plots (EBSPs; Heled and Drummond 2008) to reconstruct the evolution of ancestral population sizes of *L. alabamica* populations. EBSP is based on gene genealogies to estimate population sizes at different points in time. It uses a Bayesian MCMC simulation approach to integrate over gene genealogies and estimate the parameters of the model. An advantage of the EBSP approach over others is that the user does not have to provide information on the number of population size changes in the past—the procedure estimates this parameter from the data and produces plots of the most probable population sizes at different times in the past. For each population, we included all data sets that contained at least one SNP in the studied population (a prerequisite of the program). Using the same approach as described above for the population trees, we retained only those portions of genes that did not show evidence of recombination within each population. Substitution models (HKY for all loci), clock models (strict clock for all loci), and trees were unlinked

among genes. Molecular clocks were estimated relative to the *Adh1* gene. Two independent runs starting from different random starting points were run for 5×10^8 steps, sampling every 10,000 steps. The first 10% samples were discarded as burn-in. Convergence was assessed using the same tools as for the population tree.

Molecular Clock Calibration

The EBSPs, the MIMAR analyses, and the *BEAST analysis used the *Adh1* gene as reference with the rates of the other markers estimated relative to *Adh1*. This allows the time scales of the analyses to be directly compared. We thus calibrated the molecular clock of *Adh1* to be able to approximately date the evolutionary events on a temporal scale. We used the median sequence divergence of 0.001341 for *Adh1* from the common ancestor of *L. alabamica* and *L. crassa* populations to the present as estimated in *BEAST (see Results) and assumed that these species diverged 311,000 years ago following Joly and Schoen (accepted), an estimate inferred with a modified version of the MIMAR software (Becquet and Przeworski 2007) using the commonly used synonymous substitution rate of 1.5×10^{-8} (Koch et al. 2000). This gives a molecular clock for the *Adh1* gene in *L. alabamica* of 4.31×10^{-9} substitutions per site per year.

Genetic Diversity

Nucleotide diversity at synonymous sites (π_{syn}) averaged over all eight nuclear genes was estimated for each population. Because selfing reduces effective population sizes and thus genetic diversity ($\theta = 4N_e\mu$) (Nordborg 2000), it is of interest to compare genetic diversity among populations for a given level of population inbreeding. The genetic diversity of a population with selfing (Θ_S) with an inbreeding coefficient F could be compared with the genetic diversity of a completely outcrossing population (Θ; $F = 0$) according to the equation $\Theta_S = \Theta/(1 + F)$, where $F = s/(2 - s)$ and considering that s is the selfing rate (Pollak 1987). Given that the observed genetic diversity in populations corresponds to θ_S, θ can be obtained by multiplying θ_S by $(1 + F)$. Using π_{syn} as an estimate of θ_S, we estimated adjusted θ assuming that $s = 0$ in the SI populations (this is supported by recent estimates of the selfing rate from progeny arrays; Herman A and Schoen DJ, unpublished data) and $s = 0.523$ for the Landersville population, as estimated from progeny arrays (Busch et al. 2010). We do not have a precise selfing estimate for the a4 race populations, but this did not affect our results because there was no genetic variation in these populations (see below).

Results

Lal2 Genealogy and Diversity

In contrast to the 20–50 S-alleles maintained in the SI populations (Busch et al. 2010), only a single *Lal2* allele was found in each of the selfing populations (fig. 2). All the small-flowered individuals (race a4) sampled share the same S-allele. The *Lal2* sequences in Russellville and Tuscumbia were identical, whereas the sequence in the Morgan population differed by 1 bp mutation and presumably expresses the same S-locus specificity as plants in the other two populations. The *Lal2* haplotype from the Landersville population (race a2) is quite different from the sequence seen in the small-flowered populations, and so likely corresponds to a different S-allele. Interestingly, the *Lal2* haplotype present in Landersville is also present in the Hatton population (an outcrossing population), although we do not know if it was associated with the SC phenotype in that population because when the sequence was obtained, the plant was no longer available for phenotypic characterization. That two distinct *Lal2* sequences are associated with selfing populations having quite different floral characteristics and that these alleles are fixed in these populations together suggest that the two transitions to selfing occurred independently. These data are, however, insufficient to determine whether the *Lal2* gene harbors the mutation(s) or is in close linkage with the mutation(s) causing SC.

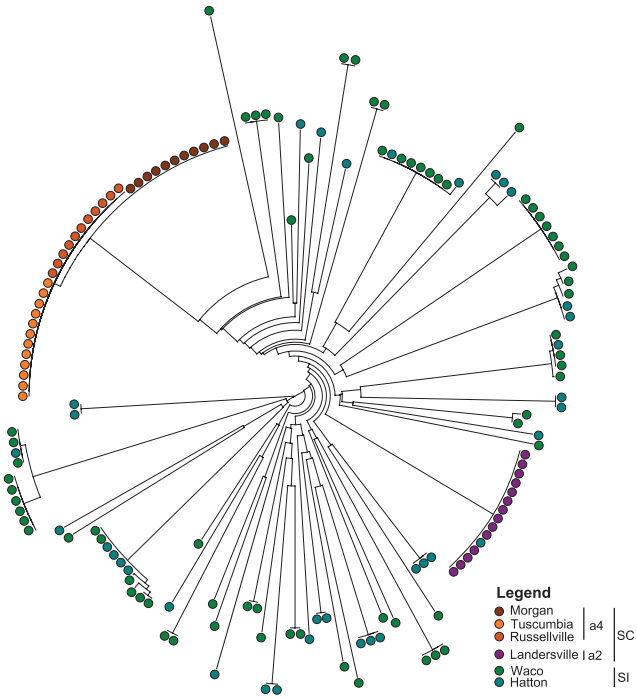

FIG. 2. Genealogy of S-linked *Lal2* sequences in *Leavenworthia alabamica*. Sequences of high similarity are assumed to share the same S-allele specificity, whereas highly divergent sequences are expected to have different S-allele specificities. The tree shows that sequences from selfing plants fall into two distinct clades. One clade includes all sequences from the a4 race, whereas the other contains all sequences from the a2 race population plus one sequence from the Hatton population.

The Origins and Genetic Basis of Selfing

The cross between the SC Tuscumbia plant (race a4) and the SI Waco plant (race a1) produced fully SC F_1 progeny. In 64 separate F_2 progeny, the SC phenotype occurred in a 3:1 ratio and cosegregated with the *Lal2-Tu* sequence (fig 3a). Plants heterozygous and homozygous for the *Lal2-Tu* sequence were fully SC. Reciprocal crosses between

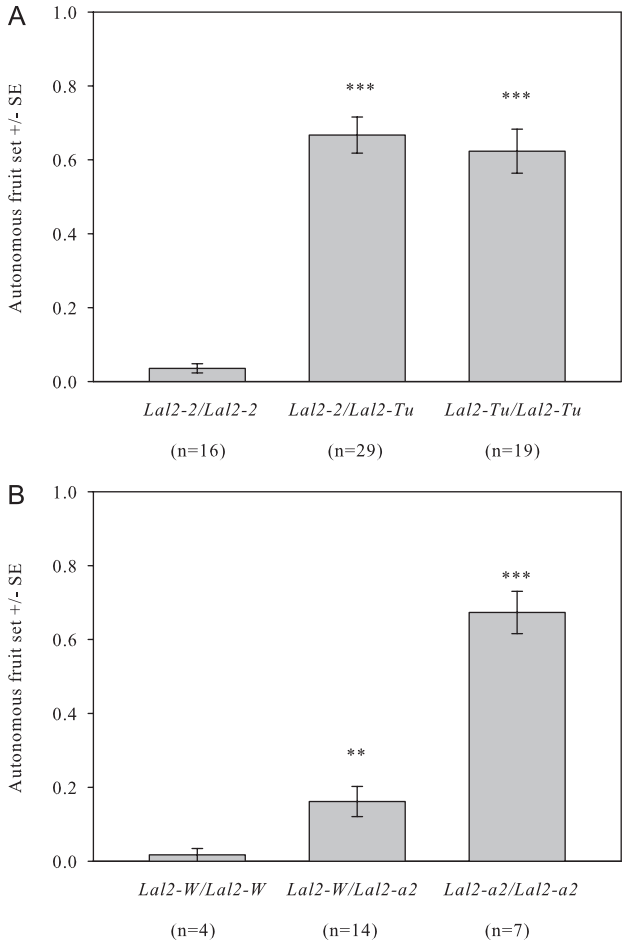

Fig. 3. Selfing is caused by S-alleles disrupting SI. (a) The proportion of flowers that become fruit is shown for S-linked *Lal2* genotypes in an F_2 generated by crossing a SI Waco and a SC Tuscumbia plant. (b) The proportion of flowers that become fruit is shown for S-linked *Lal2* genotypes in an F_2 generated by crossing a SI Waco and a SC Landersville plant. The mutation linked to *Lal2-Tu* is dominant, whereas the mutation linked to *Lal2-a2* is partially dominant. Each of these sequences cosegregates with SC. Asterisks denote *P* values associated with tests of the null hypothesis that autonomous fruit set equals 0 (**$P < 0.01$, ***$P < 0.001$).

heterozygotes and homozygotes for the functional S-allele produced contrasting results (table 1). Specifically, cross-incompatibility in heterozygotes as pistil parents, but not pollen parents, suggests that the mutation linked to the *Lal2-Tu* sequence does not suppress the SI response in

Table 1. Pollen Tube Growth in F_2 S-locus Genotypes Derived from a Cross between a SI Plant and a SC Tuscumbia Plant.

	Pollen Parent Genotype		
Pistil Genotype	***Lal2-2/Lal2-2***	***Lal2-2/Lal2-Tu***	***Lal2-Tu/Lal2-Tu***
Lal2-2/Lal2-2	4[a]/21[b]	20/21	10/10
Lal2-2/Lal2-Tu	0/11	12/14	23/23
Lal2-Tu/Lal2-Tu	14/14	15/16	22/22

[a] Numerators denote crosses where more than five pollen tubes penetrate the style.
[b] Denominators denote the total number of crosses conducted between genotypes.

Table 2. Pollen Tube Growth in F_2 S-locus Genotypes Derived from a Cross between a SI Plant and a SC Landersville Plant.

	Pollen Parent Genotype		
Pistil Genotype	***Lal2-W/Lal2-W***	***Lal2-W/Lal2-a2***	***Lal2-a2/Lal2-a2***
Lal2-W/Lal2-W	0[a]/11[b]	3/10	6/6
Lal2-W/Lal2-a2	0/18	5/24	6/6
Lal2-a2/Lal2-a2	7/7	8/8	5/5

[a] Numerators denote crosses where more than five pollen tubes penetrate the style.
[b] Denominators denote the total number of crosses conducted between genotypes.

the female component of the reaction. In contrast, the mutation suppresses the pollen component of SI.

The cross between the SC Landersville plant (race a2) and the SI Waco plant (race a1) produced partially SC F_1 progeny. In 25 separate F_2 progeny, the SC phenotype also cosegregated with the *Lal2-a2* sequence (fig 3b). Plants heterozygous for the *Lal2-a2* sequence were only partially SC and partially autonomously self-fertile, whereas plants homozygous for the *Lal2-a2* sequence were SC (fig 3b). Reciprocal crosses between heterozygotes and homozygotes for the functional S-allele produced contrasting results (table 1). Cross-incompatibility in heterozygotes as pistil parents, but not pollen parents, suggests that the mutation linked to the *Lal2-a2* sequence does not suppress the SI response in the female component of the reaction. In contrast, the mutation suppresses the pollen component of SI, although only partially in the heterozygous state.

These results suggest that the mutations that cause SC, and the breakdown of SI, are linked to the S-locus in both the naturally occurring selfing races (a2 and a4). In both cases, the mutation affects only the pollen component of the SI reaction. Moreover, the fact that two different S-alleles are fixed in the a2 and a4 races provides strong support for two independent shifts to selfing in this species.

Population Genetic Structure and Selfing Rate Estimates

Analyses in InStruct recovered four distinct population clusters. The log likelihoods of the reconstructions saturate at $K = 4$, and there is a local peak in the ΔK value. Outcrossing populations belong to one of two clusters, roughly corresponding to the western (Waco and 28 populations) and eastern (TNC7, 15, 18, and 02 populations) portions at the very center of the species range (fig. 4). The Landersville population, which is selfing and large flowered (race a2) belongs to a unique cluster, and the small-flowered selfing populations (race a4) were strongly supported as belonging to a single distinct cluster. These results are consistent with the hypothesis of multiple origins of selfing in the species. The estimate of the selfing rate using InStruct in the cluster containing the Landersville population was moderate ($s = 0.686 \pm 0.005$). This is slightly higher than a previous estimate of the multilocus selfing rate of $s = 0.523 \pm 0.074$ inferred using progeny array genotypes and mixed mating system model (Busch et al. 2010), a result that likely reflects population subdivision. A higher selfing rate was inferred

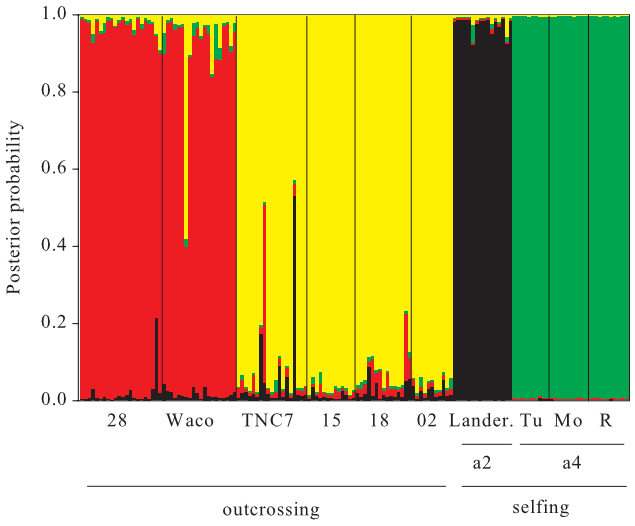

FIG. 4. Population structure inferred from InStruct analysis of microsatellites. Plot of the probability that each individual genome is derived from each of the $K = 4$ clusters. Plants from outcrossing populations are derived from one of two clusters. Plants from the partially selfing Landersville population (Lander) are derived from a single cluster, as are plants from populations of the partially selfing a4 race (Tu = Tuscumbia, Mo = Morgan, and R = Russellville).

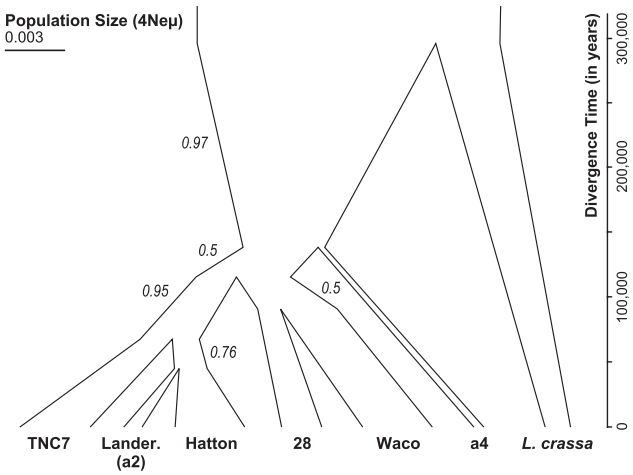

FIG. 5. Population tree obtained from eight nonrecombinant DNA regions. The tree was obtained using the Bayesian method *BEAST. Numbers denote the PPs for nodes in the population tree. The branch lengths and widths are proportional to divergence times and populations sizes, respectively.

using InStruct in the cluster containing small-flowered self-compatible plants ($s = 0.900 \pm 0.001$).

Relationships between the Selfing Populations

The three populations sampled from the a4 race were fixed for the same haplotype for all eight nuclear genes sequenced except the MtN21b gene for which the Russellville population was fixed for a haplotype different than those found in the two other populations. Because of the high similarity of these a4 populations, we treated them as a single population when reconstructing the population tree.

The branch supporting all *L. alabamica* populations is well supported (0.97 posterior probability [PP]) as well as the group of populations including TNC7, Hatton, and Landersville (a2 race) (0.95 PP; fig. 5). Although the other branches did not receive strong support, it is clear that the two selfing races are not closely related. The population tree also shows that the branches leading to the a4 and a2 races are characterized by small population sizes, with some signs of a recent increase in the case of a2. Divergence time estimates suggest that the common ancestor between the populations of the a4 race and remaining *L. alabamica* populations occurred ~150 ka ago and that the split between the Landersville (a2 race) population and the outcrossing Hatton population occurred ~48 ka ago.

Because *BEAST assumes no migration between populations, these estimates could be biased in the presence of recombination. To evaluate to what extent this might be the case, we used the software MIMAR to estimate divergence times, population sizes, and migration rates for pairs of populations involving selfing populations (fig. 6). Migration rates between the Landersville population (a2 race) and the SI populations ranged between 1.1×10^{-6} and 1.5×10^{-5}, whereas migration rates between the a4 race and the other populations were substantially smaller, ranging between 2.6×10^{-7} and 7.5×10^{-7}, suggesting that there is essentially no migration between the a4 race and the other populations. The wide posterior distributions for migration rates in some analyses involving the a4 race (e.g., TNC7 and 28) could indicate either a lack of information in the data (these populations had smaller sample sizes) or a lack of migration. The divergence times estimated by MIMAR for the Landersville population were similar to that of the population tree and the ordering was very similar (Hatton < TNC7 < Waco < pop28). The posterior distributions of the divergence times between the a4 race and the SI populations were broader, perhaps due to the little information available (race a4 was fixed for a single allele for all genes except one), but the results, although slightly larger, are not incompatible with those of *BEAST.

Past Fluctuations in Population Sizes

To investigate the ecological correlates of the shift to selfing, it is useful to infer historical fluctuations in effective population sizes. We reconstructed the evolution of past effective population sizes in selfing populations and compared it with that of outcrossing populations so as to contrast global ecological patterns that affected all populations with those that occurred only in the selfing populations. All populations surveyed except population TNC7 showed evidence of a strong bottleneck at roughly the same time in the past (fig. 7). The timing of the bottleneck closely matches the end of the last glacial period, that is, ~12.5 ka ago. Although most populations showed a strong bottleneck, the strongest apparently occurred in the Landersville population, which is selfing. Unfortunately, the lack of genetic variation in the small-flowered selfing populations (race a4) prevents the reconstruction of ancestral population sizes in this lineage.

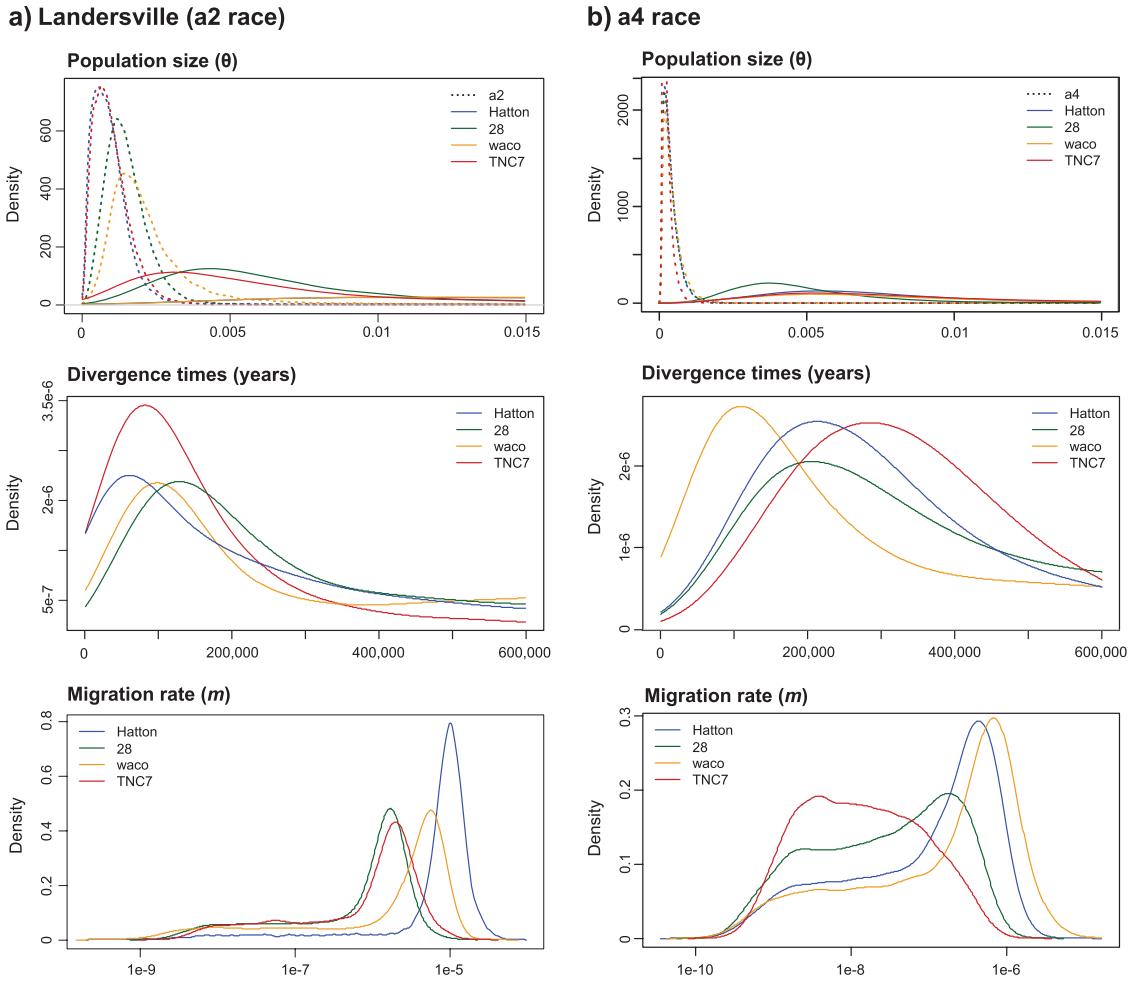

Fig. 6. Population sizes, divergence times, and migration rates estimated using a "isolation-with-migration" model. The estimates were performed with the MIMAR software for pairs of populations involving either (a) the Landersville population or (b) the a4 race. The different colors represent the results obtained from independent analyses of pairs of populations. The population sizes plots show, for each analysis (color), the estimate obtained for the selfing population (dashed line) and the outcrossing populations (solid line).

Genetic Diversity of Populations

The observed average genetic diversity was very similar across outcrossing populations, ranging from 0.020 to 0.022 (fig. 8). π_{syn} for the Landersville population (a2 race) was smaller at 0.014, whereas it was 0 for the a4 race populations. Population sizes estimated by *BEAST show a similar pattern where selfing populations had much smaller historical population sizes than outcrossing populations and ancestors (fig. 5). Interestingly, when the genetic diversity of the Landersville population was adjusted to remove the impact of inbreeding so as to be comparable with the other populations, its value was indistinguishable from those of the outcrossing populations (fig. 8). This result suggests that the lower genetic diversity observed in the Landersville population could be explained by elevated levels of inbreeding in this population; no recent bottlenecks are needed to explain this observation. These results do not seem to be affected by gene flow among populations as θ estimates with MIMAR show a similar pattern: the a4 race had a θ value close to 0, the SI populations around 0.005, and the Landersville population was intermediate (fig. 6).

Discussion

The Genetic Basis of Selfing

The loss of adaptations for outcrossing has been a major recurring evolutionary theme in plants (Holsinger 2000). It has not been clear until recently what types of mutations underlie this trend and what selective agents are responsible for it. One possibility is that mutations at the S-locus are involved in the initial evolutionary transition to selfing, as opposed to unlinked modifiers (Shimizu et al. 2008; Guo et al. 2009; Tsuchimatsu et al. 2010). It has also not been clear whether selfing rates evolve rapidly or gradually during shifts in the mating system (Igic et al. 2008). In this study, we found that there are likely two independent mutations within S-alleles that cause the suppression of a functional SI mechanism in *L. alabamica*, one in race a2 and one in race a4 populations. Although these mutations showed differing degrees of dominance, this could be a result of the different functional S-alleles with which they were combined (Busch et al. 2008).

We hypothesize that the mutations causing SC that were identified in this study lie in the *SCR* gene. Further

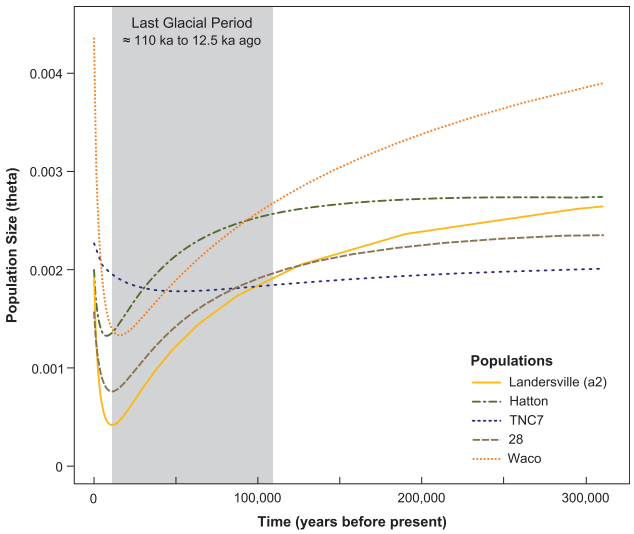

Fig. 7. EBSPs for all *Leavenworthia alabamica* populations that showed sequence variation.

support for this hypothesis could be obtained by crossing the S-allele found in the a2 race with other copies of this S-allele that are found in outcrossing populations and are therefore presumably functional; this S-allele would be expected to retain a functional pistil component of SI, given the recent evolution of selfing. The *SCR* gene encodes a pollen coat protein that triggers the SI reaction when it interacts with the protein encoded by the *SRK* gene in the stigma of the plant (Kachroo et al. 2001). Although *SCR* has yet to be identified in *L. alabamica*, it appears to be common to mustard SI systems. For both the SC mutations identified in this study, the ability of F_2 hybrid plants to autonomously self-fertilize is greatly increased by the presence of the mutant S-allele, irrespective of mutations that influence floral morphology. Indeed, the selfing rate in the a2 population has been inferred to be quite high (s_m = 0.523), even though plants in this population do not possess morphological adaptations for self-pollination (Busch et al. 2010). This result shows that the partially selfing phenotype can be caused by SC mutations, implying

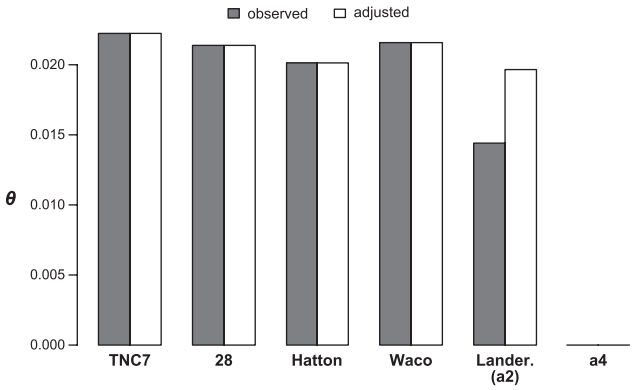

Fig. 8. Genetic diversity of populations, observed and adjusted for an inbreeding coefficient $F = 0$. Observed diversity (θ_S) was estimated from π_{syn} over all eight nuclear loci studied.

that the evolutionary shift from outcrossing to selfing can occur rapidly.

Mutations in the *SCR* gene have been implicated in the evolution of self-fertilization in *A. thaliana* and *C. rubella* (Guo et al. 2009; Tsuchimatsu et al. 2010). In *A. thaliana*, the evolution of selfing likely happened in the distant past (Tang et al. 2007), which makes it difficult to infer exactly when selfing evolved and complicates the interpretation of which mutation triggered the mating system shift. Indeed, a number of mutations disabling the SI mechanism have been identified in natural accessions (Nasrallah et al. 2004; Sherman-Broyles et al. 2007). Up to five different mutations cause selfing at the S-locus and their effects can be reversed by transformation with the *A. lyrata SRKb–SCRb* haplotype (Boggs et al. 2009), whereas other non-S-locus modifiers also weaken the SI response (Liu et al. 2007). One mutation at the *SCR* locus involves a 213 bp inversion and appears in 95% of European accessions, suggesting that it may have been the first mutation causing selfing (Tsuchimatsu et al. 2010), although this inference has been disputed given the maintenance of variation at the S-locus in this species (Tang et al. 2007). One relatively straightforward case is that of *C. rubella*, where a single S-allele bearing a *SCR* pseudogene was fixed in this lineage ~26,500 years ago, suggesting its involvement in the evolution of selfing. Interestingly, as in data presented in this paper for *L. alabamica*, crosses between *C. rubella* and its SI sister species *C. grandiflora* support the idea that self-fertility is caused by a single-dominant mutation, which suppresses the pollen component of SI (Nasrallah et al. 2007).

The fact that mutations in the *SCR* gene at the S-locus have been implicated in the evolution of selfing in *Arabidopsis*, *Capsella*, and *Leavenworthia* is perhaps not surprising. Theoretical investigations on the fate of S-alleles causing selfing have found that they have profound fertility advantages over functional S-alleles (Uyenoyama et al. 2001; Vallejo-Marin and Uyenoyama 2004; Porcher and Lande 2005). Mutant S-alleles causing selfing are expected to rapidly replace all functional S-alleles unless inbreeding depression is very strong. If inbreeding depression is strong enough, the mutant S-allele may be maintained at an intermediate frequency within populations, especially if the mutation leads to high selfing rates (Charlesworth 1988; Porcher and Lande 2005). Interestingly, mutations in the gene expressed in the pollen component of SI are expected to be more strongly selected during the evolution of selfing than mutations in the female gene. This is because they are transmitted through outcross pollen to offspring more often than other (functional) S-alleles because they evade recognition and rejection (Uyenoyama et al. 2001). Mutations in the female gene, however, will not be selected unless limited S-allele diversity causes mate limitation of seed set (Vekemans et al. 1998; Ehlers and Schierup 2008). Mutations that cause selfing at non-S-linked modifier loci cannot enjoy either of these specific fertility advantages, though they can be selected under conditions where a lack of pollinators leads to low seed set. Under conditions where pollinators are present, mutations at unlinked modifiers are therefore

likely to experience weaker natural selection and may be involved to a lesser extent in the evolution of selfing. Interestingly, dominance is also commonly observed for SC mutations, and this may be the result of a selective sieve because rare mutations may only be expressed in the heterozygous state if they are dominant (Pannell et al. 2005).

Although our results strongly suggest two independent origins of selfing, a more definitive conclusion requires the characterization of these mutations at the sequence level and tests that the SI mechanism is lost following transformation with these mutations (Nasrallah et al. 2004). This approach may be difficult for mutations located in the *SCR* gene, as this gene is highly variable and hard to amplify with PCR, given relatively few highly conserved sites. Multiple origins of selfing are consistent with the idea that transitions from SI to SC are frequent (Goldberg et al. 2010) and we predict that more examples of parallel origins of selfing will be found. If selfing generally evolves rapidly, inferences of the number of times selfing evolves using purely morphological data will likely underestimate the true number of times it has arisen. Indeed, molecular investigations have commonly identified cryptic origins of selfing (Husband and Barrett 1991; Goodwillie 1999; Foxe et al. 2010).

The Demographic Correlates of the Evolution to Selfing

The size of the demographic bottleneck associated with the evolution of selfing is of major interest because it may point to whether automatic selection or reproductive assurance favors selfing. In general, both these scenarios have been implicated by empirical studies in the mustard family. The evolution of selfing in *C. rubella* is associated with an extreme loss in genetic diversity, consistent with a bottleneck of few individuals (Foxe et al. 2009; Guo et al. 2009). It has been difficult to infer whether a bottleneck coincided with the evolution of selfing in *A. thaliana* as it occurred more than 400 ka (Bechsgaard et al. 2006; Tang et al. 2007), and there has been recent admixture that disguises historical patterns of population genetic diversity (Francois et al. 2008). In *E. paniculata* and North American *A. lyrata*, patterns of genetic diversity are consistent with moderate bottlenecks or stable populations, respectively (Foxe et al. 2010; Ness et al. 2010). In this study, the evolution of selfing in the a4 race was associated with a severe bottleneck, as there is very little diversity remaining in these populations as a whole. In contrast, the bottleneck in Landersville (race a2) was less severe, causing diversity to decline in this population by an amount consistent with the expected effects of selfing, when compared with outcrossing populations.

Regardless of why the mating system evolved in both selfing races of *L. alabamica*, it is clear that both origins were recent. Selfing in the geographically disjunct populations of race a4 most likely occurred less than 150 ka ago, which is the time when this lineage diverged from the remaining *L. alabamica* populations. Although it is difficult to identify exactly when selfing evolved in this lineage, the fact that these populations have morphological adaptations associated with self-pollination, such as introrse anthers, loss of floral scent, and reduced flower size (Busch 2005a) argues for an origin previous to that in race a2, as the mutations underlying these characteristics were likely fixed by natural selection sequentially after the fixation of the S-allele causing SC (Barrett et al. 2009). Given their floral divergence from plants in the outcrossing a1 race, Lloyd (1965) elevated the a4 race to the status of *L. alabamica* var. *brachystyla* but was not convinced that the Russellville and Tuscumbia populations shared a common origin. Based on our population structure results, the virtual absence of migration between the a4 race and the other *L. alabamica* populations, and corroborating isozyme evidence (Koelling et al. 2010), these populations should be allied as members of *L. alabamica* var. *brachystyla*. The origin of selfing in the a2 race occurred more recently, as the Landersville population appears to have diverged from the outcrossing Hatton population ~48 ka ago. The more recent timing of this event, compared with the origin of selfing in the a4 race, is supported by the fact that plants in this population have flowers with an outcrossing morphology, although they are somewhat smaller and produce less pollen per anther (Busch 2005a).

The divergence between the populations of the a4 race from outcrossing populations, estimated at 150 ka ago, is older than the start of the glacial period that took place approximately between 110 and 12.5 ka ago. This time is an upper estimate on the evolution of selfing as this event may not have coincided with the divergence of these populations. It is possible that the shift to selfing in the a4 lineage may have occurred during the glacial period, but we do not have sufficient polymorphism information to precisely date this event. The breakdown of SI and the evolution selfing in two other mustard species (North American *A. lyrata* and *C. rubella*) also appear to be recent (i.e., 10–27 ka ago) and associated with the end of the last glacial period (Foxe et al. 2009, 2010). Interestingly, both these selfing lineages appear to have profoundly different demographic histories, as genetic diversity in selfing *A. lyrata* populations is consistent with neutral expectations, whereas *C. rubella* has experienced a more than 100-fold reduction in effective population size compared with its outcrossing sister species. It is possible that reductions in the activity of pollinators during global cooling or colonization of empty habitats during the retreat of glaciers might have favored selfing in these species, as both these scenarios would select for selfing because it provides reproductive assurance. The origin of selfing in the a2 race appears to agree with these findings in other species, as it most likely occurred during the last glacial period.

The EPSPs show clear evidence of a population bottleneck in all but one population of *L. alabamica*, an event that appears to coincide with the end of the glacial period, after which all populations showed evidence of rapid growth. These population bottlenecks, apparently caused by global climate change, could well have triggered evolution toward selfing in both the a2 and a4 races. The fact

that populations of this species are restricted to cedar glades makes it difficult for plants to migrate to novel habitats in response to climate change. It has been suggested that cedar glade endemics in the southeastern United States migrated southward as temperatures cooled in the Pleistocene (Delcourt et al. 1986), although others have suggested that refugia permitted populations to persist in their present location during this cooler geologic period (Baskin and Baskin 1986). Our data support the latter hypothesis because the EBSP results suggest that some of the actual polymorphisms in the populations predate the end of the last glacial period; a result that is inconsistent with the hypothesis of a recent recolonization of these sites. Only one population (TNC7) does not show evidence of a population bottleneck. The lack of a population bottleneck in this population may be explained by its occurrence on the largest intact cedar glade in northern Alabama (Busch et al. 2010), which could have minimized the impact of the last glacial period. Migration into this population at the center of the species range might also have erased the signature of a past population bottleneck.

This study has shed some light on the likelihood of reproductive assurance causing the evolution of selfing in *L. alabamica*, although there are caveats with the approach employed in this and other studies. Declines in genetic diversity in selfing populations may reflect a bottleneck in population size that triggered the evolution of selfing, yet the evolution of selfing may also facilitate bottlenecks once it has evolved because a single selfing seed can found a new population (Baker 1955). Notwithstanding this difficulty with inferring the arrow of causality between selfing and bottlenecks, the fact that the a4 race underwent a large bottleneck in population size and is restricted to the periphery of the species range appears to collectively suggest that reproductive assurance facilitated the evolution of selfing in this lineage. The bottleneck in the Landersville population (race a2) was also substantial, but its magnitude was relatively similar to those experienced by most outcrossing populations of the species in the recent past. Given the moderate bottleneck in this lineage, it is possible that reductions in inbreeding depression within this population following this event may have contributed to the evolution of selfing (Pujol et al. 2009), as automatic selection will drive SC mutations to fixation if inbreeding depression declines below a critical value (Porcher and Lande 2005). Although automatic selection and reproductive assurance are often proposed as alternative mechanisms for the evolution of selfing, they are not mutually exclusive possibilities. Direct studies of mating system evolution should help to clarify their relative importance in the well-trod evolutionary pathway between outcrossing and selfing in plants.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Z. Wang for her laboratory work, W. Werner for microsatellite genotyping and analysis, and J. Heled for advice on *BEAST analyses. J.W.B. and S.J. were supported by McGill University Tomlinson Postdoctoral Fellowships. J.W.B. acknowledges support from the School of Biological Sciences and the College of Sciences at Washington State University. D.J.S. acknowledges support from the NSERC Discovery Grant Program and NSERC CANPOLIN Strategic Network grant.

References

Baker HG. 1955. Self-compatibility and establishment after long-distance dispersal. *Evolution* 9:347–349.

Barrett SCH. 2002. The evolution of plant sexual diversity. *Nat Rev Genet*. 3:274–284.

Barrett SCH, Ness RW, Vallejo-Marin M. 2009. Evolutionary pathways to self-fertilization in a tristylous plant species. *New Phytol*. 183:546–556.

Baskin JM, Baskin CC. 1986. Distribution and geographical/evolutionary relationship of cedar glade endemics in southeastern United States. *Assoc SE Biol Bull*. 33:138–154.

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

Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH. 2006. The transition to self-compatibility within S-haplotypes over 10 Myr. *Mol Biol Evol*. 23:1741–1750.

Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res*. 17:1505–1519.

Billiard S, Castric V, Vekemans X. 2007. A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. *Genetics* 175:1351–1369.

Boggs NA, Nasrallah JB, Nasrallah ME. 2009. Independent S-locus mutations caused self-fertility in Arabidopsis thaliana. *PLoS Genetics*. 5:e1000426.

Bruen TC, Philippe H, Bryant D. 2006. A simple robust statistical test for detecting the presence of recombination. *Genetics* 172:2665–2681.

Busch JW. 2005a. The evolution of self-compatibility in geographically peripheral populations of *Leavenworthia alabamica*. *Am J Bot*. 92:1503–1512.

Busch JW. 2005b. Inbreeding depression in self-incompatible and self-compatible populations of *Leavenworthia alabamica*. *Heredity*. 94:159–165.

Busch JW, Herlihy CR, Gunn L, Werner WJ. 2010. Mixed mating in a recently derived self-compatible population of *Leavenworthia alabamica* (Brassicaceae). *Am J Bot*. 97:1005–1013.

Busch JW, Joly S, Schoen DJ. 2010. Does mate limitation in self-incompatible species promote the evolution of selfing? The case of *Leavenworthia alabamica*. *Evolution* 64:1657–1670.

Busch JW, Sharma J, Schoen DJ. 2008. Molecular characterization of *Lal2*, an *SRK*-like gene linked to the S-locus in the wild mustard *Leavenworthia alabamica*. *Genetics* 178:2055–2067.

Charlesworth D. 1988. Evolution of homomorphic self-incompatibility. *Heredity* 60:445–453.

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

Clegg MT. 1980. Measuring plant mating systems. *Bioscience* 30:814–818.

Darwin CR. 1876. The effects of cross and self-fertilisation in the vege kingdom. London: John Murray.
Delcourt HR, Delcourt PA, Wilkins GR, Smith EN. 1986. Vegetational history of the cedar glades regions of Tennessee, Kentucky, and Missouri during the last 30,000 years. *Assoc SE Biol Bull*. 33:128–137.
Drummond AJ, Ashton B, Buxton B, et al. (12 co-authors). 2010. *Geneious, version 1.4.8*. Available from http://www.geneious.com.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acid Res*. 32: 1792–1797.
Ehlers BK, Schierup MH. 2008. When gametophytic self-incompatibility meets gynodioecy. *Genetical Res*. 90:27–35.
Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 14:2611–2620.
Fisher RA. 1941. Average excess and average effect of a gene substitution. *Ann Eugen*. 11:53–63.
Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proc Natl Acad Sci U S A*. 106:5241–5245.
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* 64:3495–3510.
Francois O, Blum MGB, Jakobsson M, Rosenberg NA. 2008. Demographic history of European populations of Arabidopsis thaliana. *PLoS Genetics*. 4:e1000075.
Gao H, Williamson S, Bustamante CD. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176:1635–1651.
Gelman A, Carlin JB, Stern HS, Rubin DB. 2009. Bayesian data analysis. 2nd ed. Boca Raton (FL): Chapman & Hall.
Glemin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc R Soc Lond B*. 273:3011–3019.
Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B. 2010. Species selection maintains self-incompatibility. *Science* 330:459–460.
Goodwillie C. 1999. Multiple origins of self-compatibility in *Linanthus* section *Leptosiphon* (Polemoniaceae): phylogenetic evidence from internal-transcribed-spacer sequence data. *Evolution* 53:1387–1395.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.
Guo YL, Bechsgaard JS, Slotte T, Neuffer B, Lascoux WD, Schierup MH. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc Natl Acad Sci U S A*. 106:5246–5251.
Hagenblad J, Bechsgaard J, Charlesworth D. 2006. Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. *Genetics* 173:1057–1073.
Heled J, Drummond AJ. 2008. Bayesian inference of population size history from multiple loci. *BMC Evol Biol*. 8:289.
Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Mol Biol Evol*. 27:570–580.
Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.
Holsinger KE. 2000. Reproductive systems and evolution in vascular plants. *Proc Natl Acad Sci U S A*. 97:7037–7042.
Husband BC, Barrett SCH. 1991. Multiple origins of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae)—inferences from style morph and isozyme variation. *J Evol Biol*. 6: 591–608.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 23(2):254–267.
Igic B, Lande R, Kohn JR. 2008. Loss of self-incompatibility and its evolutionary consequences. *Int J Plant Sci*. 169:93–104.
Jain SK. 1976. The evolution of inbreeding in plants. *Annu Rev Ecol Syst*. 24:441–466.
Joly S, Schoen DJ. Accepted. Migration rates, frequency-dependent selection and the self-incompatibility locus in *Leavenworthia* (Brassicaceae). *Evolution*.
Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB. 2001. Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. *Science* 293:1824–1826.
Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol*. 17:1483–1498.
Koelling V, Hamrick JL, Mauricio R. 2010. Genetic diversity and structure in two species of *Leavenworthia* with self-incompatible and self-compatible populations. *Heredity* doi:10.1038/hdy.2010.59.
Kondo K, Yamamoto M, Itahashi R, Sato T, Egashira H, Hattori T, Kowyama Y. 2002. Insights into the evolution of self-compatibility in *Lycopersicon* from a study of stylar factors. *Plant J*. 30:143–153.
Liu F, Zhang L, Charlesworth D. 1998. Genetic diversity in *Leavenworthia* populations with different inbreeding levels. *Proc R Soc Lond B*. 265:293–301.
Liu P, Sherman-Broyles S, Nasrallah ME, Nasrallah JB. 2007. A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Curr Biol*. 17:734–740.
Lloyd DG. 1965. Evolution of self-compatibility and racial differentiation in Leavenworthia (Cruciferae). *Contr Gray Herb Harv Univ*. 195:3–134.
Molecular Ecology Resources Primer Development Consortium et al.. 2009. Permanent genetic resources added to molecular ecology resources database 1 May 2009-31July 2009. Mol Ecol Resour. 91460–1466.
Minin VN, DormanKS FF, Suchard MA. 2005. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* 21(13):3034–3042.
Nasrallah JB, Liu P, Sherman-Broyles S, Schmidt R, Nasrallah ME. 2007. Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. *Genetics* 175:1965–1973.
Nasrallah ME, Liu P, Sherman-Broyles S, Boggs NA, Nasrallah JB. 2004. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: implications for the evolution of selfing. *Proc Natl Acad Sci U S A*. 45:16070–16074.
Ness RW, Wright SI, Barrett SCH. 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 JR, Dorken ME, Eppley SM. 2005. 'Haldane's Sieve' in a metapopulation: sifting through plant reproductive polymorphisms. *Trends Ecol Evol*. 20:374–379.
Plummer M, Best N, Cowles K, Vines K. 2010. coda: output analysis and diagnostics for MCMC. Available from http://CRAN.R-project.org/package=coda.
Pollak E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117:353–360.

Porcher E, Lande R. 2005. Loss of gametophytic self-incompatibility with evolution of inbreeding depression. *Evolution* 59:46–60.

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

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

R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from http://www.R-project.org.

Roselius K, Stephan W, Stadler T. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171:753–763.

Schoen DJ, Busch JW. 2008. On the evolution of self-fertilization in a metapopulation. *Int J Plant Sci*. 169:119–127.

Schoen DJ, Morgan MT, Bataillon T. 1996. How does self-pollination evolve? Inferences from floral ecology and molecular variation. *Phil Trans R Soc Lond B*. 351:1281–1290.

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

Shimizu KK, Shimizu-Inatsugi R, Tsuchimatsu T, Purugganan MD. 2008. Independent origins of self-compatibility in *Arabidopsis thaliana*. *Mol Ecol*. 17:704–714.

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

Stebbins GL. 1974. Flowering plants: evolution above the species level. Cambridge (MA): Belknap Press.

Takebayashi N, Morrell PL. 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, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, Hu TT, Clark RM, Nasrallah JB, Weigel D, Nordborg M. 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science* 317:1070–1072.

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

Uyenoyama MK, Holsinger KE, Waller DM. 1993. Ecological and genetic factors directing the evolution of self-fertilization. In: Futuyama DJ, Antonovics J, editors. Oxford surveys in evolutionary biology. Vol. 9. Oxford: Oxford University Press. p. 327–381.

Uyenoyama MK, Zhang Y, Newbigin E. 2001. On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics* 157:1805–1817.

Vallejo-Marin M, Uyenoyama MK. 2004. On the evolutionary costs of self-incompatibility: incomplete reproductive compensation due to pollen limitation. *Evolution* 58:1924–1935.

Vekemans X, Schierup MH, Christiansen FB. 1998. Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. *Evolution* 52:19–29.

Wright S. 1939. The distribution of self-sterility alleles in populations. *Genetics* 169:2255–2265.

Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. *Int J Plant Sci*. 169:105–118.

Wyatt R. 1988. Phylogenetic aspects of the evolution of self-pollination. In: Gottlieb LD, JainSK, editors. Plant evolutionary biology, breeding, and genetic resources. London (UK): Chapman and Hall. p. 109–131.