ORIGINAL ARTICLE

MIGRATION RATES, FREQUENCY-DEPENDENT SELECTION AND THE SELF-INCOMPATIBILITY LOCUS IN *LEAVENWORTHIA* (BRASSICACEAE)

Simon Joly[1,2,3] and Daniel J. Schoen[1]

[1]*Biology Department, McGill University, 1205 Docteur Penfield, Montreal, QC, Canada*

[2]*Institut de recherche en biologie végétale, Université de Montréal and Montreal Botanical Garden, 4101 Sherbrooke East, Montréal, QC, Canada*

[3]*E-mail: simon.joly@umontreal.ca*

Loci subject to negative frequency-dependent selection are expected to exhibit higher effective migration rates compared to reference loci. Although the number of gene copies transferred between populations by migration is the same for all genes, those subject to negative frequency-dependent selection are more likely to be retained in the immigrant population because rare alleles are selectively favored. So far, evidence for this prediction has been indirect, based on summary statistics rather than on migration rate estimates. Here, we introduce an approximate Bayesian procedure to jointly estimate migration rates at two predefined sets of loci between two populations. We applied the procedure to compare migration rate estimates at the self-incompatibility locus (S-locus) with that at 10 reference loci in two plant species, *Leavenworthia alabamica* and *L. crassa* (Brassicaceae). The maximum likelihood estimate for the proportion of migrants (m) was four times higher at the S-locus than at reference loci, but the difference was not statistically significant. Lack of significance might be due to insufficient data, but perhaps also to the recent divergence of the two species (311 ka), because we also show using simulations that the effective migration rate at the S-locus is expected to increase with increasing divergence time. These findings aid in understanding the evolutionary dynamics of negative frequency-dependent selection and they suggest that divergence time should be accounted for when employing migration rates to help detect negative frequency-dependent selection.

KEY WORDS: Balancing selection, divergence time estimates, gene flow, self-incompatibility locus.

Migration between populations may oppose or reinforce natural selection (Lenormand 2002). Indeed, the interaction between migration and selection in partially isolated populations or species is often examined with the aim of identifying loci that have played a role in adaptation and reproductive isolation (Scotti-Saintagne et al. 2004; Turner et al. 2005; Bull et al. 2006; Hey 2006; Putnam et al. 2007). At the genomic scale, this approach has been applied to scan for signatures of spatially variable directional selection under the expectation that divergence at selected loci should exceed background genomic levels (Kelley and Swanson 2008). There has been growing interest in extending the genomic scan approach to detect the signature of balancing selection (Bubb et al. 2006;

Charlesworth 2006; Andrés et al. 2009). Balancing selection is often seen as a mechanism for maintaining functional variation at loci involved in defense against diseases (Andrés et al. 2009), but its prevalence in nature is the subject of debate (Gillespie 1994; Asthana et al. 2005). Genome scans should help to clarify the importance of this type of selection in nature, but identifying such loci remains a challenge. In particular, although balancing selection may maintain genetic variation within populations and reduce divergence among populations (compared with background levels), the likelihood of detecting such signatures is influenced by a number of factors, including the ratio of the population mutation rate to population recombination rate (Nordborg and Innan 2003),

migration rates among populations (Schierup et al. 2000; Muirhead 2001), and by the strength of selection (Muirhead 2001). Consequently, using additional signatures of balancing selection could help increase the power of these searches.

Some forms of balancing selection, namely negative frequency-dependent selection, are expected to result in a higher effective migration rate at the selected locus compared to neutral genes (Schierup et al. 2000). Indeed, novel immigrant alleles are less likely to be lost by drift because their low frequencies confer upon them a selective advantage. Although migration transfers the same number of gene copies between populations, copies of genes evolving under negative frequency-dependent selection will be retained more frequently, and consequently estimated migration rates (effective migration rates) should be higher at such loci relative to neutral ones.

Clear empirical evidence for this prediction has been difficult to obtain. One expected consequence of higher effective migration is a weaker population structure. For example, major histocompatibility complex (MHC) loci are thought to evolve under negative frequency-dependent selection, yet studies in vertebrates that have employed estimates of F_{ST} to compare population structure at MHC loci with that of microsatellites have generally found little difference (reviewed in Muirhead 2001). In contrast, investigations of the self-incompatibility (SI) system of plants, the system that mediates recognition and rejection of self pollen (and for which there is strong evidence for the action of negative frequency-dependent selection; Wright 1939; Castric and Vekemans 2004), have shown that population structure is more modest at the SI locus (the S-locus) than that at neutral reference loci (e.g., Glémin et al. 2005; Ruggiero et al. 2008). Such results are compelling, but should be interpreted with caution, as other factors (e.g., mutation rate variation among the loci compared) could influence the estimate of population structure (Kronholm et al. 2010).

Recently, Castric et al. (2008) took a different approach. They estimated the parameters of the "isolation-with-migration" (IM) model (Hey and Nielsen 2004) using sequence information from several reference loci, and conducted coalescent-based simulations to calculate the expected sequence divergence at fourfold degenerate sites between sequences of the same functional alleles at the SI locus (S-locus) in *Arabidopsis halleri* and *A. lyrata*. To obtain simulated sequence divergence values for the S-locus comparable in magnitude to the observed divergence values at reference loci, Castric et al. (2008) found that the migration rate had to be increased fivefold over that estimated at the neutral loci. On the basis of this, they concluded that there is indirect evidence for increased migration rate at the S-locus compared to neutral loci. Although this approach is interesting and eliminates the potential confounding effect of varying mutation rate by relying on neutral nucleotide variation for both sets of loci, it requires as-

sumptions to be made regarding the effective population size of the S-locus.

In this study, we develop a new and explicit approximate Bayesian approach to co-estimate the migration rates, the effective population sizes, and the divergence times of two sets of loci using neutral DNA sequence variation. By integrating over probable values of divergence times and effective population sizes at both loci, the method incorporates a maximum of uncertainty in its estimates. We apply this approach to test whether the effective migration rate is higher at a locus evolving under frequency-dependent selection. The S-locus of the mustard family (Brassicaceae) was selected for this study because the evolutionary dynamics of the system are well understood (Uyenoyama 2000; Schierup et al. 2008).

Methods

STUDY SYSTEM

As part of an ongoing research project focused on the evolution and breakdown of SI, we have been investigating the population genetics of the S-locus in two closely related mustard species *Leavenworthia alabamica* and *L. crassa* (Busch et al. 2008, 2010, 2011). These winter-annual species are restricted to limestone outcrops (cedar glades) in northern Alabama (USA), and are distinguishable mainly by their fruit morphologies (Fig. 1). The species partially overlap in their geographic distribution (Lloyd 1965), produce fertile hybrids (Lloyd 1968), and hybridize spontaneously when in contact (Rollins 1963). Both species possess a sporophytic SI system, although there are also populations in both species that are self-compatible. In the Brassicaceae, the S-locus is mainly composed of two tightly linked genes involved in the SI reaction (Kusaba et al. 2001). These are the S-locus receptor kinase (*SRK*) gene, which encodes a recognition protein produced in the stigma, and the S-locus cysteine rich (*SCR*) gene, which encodes a pollen coat protein (Kachroo et al. 2001). For fertilization to occur, the parental plants must express nonmatching S-locus phenotypes in the pollen and in the stigma.

SAMPLING

Two populations of *L. alabamica* and one of *L. crassa* were sampled (Fig. 1C). This allowed the comparison of the migration rate at the S-locus against that of reference loci at two levels of divergence: between populations within species and between species. Leaves were collected from the Hatton population of *L. alabamica* (20 samples), and seeds were collected from the Waco population of *L. alabamica* (25 mothers), as well as from population 31 of *L. crassa* (20 mothers) in 2007. Additional seeds were collected in 2009 for Hatton and population 31. Seeds were germinated at 15°C and transferred to pots (1:1 of sand:promix) in
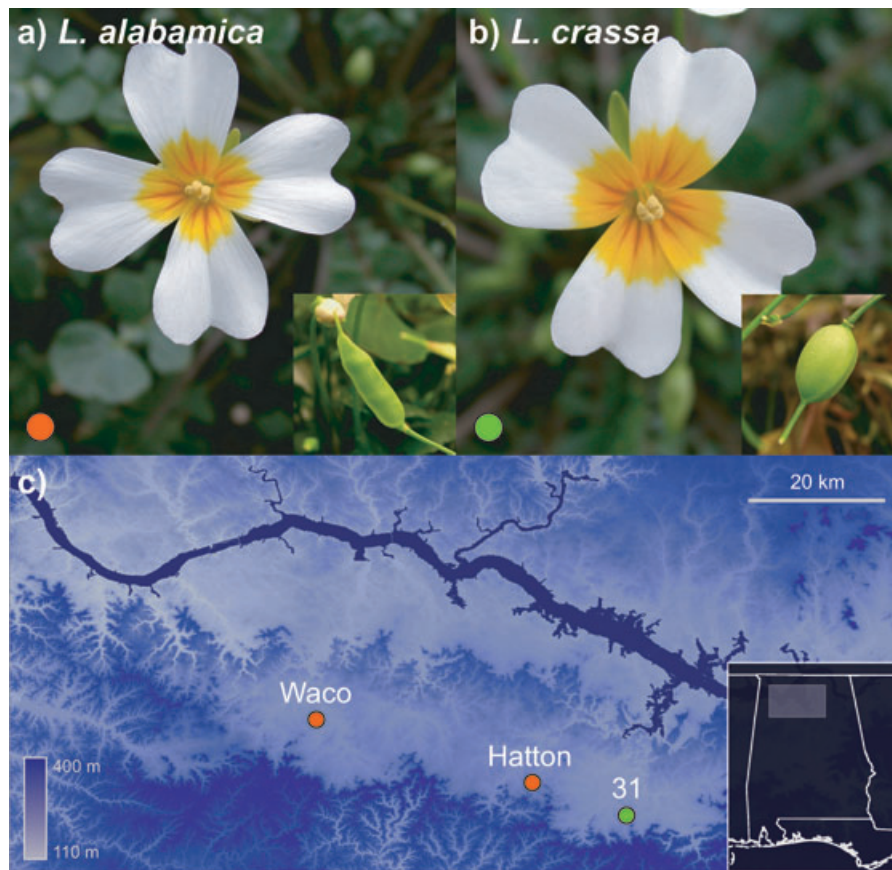
Figure 1. Flowers of (A) *L. alabamica* and (B) *L. crassa* showing diagnostic fruit characters (inset), and (C) locations of the populations sampled in this study on an elevation map. The inset on the map shows the location of the map within the state of Alabama.

greenhouses or growth chambers with supplementary lighting (22°C, 16 h day). Plants of *L. torulosa* and *L. stylosa* were grown from seed for use as outgroups. DNA was extracted from leaves using the DNeasy kit (QIAgen, Mississauga, Canada).

LOCI INVESTIGATED

Six non-S-linked nuclear loci previously used in the Brassicaceae by Joly et al. (2009a) were amplified, cloned, and sequenced using standard procedures (Joly et al. 2009a) from a few accessions to investigate variation among *L. alabamica* and *L. crassa*. Four of these were retained: malate synthase (*MS*), phosphoribulokinase (*PRK*), and two paralogs of the *MtN21* nodulation gene family (*MtN21a* and *MtN21b*). The *lal8* gene (Busch et al. 2008) was also investigated. Although *lal8* has similarities with the *ARK3* gene in *Arabidopsis thaliana*, where it is tightly linked to the S-locus, genotyping of progeny from three crosses showed that *lal8* does not segregate with SI phenotypes and thus is not linked to the S-locus in *L. alabamica* (data not shown). Primers were developed (Table S1) to amplify exon regions of 450 bp or less so that the products would be amenable for separation by single strand conformational polymorphisms (SSCP). Three *adh* genes, already optimized for SSCP analyses (Charlesworth et al. 1998),

were also investigated, for a total of eight non-S-linked loci used as reference loci.

To investigate sequence variation at the S-locus, we analyzed the gene *lal2*, the homologue of *SRK* in *Leavenworthia*. Previous studies have shown that *lal2* co-segregates with SI phenotypes in this species (Busch et al. 2008, 2011) and an ongoing *L. alabamica* genome sequencing project has generated the near-complete sequence of the *lal2* gene, which has high sequence similarity with *SRK* through the seven exons of the sequence.

SEQUENCE ANALYSES

The eight reference loci were amplified from eight to 10 individuals per population using standard conditions (Table S1). *lal2* was amplified for all available individuals as described in Busch et al. (2010). Alleles of each gene were separated by SSCP. Bands were cut out of the gel and sequenced as described in Busch et al. (2010). This process allowed us to avoid cloning as well as problems of PCR errors associated with this technique. Ambiguous allelic sequences were discarded. Because *lal2* was sometimes amplified from individuals descended from a single mother in *L. crassa* (for five maternal parents out of 25), to avoid biasing the allele frequencies in the population we retained only one sequence
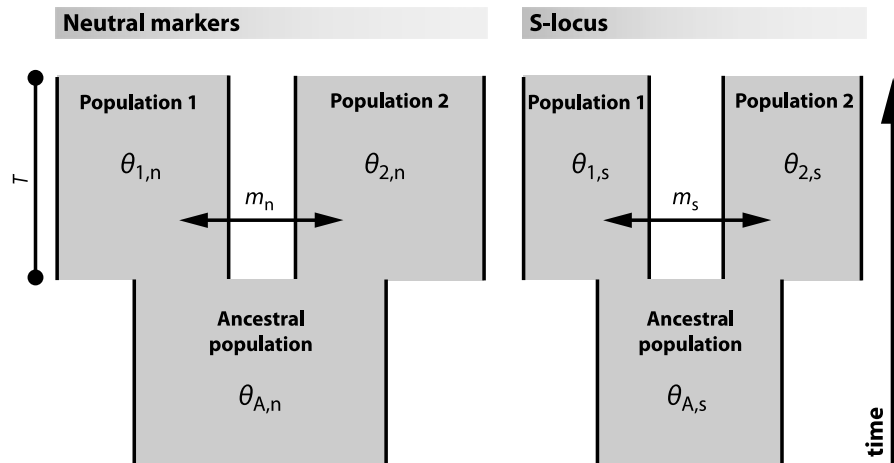
Figure 2. IMfor2 model in which two sets of loci share the same population history (i.e., divergence time [T]), but where migration rates and population sizes are independent between sets of loci.

when the same sequence was found in different individuals from the same mother. Sequences were corrected in Geneious 1.4.8 (Drummond et al. 2010) and aligned with MUSCLE (Edgar 2004). Gene genealogies were built with phyml (Guindon and Gascuel 2003) using a HKY $+ \Gamma + I$ substitution model, estimating all parameters during the search. Sequence statistics were obtained with DNAsp (Librado and Rozas 2009) from complete codons. Datasets were tested for the presence of selection using Tajima's D (Tajima 1989) and Fu and Li's D* (Fu and Li 1993), both performed in DNAsp, and the Hudson, Kreitman and Aguadé (HKA) test (Hudson et al. 1987) from synonymous substitutions for all pairs of populations, using a program written by Jody Hey (2001).

POPULATION STRUCTURE

Population structure at the S-locus, at individual S-alleles, and at the neutral genes was compared using the F_{ST} statistic. Because mutation toward new functional alleles at the S-locus is not the same as mutation for new alleles at the reference loci, we did not use an F_{ST} estimator based on allele frequencies. Instead, F_{ST} was estimated according to the formula $F_{ST} = [(\pi_1 + \pi_2)/2] / \pi_{between}$, where π_1 and π_2 are the average number of pairwise differences per site for all pairs of sequences in populations 1 and 2, respectively, and $\pi_{between}$ is the average number of pairwise differences per site among sequences of populations 1 and 2 (Hudson et al. 1992). F_{ST} statistics were estimated using only synonymous substitutions. To do this, we filtered the alignment to retain synonymous substitutions in DNAsp, and then used a custom script to estimate F_{ST} on the filtered dataset, while scaling the results according to the total number of synonymous substitutions in the sequences, as determined by DNAsp.

To add to our understanding of population structure at the S-locus, we estimated the total number of S-alleles in each population by fitting a Michaelis–Menten two-parameter model to an S-accumulation curve obtained through the Monte Carlo es-

timation of the mean number of S-alleles sampled for a given sample size (see Busch et al. 2010). The use of the Michaelis–Menten model for estimating the number of S-alleles is similar to the estimation of species richness from species accumulation curves (Colwell and Coddington 1994), and apart from the assumption that the sample is representative of the complete set of S-alleles, it does not require knowledge of the S-allele dominance relationships.

AN ISOLATION-MIGRATION MODEL FOR TWO PREDEFINED SETS OF LOCI (IMFOR2)

To compare migration rates at the S-locus and neutral loci, we expanded the standard isolation with migration (IM) model, which uses the coalescent to distinguish between migration and divergence (Nielsen and Wakeley 2001), to allow for simultaneous estimation of parameters at two predefined sets of loci. This model, henceforth called IMfor2, assumes that the two sets of loci evolved along the same population tree and split at the same time from the ancestral population. The IMfor2 model allows population sizes (θ) and migration rates (m) to be estimated independently between sets (Fig. 2).

To estimate the parameters of the IMfor2 model, we modified the software MIMAR, which uses an approximate Bayesian Markov Chain Monte Carlo approach for estimating the parameters of the IM model (Becquet and Przeworski 2007). MIMAR simulates ancestral recombination graphs (ARGs) (recombination rates could be fixed at 0 to result in bifurcating genealogies) according to the parameters of the IM model. The likelihood of the data given the model is calculated using summary statistics obtained from SNPs simulated on the ARGs. A Markov chain Monte Carlo simulation is used to estimate the posterior distribution of the parameters. Our modifications allowed the estimation of all the parameters of the IMfor2 model. Uninformative and identical priors were used for both sets of loci for all parameters to

render the posterior distributions directly comparable. The modified software, MIMARfor2, is available upon request. We considered using the software IMa2 (Hey 2010) instead of MIMAR, but we preferred MIMAR because of its capacity to handle recombination within markers and to facilitate the use of synonymous as opposed to all substitutions.

MIMARFOR2 ANALYSES

We assume that an S-allele consists of the set of functionally equivalent sequences that code for the same S-locus specificity and thus lead to pollen rejection when pollen and stigma express this specificity. If selection influences the frequencies of functional S-alleles in a population, sequences of a given S-allele are expected to be selectively neutral with respect to each other, and so their frequencies should vary due to genetic drift. Indeed, it has been shown that the genealogical process for sequences of a given functional S-allele can be treated as equivalent to one expected under selective neutrality (Vekemans and Slatkin 1994). Importantly, the effective population size (N_e) of each S-allele is expected to be smaller than that for neutral loci, as only a fraction of individuals each generation possesses each S-allele. In a system where all S-alleles are co-dominant, the effective population size of one S-allele should be roughly equal to that of the neutral alleles, divided by the number of S-alleles in the population. The fact that coalescence occurs only within S-alleles also implies that the genealogical processes at different functional S-alleles are independent. Consequently, it is possible to treat each S-allele shared between populations as an independent locus for estimating the parameters of the IMfor2 model at the S-locus, a factor that is important for obtaining good estimates of the parameters at the S-locus. This has two implications. First, genealogical independence of S-alleles implies an absence of recombination between S-alleles, although our approach allows recombination between sequences of the same S-allele. Second, using shared S-alleles as independent loci without scaling of population sizes among them implies that all S-alleles are co-dominant, which is not universally true in *L. alabamica* (Busch et al. 2008). Details of how this may affect the results are given in the discussion.

To allow a direct comparison of the S-locus and the neutral loci in the analyses, only synonymous mutations were considered. To do this, MIMAR statistics (see Tables S3 and S4) were calculated on synonymous sites using outgroup(s) sequence(s) to identify derived SNPs. Sequence length was the mean number of synonymous sites in the sequences. We tested for evidence of recombination among synonymous mutations using the Hudson and Kaplan (1985) algorithm for all loci. When there was evidence of recombination, ARGs (instead of bifurcating genealogies) were simulated using the default settings of MIMAR. In interspecific analyses, we used flat priors for θ (actual: 0.0001 to 0.05; ancestral: 0.0001 to 0.1), divergence times (10 to 5×10^6 years), and

M ($= 4N_e m$) (e^{-10} to e^2). 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. The intraspecific analyses used the same settings except for priors for θ (actual: 0.0001 to 0.1, ancestral: 0.0001 to 0.15) and the divergence time (10 to 5×10^5 years). Three chains of 5×10^6 steps were run for each analysis with samples taken every 200 steps, removing 20% as burn-in. The R package coda (Plummmer et al. 2010) was used to confirm that all parameters had reached the stationary phase and that the chains converged (Gelman's R < 1.01 [Gelman et al. 2009]).

GOODNESS-OF-FIT (GOF) TESTS FOR THE MIMARFOR2 ANALYSES

To test whether the IMfor2 model represented a good fit of the data, we performed GOF tests following Becquet and Przeworski (2007). We generated posterior predictive datasets by simulating data under the IMfor2 model using the posterior distributions obtained from MIMARfor2. The model was considered to provide a poor fit of the data when the observed statistics fell outside the 95% interval of the simulated ones. Several statistics were considered for the GOF tests: the statistics used in MIMARfor2 were used (summed over all loci), but also the average π (for each population), the mean F_{ST}, and the minimum pairwise sequence distance between populations per loci, averaged over all loci (Mindist). The minimum pairwise distance among sequences from the two populations was computed because it has been shown to be a good predictor of hybridization (Joly et al. 2009b).

TESTING THE MIMARFOR2 APPROACH USING DATA SIMULATED IN ABSENCE OF MIGRATION

To test whether negative frequency-dependent selection acting at the S-locus led to migration estimates that are artificially biased upward, we tested the MIMARfor2 program on datasets that were simulated without migration. In such a case, there should be no difference between the migration rate estimates of neutral genes and the S-locus. Backward simulation of DNA sequences using the coalescent at the S-locus is complex because two coalescent rates interact, both within and between S-alleles (Takahata 1990). To circumvent this problem, we used a two-step approach. First, we simulated S-alleles under a co-dominant sporophytic SI system in a forward simulation as described below (see Forward Simulation section). For this, we set the migration rate to zero, the mutation rate to new S-alleles to 2×10^{-7}, and the population sizes of the ancestral and sister populations to 20,000 diploid individuals. Two populations were allowed to evolve for 300,000 generations following divergence (the time since the split between *L. alabamica* and *L. crassa*; see results) from an ancestral population that was first evolved to equilibrium starting from a large (10,000) number of S-alleles. These parameters were selected so as to yield numbers of shared alleles between populations that were

similar to that observed between *L. alabamica* and *L. crassa* (i.e., between five and 10 sampled shared alleles). After 300,000 generations, we simulated empirical sampling of S-alleles by randomly sampling 100 chromosomes per population. We then counted the number of shared S-alleles between the populations and the number of times each S-allele was sampled in each population. Using this information, coalescent simulations were performed with the program *ms* (Hudson 2002) to obtain DNA sequences for each of these shared S-alleles. We also simulated DNA sequences at eight neutral loci (10 sequences in each population). For the neutral co-alescent simulations, the θ value used in *ms* was that estimated for *L. alabamica* from the data (0.0174; see Results), multiplied by the sequence length (set to 75 bp for the neutral loci, which is similar to the number of synonymous sites at empirical neutral loci). The θ used for the S-loci was estimated by dividing the θ value of the neutral genes for *L. alabamica* by 75 (to reflect the smaller effective population size of each S-allele; 75 is the equilibrium number of S-alleles per population in these forward simulations) multiplied by a sequence length of 130 bp for S-alleles. The co-alescent time since population splitting was set to 3.75 for the neutral genes (i.e., the number of generations [300,000] divided by four times the population size [20,000]). Coalescent time for S-alleles was set to 281 for the S-alleles, because the population size at each S-allele is approximately 75 times smaller than that at neutral genes. MIMAR statistics were then calculated for all these genes, and input into MIMARfor2 to estimate S-locus and neutral migration rates. MIMARfor2 settings were the same as that for the empirical data, except for the prior distribution for the migration rate parameter *M*, which was flat on a logarithmic scale from e^{-30} to e^2.

VARIATION IN S-LOCUS EFFECTIVE MIGRATION FOLLOWING POPULATION DIVERGENCE

Forward simulations were performed to investigate whether (and how) migration rates at the S-locus experiencing negative frequency-dependent selection fluctuate following population divergence. Mating was simulated according to a sporophytic SI system with co-dominant S-alleles. A maternal parent was chosen randomly from the population and a potential pollen parent was similarly picked. When the cross was compatible, gametogenesis and syngamy were simulated. When the cross was incompatible, another pollen parent was chosen, and so on until a compatible match was obtained.

We used constant mutation rates at the S-locus (μ_s; resulting in a new S-allele). Mutations within functional S-alleles and at non-S-linked loci were assumed to be neutral. The number of mutation events per generation in each population was determined by sampling from a Poisson distribution with mean $2N\mu$. The first phase of the procedure involved simulating a single ancestral population of *N* individuals that initially possessed 2*N* different

S-alleles for 50,000 generations, which was sufficient to achieve mutation–drift equilibrium. Next, two populations, each of size *N,* were formed by random sampling. After this split, two daughter populations exchanged migrants at rate *m* per generation. This second phase of the simulation ran for 200,000 generations. Migration was assumed to occur only through seeds, which is appropriate for these *Leavenworthia* species whose populations occur in a patchily-distributed habitat, that is, cedar glades isolated from one another often by many kilometers. The number of individuals that migrated from one population to the other each generation (in one direction) was determined by sampling from a Poisson distribution with a mean of *Nm*.

Effective migration rates were estimated by calculating the proportion of gene copies in one population that was present in the other population 1000 generations earlier, and were recorded separately for each locus. The population size, the S-locus mutation rate, and the migration rate were varied for different simulation runs to investigate their effects on the effective migration rates. Simulations were repeated 5000 times for each parameter setting, and mean migration rates for each time frame were recorded. Scripts for the simulations were written independently and results verified by both authors as a check against coding errors.

Results

SEQUENCE VARIATION AT REFERENCE LOCI

Sequences generated in this study were deposited in Genbank under accession numbers GU587256–GU587717. No evidence of selection was detected at the reference loci using D and D* statistics after Bonferroni correction (Table S2) as well as with the HKA test ($P > 0.05$); these loci are thus assumed to be evolving under neutrality. Nucleotide diversity per site (π) within populations was comparable to Waterson's θ (θ_W) and had a mean of 0.023 (0.002 to 0.050) for synonymous sites (Table S2). The average number of synonymous substitutions per site between populations was 0.041 between *L. alabamica* and *L. crassa* and 0.024 between the two *L. alabamica* populations (Table S3). Phylogenetic analyses showed that the species were typically not reciprocally monophyletic at all loci, although identical alleles were more often found among individuals within species (Fig. S1).

S-LOCUS SEQUENCE VARIATION

A region of 626 bp of the S-domain (exon 1) of *lal2* was sequenced for 112 and 55 chromosomes in *L. alabamica* and *L. crassa*, respectively (Table S2). One S-allele in *L. crassa* was removed from the analyses because it segregated with self-compatibility (S. Joly and D. J. Schoen, unpubl. data), which strongly suggests that this allele is nonfunctional. This conclusion is supported by the existence of other independent transitions toward selfing in

L. alabamica that have apparently occurred via S-locus mutations (Busch et al. 2011). Mean synonymous π at *lal2* was 0.165, that is, more than seven times larger than the mean value for the neutral loci (Table S2), consistent with the hypothesis that *lal2* is under negative frequency-dependent selection. Tajima's D and Fu and Li's D* statistics were generally negative although not significant for *lal2* (Table S2). These results contrast with the expectation of positive test statistic values for loci evolving under balancing selection. A possible explanation is the large number of alleles maintained at the S-locus, which differs from classic application of these tests to loci under balancing selection where few alleles are maintained in a population. The star-like genealogy in the case of an S-locus with many alleles (Fig. 3A) tends to increase the number of segregating sites relative to pairwise allelic mismatches and thereby results in a negative test statistic. The *lal2* genealogy is characterized by trans-specific polymorphisms (Fig. 3A), and exhibits a pattern of coalescence typical of the S-locus in the Brassicaceae (Takahata 1990). The distribution of sequence divergence between *lal2* sequences was clearly bimodal (Fig. 3B), reflecting the genealogy. Previous studies have shown that sequence divergence within S-alleles is generally low whereas that among S-alleles is very high (Castric et al. 2008), with different S-alleles sometimes having as little as 58% amino acid identity in region responsible for pistil SI specificity (Schierup et al. 2001). Consequently, highly similar ($\pi < 0.025$) and highly divergent ($\pi > 0.09$) sequences are henceforth assumed to belong to the same and different S-alleles, respectively (Castric et al. 2008).

POPULATION STRUCTURE

Nine S-alleles were shared among populations, and five between the two species. Estimation of S-alleles numbers in populations suggests that 28 of 54 total S-alleles were sampled from the Waco population, 24 of 46 total S-alleles were sampled from the Hatton population, and 15 of total 21 S-alleles were sampled from population 31 (Fig. 3C). Accumulation curves for S-alleles suggest that a much larger sampling effort would be required to achieve an exhaustive survey of S-allele diversity in each population. Given that all S-alleles have not been sampled, the number of observed shared alleles is likely an underestimate of the true extent of S-allele sharing.

F_{ST} estimates show that neutral genes have higher population structure than the S-locus (Table 1), as expected from theory. However, the F_{ST} of single S-alleles (different allelic variants of the same specificity) is even higher than that of neutral genes (Table 1), which likely reflects the small population size at these genealogically independent loci.

MIGRATION RATES

Parameter estimation of the IMfor2 model for the interspecific comparison was done in three independent analyses: two of these
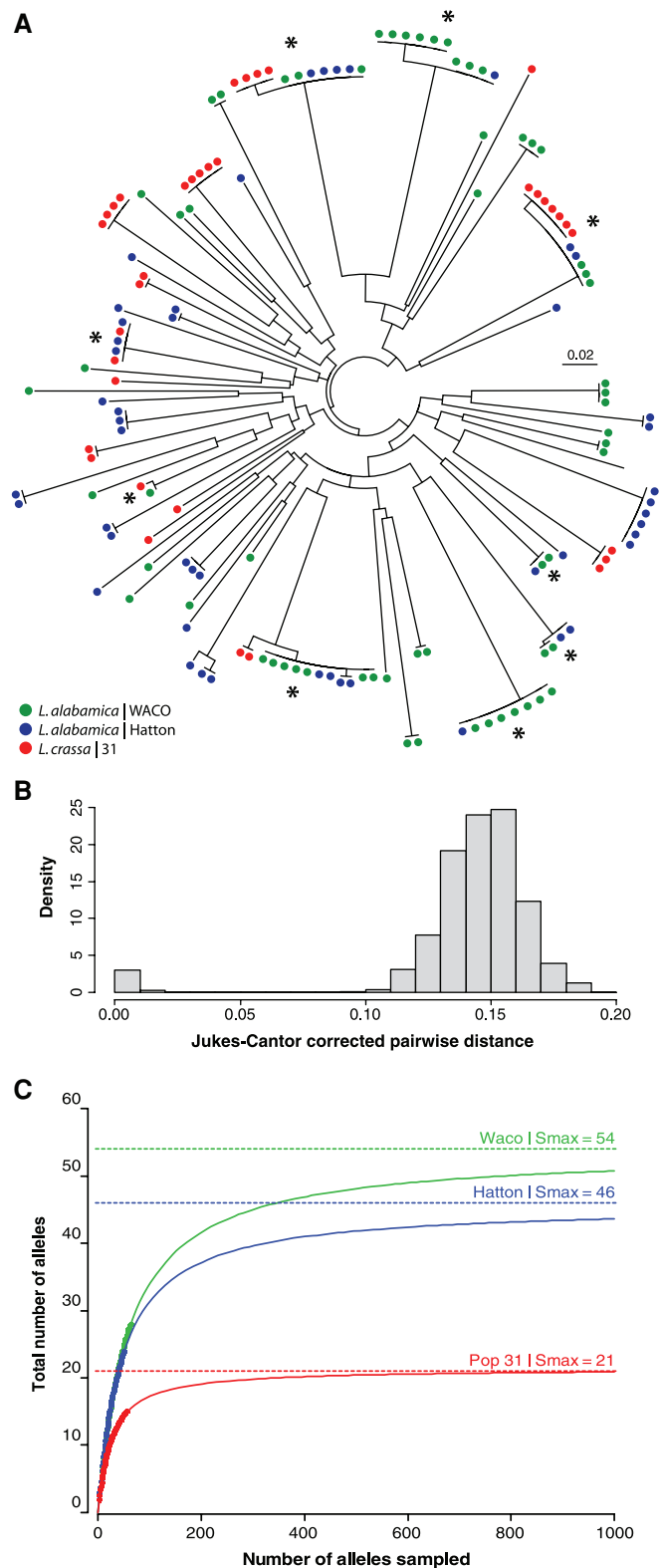

Figure 3. (A) *lal2* genealogy, where S-alleles shared among populations are denoted with asterisks, (B) average Jukes-Cantor corrected pairwise distances between all pairs of sequences, and (C) estimates of total numbers of S-allele from S-allele accumulation curves in the three populations obtained by fitting a Michaelis–Menten model on the observed data (dots).

Table 1. F_{ST} between populations for the neutral genes, the S-locus, and individual S-alleles.

	F_{ST}		
	Neutral genes[1]	S-locus	S-alleles[1]
Hatton vs. Waco	0.070208	0.01730	0.136905
crassa vs. Hatton	0.444148	0.04393	0.694444
crassa vs. Waco	0.404511	0.02924	1
crassa vs. *alabamica*	0.414634	0.02895	0.772308

[1]Mean estimates over all loci. For individual loci estimates, see Tables S3 and S4.

compared the *L. crassa* population with each of the two *L. alabamica* populations separately, and the third assumed that the Waco and Hatton populations of *L. alabamica* formed a single panmictic population. Numbers reported in the text are the mean of the two analyses using each of the *L. alabamica* populations, although the analysis with the combined *L. alabamica* populations gave similar results, suggesting that the results are robust to the assumption of panmixia within *L. alabamica* (Fig. 4). Assuming a synonymous substitution rate of 1.5×10^{-8} per generation for the Brassicaceae (Koch et al. 2000) (generation time is assumed to be of one year in these annual *Leavenworthia*), *L. alabamica* and *L. crassa* likely diverged ∼311,000 years ago (Table 2). The effective population size for the neutral loci was slightly higher for *L. alabamica* (∼300,000) than for *L. crassa* (∼200,000). Population size estimates per S-allele were about 39 times smaller than those of neutral loci (Table 2), a value that suggests the presence of 39

S-alleles in these populations in a completely co-dominant system. This is in the same range as the S-allele number estimates obtained with the Michaelis–Menten model (Fig. 3C) for the *L. alabamica* populations (46 and 54 S-alleles), but higher than that obtained for *L. crassa* (21 S-alleles). Because dominance among S-alleles is seen in some *Leavenworthia* S-alleles (Busch et al. 2008), this could explain the discrepancy.

An informative measure for the migration rate is the proportion of migrants (*m*), as it is not affected by different population sizes of the different sets of loci. Inspection of the posterior distributions of *m* for the S-locus and the neutral loci shows that the modes are distinct for the two distributions, with the S-locus showing higher migration rates (Fig. 4), although the credible intervals overlap. This lack of significance could be due to an insufficient amount of information in the data. Indeed, both distributions have long left tails, reflecting the flat prior, and suggesting that it is difficult to confidently reject small migration rates at both sets of loci. Nevertheless, the mean posterior probability over all analyses that the S-locus migration rate is higher is 78%, as estimated by the proportion of steps in the stationary phase of the Markov chain for which the S-locus migration rate was higher than that of the reference loci. Considering the maximum likelihood estimates for *m* (the mode of the posterior distribution with flat priors), the proportion of migrants per generation (*m*) was estimated to be four times higher at the S-locus than at neutral loci (Table 2 and Fig. 4).

The intraspecific analysis between the two *L. alabamica* populations suggested that the divergence between these populations occurred ∼28,800 years ago (Fig. S2). Unfortunately, the posterior distributions for the other parameters were not informative
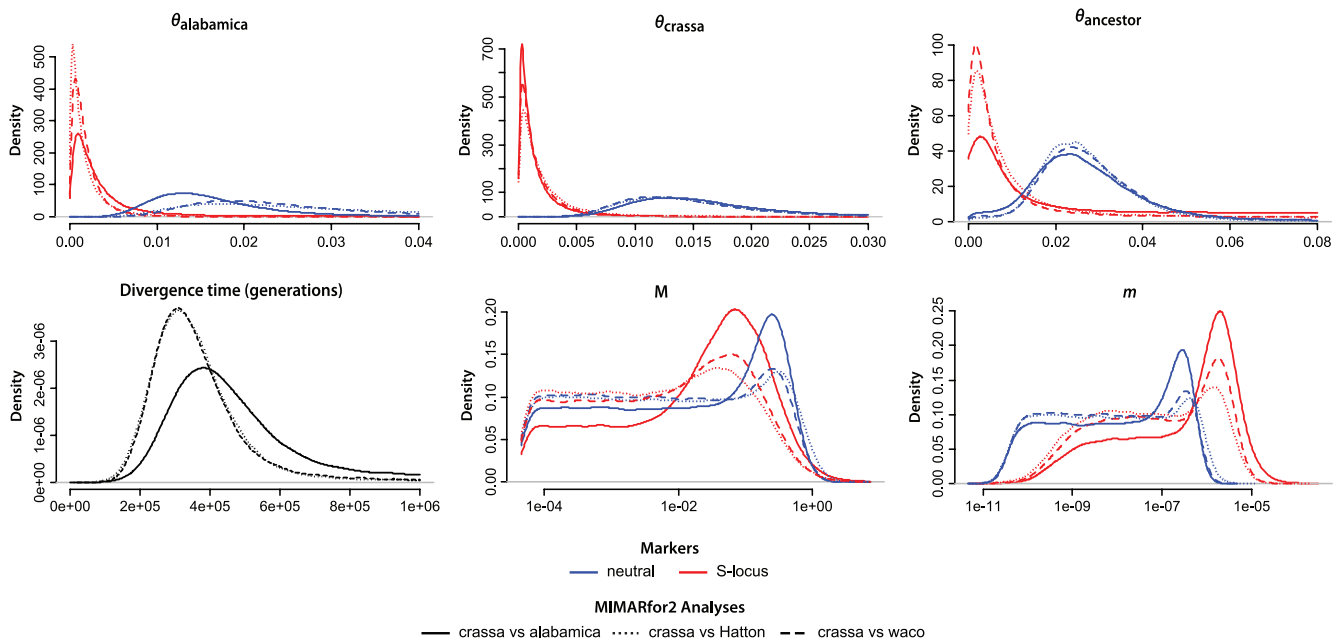

Figure 4. Posterior distributions of the parameters of the IMfor2 model (see Fig. 2) obtained from the analyses between *L. alabamica* and *L. crassa*. The parameter $M = 4N_{alabamica}m$, where $N_{alabamica}$ is the effective population size of *L. alabamica*.

Table 2. IMfor2 model maximum likelihood parameter estimates (MLE) and their 95% posterior probability intervals (PI).

Parameter	Non S-linked loci		S-linked loci	
	MLE	95% PI	MLE	95% PI
L. alabamica				
θ_{ala}	0.01747	[0.009556, 0.04632]	0.0004492	[0.0001542, 0.008429]
N_{ala}	291 186	[159 267, 772 000]	7486	[2570, 140 483]
L. crassa				
θ_{cra}	0.01212	[0.006503, 0.03103]	0.0003074	[0.0001403, 0.01212]
N_{cra}	202 065	[108 383, 517 167]	5123	[2338, 202 000]
Ancestor				
θ_A	0.02409	[0.009169, 0.05868]	0.001668	[0.0002829, 0.08875]
N_A	401 554	[152 817, 978 000]	27 797	[4715, 1 479 167]
Divergence time				
T (generations)	311 437	[183100,3527000]	311 437	[183100, 3527000]
Migration rate				
M ($4N_{ala}m$)	0.2904	[5.85×10^{-5}, 0.6489]	0.05521	[5.79×10^{-5}, 0.4223]
m	3.73×10^{-7}	[5.73×10^{-11}, 7.55×10^{-7}]	1.58×10^{-6}	[3.16×10^{-10}, 4.53×10^{-6}]

Note: These estimates were obtained by combining all chains from the analysis of (crassa vs. Waco) and (crassa vs. Hatton). Because all parameters had flat priors, the maximum likelihood estimate of a parameter corresponds to the mode of its distribution.

(Fig. S2), probably because of a lack of sufficiently informative data. Thus, it was not possible to draw conclusions regarding the S-locus migration rate at this level of divergence.

To test whether negative frequency-dependent selection acting at the S-locus leads to migration estimates that are artificially biased upward, we tested the MIMARfor2 program on datasets
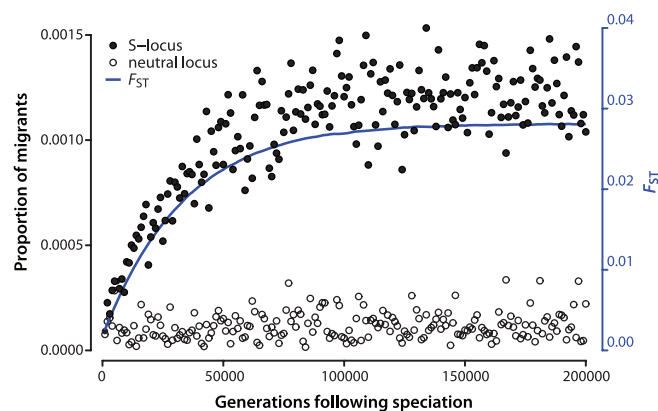

Figure 5. Results from forward computer simulations showing the proportion of migrants per population observed in time frames of 1000 generations following the speciation event. The filled circles indicate values for the S-locus and the empty circles the values for the neutral loci. The population size of the ancestral and daughter populations were set to 1000, the migration rate to 1×10^{-7}, and the mutation rates for S-linked loci to 1×10^{-6}. Each point represents the mean value from 5000 independent simulations. Mating was simulated according to a sporophytic system with all alleles co-dominant. See Methods for more details regarding the simulations.

that were simulated without migration. The posterior distributions of these analyses were several orders of magnitudes smaller than those obtained with the empirical data, were flat (uninformative), and virtually identical for the two sets of loci (Fig. S3). Thus, the approach we used does not appear to yield artificially high migration rates for the S-locus, and we conclude the higher migration rates observed with the empirical data is not a methodological artifact.

GOF tests performed from posterior distributions of all analyses showed that the IMfor2 model used provided a good fit to the data for all analyses (see Fig. S4 for the results of the *L. crassa* vs. *L. alabamica* (Waco) analysis). Statistics for the individual loci used in the GOF tests can be found in Tables S3 and S4.

SIMULATION OF EFFECTIVE MIGRATION AT THE S-LOCUS FOLLOWING DIVERGENCE

In contrast to the effective migration rate measured at neutral loci, the effective migration rate at the S-locus was not constant through time. In the first generations following the divergence, the effective migration rate at the S-locus is similar to that of the neutral locus but it increases steadily with time before eventually reaching a plateau many tens of thousands of generations later (Fig. 5). The same pattern was observed for a large range of parameter values, although the precise point at which the migration rate at the S-locus stabilizes is parameter dependent (Fig. S5). Briefly, lower migration rates, higher population sizes, and smaller mutation rates for new S-alleles reduce the pace at which the effective migration rate stabilizes. F_{ST} estimates at the S-locus throughout the simulations show that the S-locus effective migration rate is tightly

linked to population differentiation at the S-locus, suggesting that increased population differentiation should result in higher effective migration rates for loci evolving under frequency-dependent selection.

Discussion

Migration rates are not expected to be uniform across the genome. In plant SI systems, negative frequency-dependent selection should increase the effective migration rate at the S-locus compared to neutral loci and protect novel migrant S-alleles from loss due to drift. So far, there has been only indirect evidence for this prediction, as migration rates have not been directly estimated simultaneously for both types of loci. Many studies have estimated population differentiation using microsatellites for reference loci and allele frequencies or nucleotide diversity for the locus under frequency-dependent selection, which makes comparisons difficult because these markers have different mutation rates. To avoid this potential bias, we estimated F_{ST} from synonymous substitutions for both the reference loci and the S-locus, and found that the population structure at the S-locus was reduced compared with that at reference loci. Although this suggests that effective migration rates are higher at the S-locus, it is difficult to draw strong conclusions or to quantify the migration rate difference with F_{ST} estimates. According to Wright's island model, F_{ST} is related to migration using the formula $F_{ST} \approx 1/(4N_em+1)$. But Wright's island model makes several assumptions that are generally violated (Whitlock and McCauley 1999). In the present case, one most problematic aspect is that the effective population size is likely strikingly different for the reference loci compared with the S-locus. Indeed, the genealogy at the S-locus is influenced by two independent processes that occur within and among S-alleles, at slower and faster rates, respectively, compared to neutral loci (Takahata 1990). This and other violations of Wright's island model limit the use of F_{ST} in comparing migration rates at sets of loci influenced by different selection pressures.

We have presented a model that allows the co-estimation of migration rates and population sizes for two sets of loci from synonymous nucleotide substitutions. Although this approach could be used for any two sets of loci determined a priori, we have used it to co-estimate the effective migration rates at a locus evolving under frequency-dependent selection and at reference loci. We chose to use the SI system of plants for this purpose as it has the advantage of being easy to study because it consists of a long stretch of nonrecombining DNA subject to frequency-dependent selection (Charlesworth et al. 2003). In the analyses between *L. alabamica* and *L. crassa*, we observed that posterior distributions for the migration rates at the S-locus and at reference loci had distinct modes. The maximum likelihood estimate of the migration rate at the S-locus was approximately four times higher

than that at neutral loci between *L. alabamica* and *L. crassa*, a result that supports theoretical predictions. This result is quantitatively similar to the findings obtained for *A. halleri* and *A. lyrata*, where the migration rate of the S-locus was proposed to be five times higher than that of neutral loci (Castric et al. 2008). Interestingly, the average number of synonymous substitutions per site at neutral genes between *L. alabamica* and *L. crassa* (0.041) is approximately half that between *A. lyrata* and *A. halleri* estimated at 0.089 from eight loci (using data from Ramos-Onsins et al. 2004). This difference in divergence does not seem to be caused by different migration rates, as both species pairs appear to have maintained similar migration rates at neutral loci since the speciation event (3.7×10^{-7} for *Leavenworthia* vs. 2.8×10^{-7} for *Arabidopsis*). Instead, it appears to reflect differences in divergence times as the split between the two *Arabidopsis* species is estimated to have occurred approximately 2.5 million years ago (mya, Castric et al. 2008), compared to ∼0.3 mya for the two *Leavenworthia* species.

At first sight, it might be surprising to see that the difference in migration rates between the S-locus and the neutral loci is similar for the *Arabidopsis* species studied by Castric et al. (2008) and the *Leavenworthia* species studied here, because migration rates are generally thought to be higher between more recently diverged species. We suspected that this discrepancy could be caused by differences in population divergence at the S-locus, which might affect effective migration rates at the S-locus. This could occur if differentiation arising from drift and mutation at the S-locus results in increased opportunities for migration to introduce novel S-alleles. To test this hypothesis, we conducted forward simulations in a sporophytic self-incompatible two populations system with a constant gene flow, and monitored how effective migration rates varied at the S-locus and at a neutral locus as a function of the time since population divergence. The simulations confirmed that in contrast to the migration rate measured at a neutral locus, the effective migration rate at the S-locus increases steadily with time before eventually reaching a plateau. We also found that effective migration rate and population structure at the S-locus are closely associated, which supports our expectations. To our knowledge, this is the first time that a correlation between effective migration rate at the S-locus and the time since population divergence has been proposed. This phenomenon probably applies to other genes evolving under negative frequency-dependent selection, such as the MHC locus and some classes of plant resistance genes.

Although these forward simulations do not precisely replicate the history of *Leavenworthia*, they help to interpret the results. The migration rate used in the simulations was similar to that estimated from the neutral genes, but the population sizes were smaller due to constraints imposed by available computing time. Increasing the population sizes in the simulations

lengthened the point in time at which the effective migration rate at the S-locus stabilizes (Fig. S5), probably by minimizing genetic drift in populations. Moreover, a mutation rate of 1×10^{-6} as used in the simulations presented in Figure 5, likely represents an upper bound (Vekemans and Slatkin 1994), whereas smaller mutation rates slows down the rate at which the equilibrium of effective migration is achieved. Thus, it seems plausible that the effective migration rate at the S-locus between *L. alabamica* and *L. crassa*, which have been diverging for about 300,000 generations, has not yet reached its equilibrium value. This could also help account for why the relative migration rate at the S-locus versus neutral genes is slightly higher between the *Arabidopsis* species than observed here between the *Leavenworthia* species.

Although the approach used here represents a step forward toward understanding S-locus dynamics, it also makes a number of simplistic assumptions. For instance, it is assumed that all S-alleles are co-dominant, and thus that they all have the same effective size, even though this is not universally true in *Leavenworthia* (Busch et al. 2008). Recessive S-alleles are more frequent than dominant S-alleles in populations (Schierup et al. 2008; Castric et al. 2010), which means that they have higher effective population sizes, deeper genealogies, and lower effective migration rates, due to the lower probability of their being lost from the population (Schierup et al. 1997, 2000, 2008). Once additional information on S-allele dominance is obtained in *Leavenworthia*, either via allele frequencies of controlled crosses, it would be interesting to modify the model to incorporate this information. Some of the S-alleles that were found to be shared between *L. alabamica* and *L. crassa* appear to be relatively frequent (Fig. 3), suggesting they are recessive. If true, this could have biased downward the estimated migration rates at the S-locus. In general, migration rates of most recessive alleles and most dominant alleles are likely to be lower and higher, respectively, than the value reported here for the whole S-locus.

Our results are important for directing searches of loci experiencing balancing selection in an evolutionary framework. They suggest that a model that assumes constant migration following speciation does not fit the reality of genes evolving under negative frequency-dependent selection. More generally, speciation models may benefit from relaxing the assumption of fixed migration rate following speciation, as even migration rates at neutral genes could vary in function of time, for example, as may occur when reproductive isolation builds-up between two sister species.

LITERATURE CITED

Andrés, A. M., M. J. Hubisz, A. Indap, D. G. Torgerson, J. D. Degenhardt, A. R. Boyko, R. N. Gutenkunst, T. J. White, E. D. Green, C. D. Bustamante, et al. 2009. Targets of balancing selection in the human genome. Mol. Biol. Evol. 26:2755–2764.

Asthana, S., S. Schmidt, and S. Sunyaev. 2005. A limited role for balancing selection. Trends Genet. 21:30–32.

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

Bubb, K. L., D. Bovee, D. Buckley, E. Haugen, M. Kibukawa, M. Paddock, A. Palmieri, S. Subramanian, Y. Zhou, R. Kaul, et al. 2006. Scan of human genome reveals no new loci under ancient balancing selection. Genetics 173:2165–2177.

Bull, V., M. Beltran, C. Jiggins, W. O. McMillan, E. Bermingham, and J. Mallet. 2006. Polyphyly and gene flow between non-sibling *Heliconius* species. BMC Biol. 4:11.

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

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

Busch, J. W., S. Joly, and D. J. Schoen. 2011. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. Mol. Biol. Evol. doi:10.1093/molbev/msq352 [Epub ahead of print].

Castric, V., and X. Vekemans. 2004. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. Mol Ecol. 13:2873–2889.

Castric, V., J. Bechsgaard, M. H. Schierup, and X. Vekemans. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genet. 4:e1000168.

Castric, V., J. S. Bechsgaard, S. Grenier, R. Noureddine, M. H. Schierup, and X. Vekemans. 2010. Molecular evolution within and between self-incompatibility specificities. Mol. Biol. Evol. 27:11–20.

Charlesworth, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2:e64.

Charlesworth, D., F. Liu, and L. Zhang. 1998. The evolution of the alcohol dehydrogenase gene family by loss of introns in plants of the genus *Leavenworthia* (Brassicaceae). Mol. Biol. Evol. 15:552–559.

Charlesworth, D., C. Bartolome, M. H. Schierup, and B. K. Mable. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. Mol. Biol. Evol. 20:1741–1753.

Colwell, R.K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philos. Trans. R. Soc. Lond. B 345:101–118.

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, et al. 2010. Geneious. Available at: http://www.geneious.com (accessed March 4, 2010).

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2009. Bayesian data analysis 2 ed. Chapman & Hall, Boca Raton, FL.

Gillespie, J. H. 1994. The causes of molecular evolution. Oxford Univ. Press, New York.

Glémin, S., T. Gaude, M. Guillemin, M. Lourmas, I. Olivieri, and A. Mignot. 2005. Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. Genetics 171:279–289.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Hey, J. 2001. HKA: a computer program for tests of natural selection. Rutgers Univ., New Brunswick, NJ. Available at http://genfaculty.rutgers.edu/hey/software (accessed January 26, 2011).

———. 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27:905–920.

———. 2006. Recent advances in assessing gene flow between diverging populations and species. Curr. Opin. Genet. Dev. 16:592–596.

Hey, J., and R. Nielsen. 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.

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164.

Hudson, R. R., M. Kreitman, and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Hudson, R. R., M. Slatkin, and W. P. Manning. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589.

Joly, S., P. B. Heenan, and P. J. Lockhart. 2009a. A pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. Mol. Phylogenet. Evol. 51:365–372.

Joly, S., P. A. McLenachan, and P. J. Lockhart. 2009b. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174:e54–e70.

Kachroo, A., C. R. Schopfer, M. E. Nasrallah, and J. B. Nasrallah. 2001. Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 293:1824–1826.

Kelley, J. L., and W. J. Swanson. 2008. Positive selection in the human genome: from genome scans to biological significance. Annu. Rev. Genom. Human Genet. 9:143–160.

Koch, M. A., B. Haubold, and T. Mitchell-Olds. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol. Biol. Evol. 17:1483–1498.

Kronholm, I., O. Loudet, and J. de Meaux. 2010. Influence of mutation rate on estimators of genetic differentiation—lessons from *Arabidopsis thaliana*. BMC Genet. 11:33.

Kusaba, M., K. Dwyer, J. Hendershot, J. Vrebalov, J. B. Nasrallah, and M. E. Nasrallah. 2001. Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. Plant Cell. 13:627–643.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 17:183–189.

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

Lloyd, D. G. 1965. Evolution of self-compatibility and racial differentiation in *Leavenworthia* (Cruciferae). Contrib. Gray Herb Harv. 195:3–134.

———. 1968. Partial unilateral incompatibility in *Leavenworthia* (Cruciferae). Evolution 22:382–393.

Muirhead, C. A. 2001. Consequences of population structure on genes under balancing selection. Evolution 55:1532–1541.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. Genetics 158:885–896.

Nordborg, M., and H. Innan. 2003. The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. Genetics 163:1201–1213.

Plummmer, M., N. Best, K. Cowles, and K. Vines. 2010. Coda: output analysis and diagnostics for MCMC. Available at http://CRAN.R-project.org/package=coda (accessed December 9, 2010).

Putnam, A. S., J. M. Scriber, and P. Andolfatto. 2007. Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. Evolution 61:912–927.

Ramos-Onsins, S. E., B. E. Stranger, T. Mitchell-Olds, and M. Aguade. 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics 166:373–388.

Rollins, R. C. 1963. The evolution and systematics of *Leavenworthia* (Cruciferae). Contrib. Gray Herb Harv. 192:3–98.

Ruggiero, M. V., J. Jacquemin, V. Castric, and X. Vekemans. 2008. Hitch-hiking to a locus under balancing selection: high sequence diversity and low population subdivision at the S-locus genomic region in *Arabidopsis halleri*. Genet Res. 90:37–46.

Schierup, M. H., X. Vekemans, and F. B. Christiansen. 1997. Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. Genetics 147:835–846.

Schierup, M. H., X. Vekemans, and D. Charlesworth. 2000. The effect of subdivision on variation at multi-allelic loci under balancing selection. Genet Res. 76:51–62.

Schierup, M. H., B. K. Mable, P. Awadalla, and D. Charlesworth. 2001. Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. Genetics 158:387–399.

Schierup, M. H., J. S. Bechsgaard, and F. B. Christiansen. 2008. Selection at work in self-incompatible *Arabidopsis Lyrata* II. Spatial distribution of s-haplotypes in Iceland. Genetics 180:1051–1059.

Scotti-Saintagne, C., S. Mariette, I. Porth, P. G. Goicoechea, T. Barreneche, C. Bodenes, K. Burg, and A. Kremer. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. Genetics 168:1615–1626.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Takahata, N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA. 87:2419–2423.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:e285.

Uyenoyama, M. K. 2000. Evolutionary dynamics of self-incompatibility alleles in *Brassica*. Genetics 156:351–359.

Vekemans, X., and M. Slatkin. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. Genetics 137:1157–1165.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{ST} \approx 1/(4Nm+1)$. Heredity 82:117–125.

Wright, S. 1939. The distribution of self-sterility alleles in populations. Genetics 24:538–552.

Associate Editor: J. Pannell

Supporting Information

The following supporting information is available for this article:

Figure S1. Neutral gene genealogies of *Leavenworthia alabamica* (populations Waco and Hatton) and *L. crassa* (population 31).
Figure S2. Posterior probability distributions of the parameters of the IMfor2model obtained from the analysis of the two *L. alabamica* populations (Hatton and Waco).
Figure S3. Posterior probability distributions of the number of migrant per generation (*M*; in one direction) and the symmetrical migration rate (*m*) for 10 datasets simulated without gene flow.
Figure S4. Results from the goodness-of-fit tests performed from the posterior distribution of the MIMARfor2 analysis between *L. alabamica* (pop Waco) and *L. crassa*.
Figure S5. Results from forward simulations showing the effect of (A, B) the symmetric migration rate, (C) the population size, and (D) the mutation rate for new S-alleles on the proportion of migrants per population observed in time frames of 1000 generations following the speciation event.
Table S1. Primers and PCR conditions for the loci used in this study
Table S2. Sequence characteristics of the nine loci studied.
Table S3. Statistics specific to the neutral loci used in the IMfor2 analyses.
Table S4. Statistics specific to the S-locus used in the IMfor2 analyses.

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.