The Effect of Mating System Differences on Nucleotide Diversity at the Phosphoglucose Isomerase Locus in the Plant Genus Leavenworthia

F. Liu,* D. Charlesworth† and M. Kreitman*

**Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637 and †Ashworth Laboratory, Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom*

Manuscript received May 13, 1998
Accepted for publication October 12, 1998

ABSTRACT

To test the theoretical prediction that highly inbreeding populations should have low neutral genetic diversity relative to closely related outcrossing populations, we sequenced portions of the cytosolic phosphoglucose isomerase (*PgiC*) gene in the plant genus Leavenworthia, which includes both self-incompatible and inbreeding taxa. On the basis of sequences of intron 12 of this gene, the expected low diversity was seen in both populations of the selfers *Leavenworthia uniflora* and *L. torulosa* and in three highly inbreeding populations of *L. crassa*, while high diversity was found in self-incompatible *L. stylosa*, and moderate diversity in *L. crassa* populations with partial or complete self-incompatibility. In *L. stylosa*, the nucleotide diversity was strongly structured into three haplotypic classes, differing by several insertion/deletion sequences, with linkage disequilibrium between sequences of the three types in intron 12, but not in the adjacent regions. Differences between the three kinds of haplotypes are larger than between sequences of this gene region from different species. The haplotype divergence suggests the presence of a balanced polymorphism at this locus, possibly predating the split between *L. stylosa* and its two inbreeding sister taxa, *L. uniflora* and *L. torulosa*. It is therefore difficult to distinguish between different potential causes of the much lower sequence diversity at this locus in inbreeding than outcrossing populations. Selective sweeps during the evolution of these populations are possible, or background selection, or merely loss of a balanced polymorphism maintained by overdominance in the populations that evolved high selfing rates.

SEVERAL factors are predicted to lead to low genetic diversity in highly inbreeding populations. Such populations have increased frequencies of homozygotes, resulting in reduced effective population size [complete inbreeding leads to a halving of the effective population size (Pollak 1987)] and lowered effective rates of genetic recombination. Reduced recombination is associated with (1) an increased effect of adaptive gene substitutions on neutral variability at linked sites (hitchhiking or selective sweeps; see Hedrick 1980) and (2) an increased effect of selection against deleterious alleles on neutral variation at linked sites (background selection). These processes both tend to reduce neutral genetic diversity (reviewed in Charlesworth *et al.* 1993). Bottlenecks may also be more extreme in inbreeders, in which a single seed can found a new population, than in outcrossing species. Some evidence suggests greater variability in effective population size in inbreeders than outbreeders (Schoen and Brown 1991), which might suggest that bottlenecks have been important, although other explanations for the findings are possible (Charlesworth *et al.* 1997). Finally, polymorphisms maintained by heterozygote advantage will tend to be lost in populations that are highly inbreeding (Kimura and Ohta 1971; Charlesworth and Charlesworth 1995).

Partial selfing in plants is indeed correlated with reduced within-population allozyme variability (Brown 1979; Hamrick and Godt 1990, 1996; Schoen and Brown 1991). The data indicate that allozyme diversity in selfing plant populations is ~50% of that of obligate outcrossers (Hamrick and Godt 1990; Schoen and Brown 1991). This effect is, however, much smaller than that expected if all the factors described above are taken into account (reviewed in Charlesworth *et al.* 1993), which suggests that allozyme diversity may be selectively maintained.

The aim of the work described here is to compare sequence polymorphism at the DNA level in a phosphoglucose isomerase gene between species with different outcrossing rates. Phosphoglucose isomerase (PGI; E.C. 5.3.1.9) catalyzes the reversible isomerization of glucose-6-phosphate and fructose-6-phosphate in the glycolytic pathway. Plants have at least two phosphoglucose isomerase genes, the cytosolic *PgiC* and a plastid-expressed locus that is so different in sequence that neither PCR-based methods nor Southern blotting have yielded clones from any plant species (Ford *et al.* 1995). Both loci are nuclear. Generally, plant species have a single cytosolic *PgiC* locus (Gottlieb 1982; Terauchi *et al.* 1997), but some species of Clarkia have been found to

Corresponding author: D. Charlesworth, Ashworth Lab, Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Bldgs., W. Mains Rd., Edinburgh EH9 3JT, United Kingdom.
E-mail: deborah.charlesworth@ed.ac.uk

have two, the result of a gene duplication that appears to have originated within the genus (Gottlieb and Weeden 1979; Ford *et al.* 1995; Gottlieb and Ford 1996). Isozyme electrophoresis has shown that *PgiC* is highly polymorphic in many plant species (Gottlieb and Greve 1981; Terauchi *et al.* 1997) including the genus studied here, Leavenworthia (Charlesworth and Yang 1998), and balancing selection has been invoked to explain the maintenance of the polymorphisms (Gillespie 1991; Terauchi *et al.* 1997). In two plant species, *Clarkia lewisii* (two alleles only) and *Dioscorea tokoro*, both allozyme and DNA polymorphism data have been compared; in both species, low levels of DNA diversity were found at silent sites and in intron regions in the *PgiC* locus (Thomas *et al.* 1993; Terauchi *et al.* 1997). We here describe estimates of sequence diversity within and between members of a group of species in the genus Leavenworthia, a classic example of breeding system evolution (Rollins 1963; Lloyd 1965). Intron regions were chosen for study both because their sequence variability might be expected to behave neutrally and would thus be most suitable for tests of the theory, and because we expected both replacement and silent diversity to be low in exons, on the basis of the findings just mentioned and those in *Drosophila melanogaster* (J. H. McDonald and M. Kreitman, unpublished results cited in Moriyama and Powell 1996; Terauchi *et al.* 1997).

MATERIALS AND METHODS

The genus Leavenworthia: Leavenworthia is a small genus of eight diploid annual species in section Arabidae of the Brassicaceae. The taxonomy of this family is not yet well worked out (Price *et al.* 1994), and the closest relatives of this genus are not certain, though it is thought to be closely related to Cardamine (Rollins 1963). A taxonomy of the genus supports the view that the species fall into three groups according to their chromosome numbers (Christiansen 1993).

In the Leavenworthia species with $n = 15$, selfing appears to have evolved twice, very recently in the case of *L. torulosa* (see below). This is in addition to the independent origins of selfing in the $n = 11$ species, *L. crassa* and *L. alabamica* (Rollins 1963; Lloyd 1965). This is consistent with findings from other genera, in which evolution of selfing also appears to be a fairly frequent occurrence, and selfing taxa appear often to be of recent origin (Stebbins 1957; Wyatt *et al.* 1992; Charlesworth *et al.* 1993; Barrett *et al.* 1996; Schoen *et al.* 1997). The multiple independent evolutionary losses of self-incompatibility in Leavenworthia give us the opportunity to test whether the evolution of inbreeding shows a repeatable tendency to lead to loss of genetic variability. To compare across the greatest possible contrast in mating systems, but preserve similarity in other characters, we estimated the effect of breeding system in sets of populations in two of the three chromosome number groups in the genus. The first set includes *L. stylosa* (self-incompatible, fully outbreeding) and its highly selfing relatives *L. uniflora* and *L. torulosa*, and the second consists of populations of *L. crassa*, whose selfing rates range from very low to close to 100% (some populations are self-incompatible, some are intermediate in their selfing rate and polymorphic for self-incompatibility, while others are highly self-compatible; see Lloyd 1965). All these species are reproductively isolated from one another (Rollins 1963).

Population samples: Population samples were grown from seeds collected in the field or supplied by L. L. Lyons, G. Hilton, and T. E. Hemmerley, from four populations of *L. stylosa* (Gray), two populations of *L. uniflora* (Michx.) Britton, one of *L. torulosa* (Gray), and seven populations of *L. crassa* (Rollins). Table 1 summarizes the populations studied here, which are described in more detail in Liu *et al.* (1998) and Charlesworth and Yang (1998). Note that population 95008 was previously thought to be *L. uniflora*, but now appears to be *L. torulosa* (see below). Population Hem 2 (not used in the previous work) was collected by Dr. Hemmerley at Cedar Forest, Tennessee. For populations for which breeding system information was not already available, selfing rates were estimated by measuring autogamous fruit set in the greenhouse and from hand pollinations to test self-compatibility (Charlesworth and Yang 1998).

Allozyme electrophoresis and studies of the inheritance of *PgiC* variants: Allozyme genotypes were determined by cellulose-acetate electrophoresis (Hebert and Beaton 1989) using Tris-glycine buffer as described by Charlesworth and Yang (1998), who tested single-locus inheritance of the *PgiC* variants by raising families from crosses between plants of known allozyme genotypes. Inheritance studies were also performed at the DNA level, to test segregation of putative heterozygotes (see below). Note that the populations studied by allozyme methods are fewer in number than those for which sequences were obtained, because seeds from some populations did not germinate; we could nevertheless extract DNA from seeds from these populations.

Molecular methods: *Cloning and sequencing of PgiC cDNA from L. crassa:* To study the Leavenworthia cytosolic phosphoglucose isomerases (*PgiC*), sequences from *Arabidopsis thaliana* (accession no. X69195) and *C. lewisii* (accession no. X64332) were used to design degenerate and nondegenerate primers. Total RNA was isolated from *L. crassa* and *A. thaliana* leaves using the acid guanidium thiocyanate-phenol-chloroform extraction method (Chomczynski and Sacchi 1987). The StrataScript (Stratagene, La Jolla, CA) reverse transcription (RT)-PCR kit was used to synthesize first-strand cDNA from total RNA. A 745-bp fragment of the gene from between exon 11 and exon 21 was amplified from cDNA with the "plus" primer S2 (5′ TTTGCATTTTGGGACTGGGT 3′) and the "minus" primer R1 [5′ AC(A,T,C,G)CCCCA(C,T)TG(A,G)TC(A,G)AA 3′]. PCR amplifications were carried out using 2.0 mm [Mg^{2+}]. Reaction conditions were 2 min at 95°, followed by 30 cycles of 15 sec at 94°, 30 sec at the annealing temperature (set at $T_m - 2°$, where T_m is the melting temperature, determined from the A + T/G + C content by $T_m = 4x[G + C] + 2x[A + T]$), and 2 min at 72°. One 745-bp band was seen in both *L. crassa* and *A. thaliana*. The products were cloned using the Original TA cloning kit (Invitrogen, San Diego). Plasmids from single colonies were prepared as templates for cycle sequencing using the modified mini alkaline-lysis/PEG precipitation procedure (P/N 901497; ABI, Columbia, MD). Sequencing reactions were performed using 1 μg of template plasmid, 50 ng of sequencing primer (universal primers of the vector: M13 reverse primer or M13 −20 primer), and 9.5 μl of fluorescent dideoxy terminator mix per reaction. The cycle sequencing procedure consisted of 25 cycles each of 15 sec at 95°, 30 sec at 56°, and 4 min at 60°. Sequences were analyzed on an ABI 373A sequencer.

To obtain sequences 3′ and 5′ to those obtained with the primer pair S2 and R1, internal *L. crassa*-specific primers were designed for 5′ and 3′ rapid amplification of cDNA ends (RACE; Life Technologies). The 5′ RACE system was used to obtain clones of the 5′ end of the *L. crassa PgiC* locus, and the

TABLE 1

Breeding systems of plants from the Leavenworthia populations studied

Species	Population numbers	Breeding system	Numbers of alleles studied	
			Sequence	SSCP only
L. stylosa	95007	Self-incompatible	5	Not done
	9113[a]	Self-incompatible	8	Not done
	Hem1[b]	Self-incompatible	8	Not done
	Hem2[b]	Self-incompatible	5	Not done
L. uniflora	95011	Self-compatible	3	2
	9108[a]	Self-compatible	4	2
L. torulosa	95008	Self-compatible	3	2
L. crassa	95003	Self-compatible	3	2
	95004	Self-compatible	4	1
	9107[a]	Self-compatible	3	3
	95005	Intermediate	5	2
	95010	Intermediate selfing	4	2
	8921[c]	Intermediate selfing	3	1
	8919[a]	Self-incompatible	7	0
	95014	Self-incompatible	6	1

[a] Seeds supplied by Dr. E. E. Lyons.
[b] Seeds supplied by Dr. T. E. Hemmerly.
[c] Seeds supplied by Dr. G. Hilton.

3′ end was obtained by amplifying with poly(T)18(A,C,G)*N* (where *N* can be A, C, G, or T, *i.e.*, four different primers) as the anchor primer. The amplified products were cloned and sequenced using the methods described above. For sequencing, direct PCR amplifications from white colonies using the pair of universal primers (M13 Reverse and M13 −20) were also performed, as described above. The products were then purified for cycle sequencing using the QIAquick-spin PCR purification kit (QIAGEN, Chatsworth, CA). Using these methods, the complete *L. crassa PgiC* cDNA sequence was obtained (GenBank accession number AF054455).

PCR amplification from genomic DNA and single-strand conformation polymorphism analysis: Using the *L. crassa PgiC* cDNA sequence, internal primers were designed for amplification from genomic DNA. Genomic DNA was prepared from leaves of individual plants by a modified CTAB plant miniprep method, or from seeds using a modified Puregene DNA isolation protocol (Gentra Systems, Research Triangle Park, NC). The modification consisted of adding two chloroform extractions of the lysates after protein precipitation, which helped to remove enzyme-inhibiting contaminants in the seeds (Murray and Thompson 1980).

For polymorphism analysis, we amplified a small genomic DNA fragment (270–320 bp) corresponding to the region between exon 12 and exon 13 of the *A. thaliana PgiC* gene, using primers PgiC.P1 5′ AGTATGGCTTCTCCATGGTT 3′ and PgiC.P2R 5′ ATGTGGACTTGAAATGCTG 3′. We refer to this in what follows as the intron 12 region. To obtain *PgiC* sequences from regions between exons 11 and 14 of the *PgiC* gene, the plus primer S2 and the minus primer PgiC.P3R (5′ TCCATACACTCAACAATCCTA 3′) were used. The fragments amplified from individual plants were sequenced and/or subjected to single-strand conformation polymorphism analysis (SSCP), using the method of "cold SSCP" (Hongyo *et al.* 1993), which is expected to be capable of detecting single differences in PCR products up to ~350 bases. Figure 1 shows some results for some intermediate selfing and outcrossing *L. crassa* populations and some highly selfing populations of three different species. Heterozygotes show three- or four-banded patterns and can thus be distinguished from homozygotes, which always show two-banded patterns. Sequences of alleles identified by their SSCP conformations were obtained by direct sequencing of both strands. In the case of heterozygotes, the PCR product was cloned, and five to eight clones were sequenced. There may therefore be a small proportion of errors in the sequences from these individuals, but these should be minor and should not affect our results overall. The GenBank accession numbers of the sequences are AF054456–AF054484 and AF054486–AF054495.

Sequence analyses: The numbers of alleles studied for each population are listed in Table 1. Nucleotide diversities in *L. crassa*, *L. uniflora*, and *L. torulosa*, in which lower diversity was seen, on the basis of initial sequence data, were estimated by a combination of SSCP analysis and direct sequencing. Two or more alleles of each SSCP phenotype were sequenced from several individuals of each population, either for the smaller (intron 12) region, or for a longer fragment of the gene, including introns 11–13. Complete sequence identity was found between 2 to 10 alleles from each of 10 different SSCP phenotypes (Table 1). We therefore used SSCP analysis to estimate the number of alleles of each SSCP phenotype, together with direct sequencing of alleles of each type. This will, at most, slightly underestimate diversity in the most variable populations (which is conservative for our estimates of the differences between inbreeding and outcrossing populations).

ClustalW was used to align the intron sequences, followed by manual adjustment to further reduce the number of substitutions or insertions and deletions. After removing the primer sequences, numbers of pairwise differences between sequences (*i.e.*, per base estimates of silent nucleotide diversity, π; see Nei 1987) and mean numbers of segregating sites, S_n, for silent and nonsynonymous sites, were calculated using a Fortran program written to analyze within- and between-population diversity (see Liu *et al.* 1998). S_n was used to estimate the scaled neutral mutation rate $\theta = 4Ne\mu$ (see Tajima 1993). Each variable insertion/deletion region in a population was treated as a single polymorphic site, without reducing the total number of bases in the calculations of diversity. Calcula-

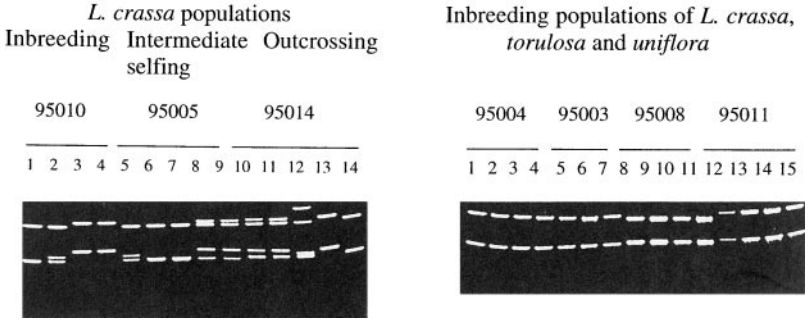


Figure 1.—SSCP gels of a portion of the phosphoglucose isomerase (*PgiC*) gene from outcrossing and inbreeding populations.

tions were done for each population separately, yielding within-population and total diversities π_S and π_T (see Nei 1987). With conservative migration, π_S depends on the metapopulation size, not that of local populations (*e.g.*, Maruyama 1971). The component of diversity between subpopulations was measured as $\pi_T - \pi_S$ (Charlesworth *et al.* 1997). Divergence values between species and haplotypes and their variances were calculated using DnaSP 2.5.2 (Rozas and Rozas 1997) with Jukes and Cantor correction (Nei 1987).

The sequences were tested for departure from neutral expectations by Tajima's (1989), Fu and Li's (1993), and HKA tests (Hudson *et al.* 1987). Linkage disequilibria between variants at different polymorphic sites and Hudson and Kaplan's (1985) estimate of the minimum number of recombination events were estimated using DnaSP. All pairs of informative (nonsingleton) polymorphic sites were tested, excluding sites involved in insertion/deletion polymorphisms. In addition, a program was written to calculate the measure of overall disequilibrium Z_{nS}, and to test this against the neutral expectation assuming no recombination (Kelly 1997). Finally, PAUP version 3.1 was used to infer the evolutionary relationships among the sequenced *PgiC* alleles, using maximum parsimony analyses to generate a 50% majority rule consensus tree with 100 bootstrap iterations (Swofford 1991).

RESULTS

Evidence for a single *PgiC* locus in Leavenworthia species: *Phosphoglucose isomerase isozymes in Leavenworthia species:* Two phosphoglucose isomerase isozyme systems, A and B, were seen in Leavenworthia plants. Examination of the bands from pollen, which does not have plastids, indicated that in all species system A corresponds to the cytosolic phosphoglucose isomerase (usually denoted by *PgiC*; see Ford *et al.* 1995). Three populations of *L. stylosa* were studied, and all were polymorphic for three to five *PgiC* alleles, while four of five populations of *L. crassa* studied (all moderately to highly selfing) were polymorphic for this locus, with two to four alleles segregating. All the highly selfing populations surveyed were monomorphic. The *PgiC* variants segregated as expected for a single locus (Charlesworth and Yang 1998).

The PgiC gene sequence in Leavenworthia crassa: The complete 1680-nucleotide sequence of the *PgiC* gene from *L. crassa* was obtained from cDNA as described in materials and methods. The deduced amino acid sequence is 560 amino acids long, the same as the *A. thaliana PgiC* gene. Based on a single *L. crassa* individual for which the entire coding sequence was obtained, the amino acid identity for these two species is 93.6%, showing that the cDNA sequence corresponds to cytosolic *PgiC*, rather than the very different plastid enzyme. The nucleotide sequence differs from that of *A. thaliana PgiC* at 19.5% of third positions of codons. Cloning of different regions of the gene from cDNA of the *L. crassa* plant studied consistently produced only one type of *L. crassa PgiC* sequence, suggesting that only one locus is present and that the plant studied was a homozygote.

Length variants of PgiC sequences in Leavenworthia species, and evidence for a single PgiC locus in L. stylosa: Using primers PgiC.P1 and PgiC.P2R for the intron 12 region, PCR products from genomic DNA of Leavenworthia species yielded one or two clearly distinct bands in 2.0% agarose electrophoresis gels. All the *L. crassa* plants studied had short (S) bands (~270 bp). *L. torulosa* plants from population 95008 yielded S bands of approximately the same size, while *L. uniflora* (populations 9108 and 95011) gave bands of ~300 bp (medium, M). In *L. stylosa*, however, two band lengths were seen within all four populations. Some plants produced bands ~320 bp in length (long, L), while others yielded 270-bp (S) bands, and some were two-banded and appeared to be heterozygous L/S. Sequencing revealed two different L types (L1 and the 7-bp longer L2; these are indistinguishable without further analysis, as these gels cannot resolve such a small size difference).

L. stylosa thus either has a duplication of this locus, or else it is highly heterozygous for alleles with different lengths of the intron 12 region. The variants are referred to as haplotypes. There are extensive sequence differences between them, which are described in detail below, but before doing so it is essential to establish whether the length variants are allelic. If a duplication is present in some or all *L. stylosa* plants, some individuals should have more than two sequences because they would often be heterozygous at least at one of the two loci in these highly outcrossing populations. We tested this using SSCP. Each allele sequence should yield a two-banded pattern in SSCP gels, so if there is a duplication, more than two sequences will be present, and more than four bands should be seen in some individuals. No plant, however, yielded more than four bands (Table

TABLE 2

Band phenotypes of the *PgiC* intron 12 region alleles of *L. stylosa* plants from different populations, and SSCP band patterns of their PCR products

Population	Number of plants	Phenotype	Number of SSCP bands
95007	1	L2/S	4
95007	1	L2/—[a]	4
95007	2	L/S[b]	4
95007	1	L1/S	4
9113	2	L1/S	4
9113	2	L2/S	4
9113	3	L2/L2	2
9113	2	S/S	2
Hem1	1	L1/L1	3
Hem1	1	S/S	3
Hem1	1	S/S	4
Hem1	2	S/S	2
Hem2	1	L2/—[a]	4
Hem2	2	L1/S	4
Hem2	1	L/S[b]	4

[a] Presence of L2 determined by PCR using L2-specific primers. No further information available.

[b] Amplified only with primers not specific for L2. No sequencing was done to distinguish between L1 and L2.

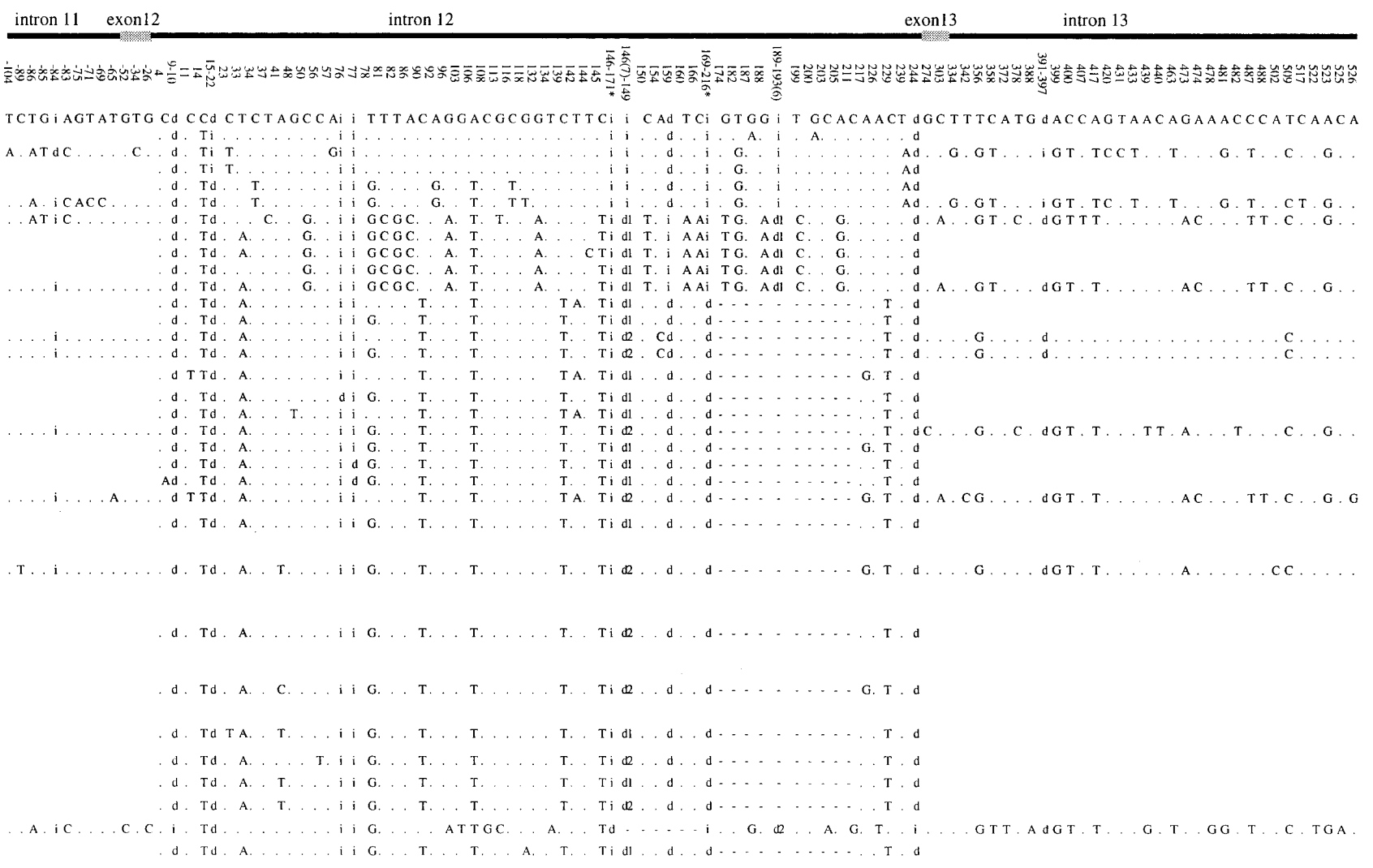
2). Furthermore, direct sequencing of each of the three two-banded individuals produced a single sequence, while cloning and sequencing of 5 to 30 positive colonies from three- or four-banded individuals produced only two types of sequences, with lengths corresponding to one or the other of the haplotypes just described. Thus all individuals with more than two bands appear to be heterozygotes for two of the three haplotypes.

These tests do not show conclusively that there is a single *PgiC* locus, because individuals with two different haplotype sequences could be double homozygotes, *e.g.*, L1/L1 S/S or L2/L2 S/S. Although this seems unlikely in a highly outcrossing plant, it should be tested. Such L/S phenotype plants should not segregate when crossed with S or L plants, whereas Mendelian segregation is expected if they are heterozygotes. Individuals of the L/S type (either L1/S or L2/S) were therefore crossed with plants that had only one band. DNA was extracted from the seeds produced from the crosses, amplified with primers PgiC.P1 and PgiC.P2R, and the band patterns of their PCR products scored electrophoretically. PCR amplifications from some of the L/S parental individuals were done using L1-specific primers (PgiCSTL1, 5′ AAGTAATGCATATTTTGTCC 3′; and PgiCSTL1.R, 5′ GAACGTTAAATCTCTCCAGT 3′) that distinguish the L1 or L2 haplotypes. All L/S plants, both those with L1 alleles and those with L2 alleles, segregated in a set of 17 families involving 11 parental plants, including 6 different L/S parents originating from several different populations. The pooled segregation ratio for reciprocal crosses L/S × S was S:LS = 57:55 and that for L/S × L was LS:L = 42:36. These results agree with single-locus Mendelian segregation (probabilities for χ^2 tests with 1 d.f. of 0.85 and 0.50, respectively), and confirm that L1, L2, and S are all allelic, consistent with the allozyme inheritance results above. There is thus no evidence for a duplicated locus.

Polymorphism pattern, linkage disequilibrium, and recombination in the intron 12 and 13 regions of the *L. stylosa PgiC* gene: The 26 intron 12 region allelic sequences from *L. stylosa* fall into three length variants, as explained above, and Figure 2 shows the details of the extensive differences between these sequences, listing all alleles, both those determined directly and those inferred from SSCP phenotypes. S-type alleles are distinguished from L1 and L2 not only by the deletion from site 169 to 216 but also by three fixed nucleotide substitutions, and L1 has fixed differences from L2 at 14 nucleotide sites and three small indels. Insertion/deletion differences are also seen when intron 12 region sequences from the different species are compared. S types of both *L. crassa* and *L. torulosa* have the same intron size and insertion/deletion variants as the S type of *L. stylosa* (see Figure 2), while the *L. uniflora* M type is more similar to the *L. stylosa* L1 type.

Figure 3 summarizes mean pairwise nucleotide differences within and between species and haplotype classes in the four Leavenworthia species. Diversity values within each haplotype, even between different species, are several times lower than between haplotypes, and values are no higher between haplotypes when the alleles compared are from different species than when they are from the same species. In the consensus parsimony tree based on the *PgiC* intron 12 region (Figure 4), the *L. stylosa* L1 and L2 alleles form two distinct clades. The *L. uniflora* M alleles form a clade with the *L. stylosa* L1 alleles. Although the *L. torulosa* S-type alleles form part of a clade containing all the S-type alleles from both *L. crassa* and *L. stylosa*, sequence divergence data from six loci (Table 3) show that this species is more closely related to *L. stylosa* than is *L. uniflora*. This conclusion is consistent with the chromosome numbers of these species.

As explained above, the variation in the intron 12 region within *L. stylosa* exhibits an evident haplotype structure. We therefore examined the pattern and organization of linkage disequilibrium among segregating nucleotide sites in this region, in the alleles sequenced from this species. Significant linkage disequilibrium was found at the 5% level for >30% of the pairwise comparisons using Fisher's exact test (Sokal and Rohlf 1981). Thirteen percent of the tests remained significant using the Bonferroni procedure to correct for multiple tests (see Weir 1996). Linkage disequilibria are common between any two of the three distinct haplotypic classes, L1, L2, and S.


intron 11
exon12
intron 12
exon13
intron 13
Alleles
Population of origin (number of alleles)
Haplotype
ST-2 95007 (1) L2
ST-1 95007 (1) L2
ST-3 9113 (1) L2
ST-4 9113 (1) L2
ST-5, 6 9113 (2) L2
ST-7 Hem2 (1) L2
ST-8 95007 (1) L1
ST-9 Hem1 (1) L1
ST-10 Hem1 (1) L1
ST-11 Hem2 (1) L1
ST-12 Hem2 (1) L1
ST-13 Hem1 (1) S
ST-14 Hem1 (1) S
ST-15 95007 (1) S
ST-16 95007 (1) S
ST-17, 19, 25 Hem1 (2) Hem2 (1) S
ST-18 Hem1 (1) S
ST-20 Hem1 (1) S
ST-21 9113 (1) S
ST-22 9113 (1) S
ST-23 9113 (1) S
ST-24 9113 (1) S
ST-26 Hem2 (1) S
CR-1 95004 (5), 95003 (5) S
CR-2 9107 (6), 8921 (2), 95005 3), 95010 (4) S
CR-3 8921 (1), 95005 (1), 95010 (1), 8919 (2) S
CR-4 8921 (1), 95005 (1), 95010 (1) S
CR-5, 11 95014 (3), 8919 (1) S
CR-6 95014 (2) S
CR-7, 8, 10 95014 (2), 8919 (2) S
CR-9, 12 8919 (2) S
UN-1 95011 (5), 9108 (6) M
TO-1 95008 (5) S

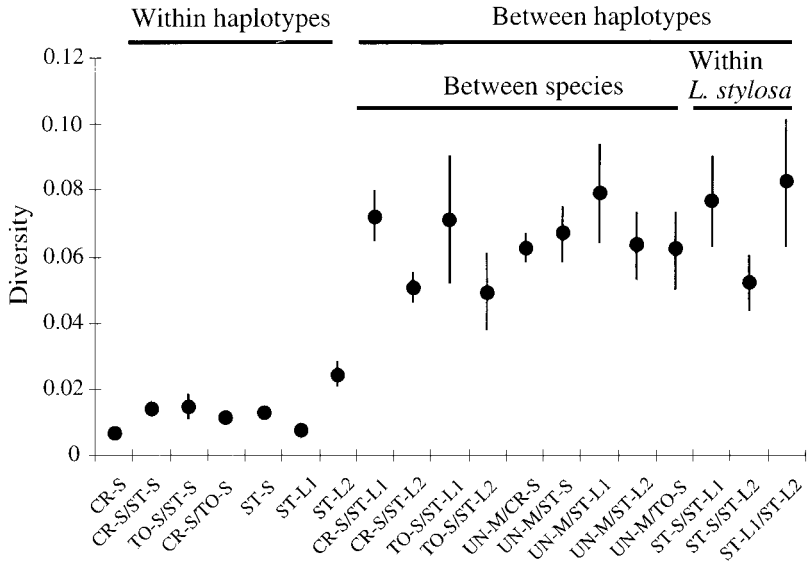


Figure 3.—Mean pairwise nucleotide divergence values within and between Leavenworthia haplotype classes of the same and different species. Species and haplotype abbreviations are the same as in Figure 2. The species are denoted by two-letter abbreviations as follows: CR, *L. crassa*; ST, *L. stylosa*; TO, *L. torulosa*; and UN, *L. uniflora*.

To check whether the same haplotype structure holds outside intron 12 in *L. stylosa*, we sequenced a smaller number of alleles (nine from *L. stylosa* and one each from *L. crassa* and *L. uniflora*) for a larger region 5′ and 3′ of the intron 12 region, giving a total of 128 nucleotides of coding and ~500 nucleotides of intron sequences, spanning introns 11 to 13, and starting 124 nucleotides before the start of intron 12 (see Figure 2). In the coding sequences, we found only one replacement polymorphism (a singleton polymorphism at position 274 in exon 13 of the S haplotype) and two synonymous differences (one of them a singleton polymorphism within the L2 haplotypes at position −34 in exon 12 and another at position 303 in exon 13).

With just these nine alleles, no linkage disequilibria were significant after Bonferroni correction, probably because statistical power to detect linkage disequilibria for polymorphic sites with very asymmetrical allele frequencies is low, given the small number of alleles analyzed (Brown 1975; Lewontin 1995). Nevertheless, the haplotype structure of the variation in intron 12 was discernible in terms of a much larger number of tests significant at the 5% level (43 of 162 pairwise tests, *i.e.*, 27%), compared with tests between sites in other regions, or between sites in intron 12 and those in flanking regions. For comparison, only 3% of the polymorphic sites within intron 12 showed significant nonrandom associations with sites in regions 5′ or 3′ to this intron, and 19% of comparisons between sites in intron 13 were significant, but no significant disequilibria were found between sites in introns 12 and 13. The nonsingleton synonymous polymorphism at position 274 in exon 13 showed nonrandom associations with some polymorphic sites within intron 13 (none significant after Bonferroni correction), but not those within intron 12 (see Figure 5).

Hudson and Kaplan's (1985) method estimates a minimum of six recombination events in the history of the nine *L. stylosa* alleles for which the longer sequence was available (between sites 33–50, 78–90, 106–239, 303–358, 358–378, and 378–474), or at least three using the larger number of alleles for which the shorter sequence is available. Assuming neutrality, the estimated ratio of recombination rate to mutation rate, on a per nucleotide basis, is 2.85, suggesting that recombination in this region is frequent enough to break up nonrandom associations caused by mutation.

Statistical tests of neutrality: Several statistical tests for selection on the polymorphism of the *L. stylosa* intron 12 region of *PgiC* failed to detect deviations from neutrality. The HKA test is based on the null hypothesis that the relative levels of intraspecific polymorphism and interspecific divergence for two loci or regions are as expected if the loci are evolving neutrally (Hudson *et al.*

Figure 2.—Polymorphic sites in the alleles from the four species studied. The allele identification numbers are shown in the left-hand column. Most alleles were sequenced for only the intron 12 region, and 11 were sequenced for a somewhat longer region (see text). The shorter sequences are blank outside the sequenced region. Sequences found multiple times are shown only once, and the numbers of alleles with each sequence are given in the figure. The populations in which each type of allele was found are given in the second column, together with the numbers of instances of each sequence type. The base positions of each variable site are shown above the details of the sequences, numbered relative to the first base of intron 12 in the sequence alignment and using the longest sequence found as a reference. An "i" denotes an insertion, "d" a deletion, and the notations "d1" and "d2" denote sequence differences in an indel. Where a range of base positions is noted, this means that the extent of the indel was variable.

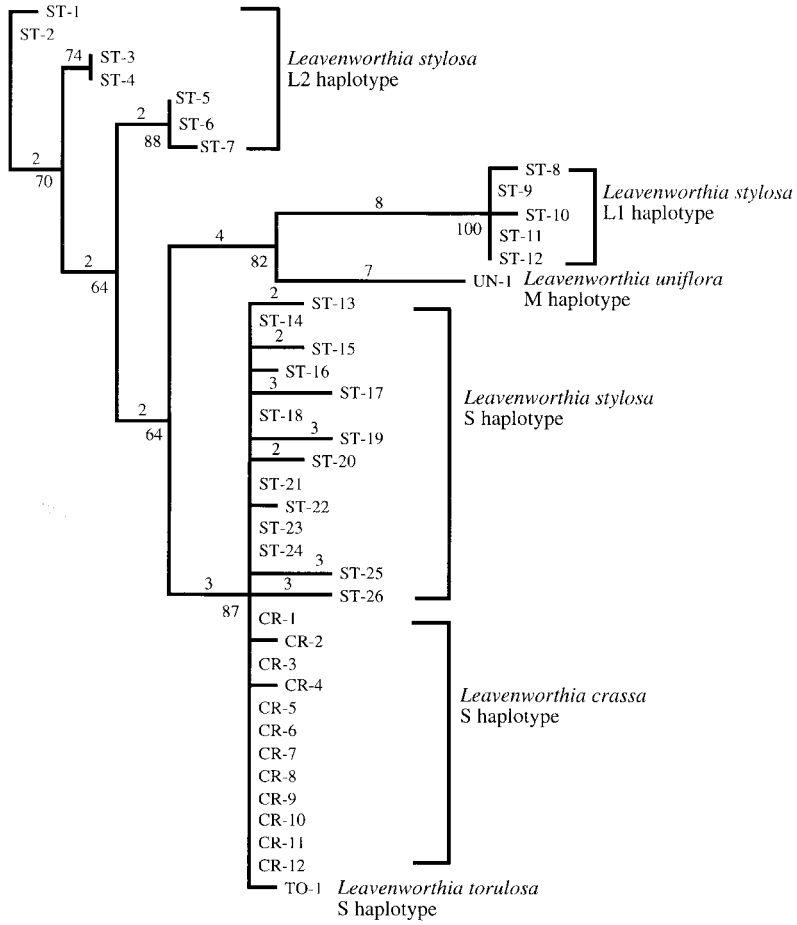


Figure 4.—Consensus parsimony tree based on sequences of the Leavenworthia *PgiC* intron 12 region.

1987). Using the *L. uniflora* M haplotype as the outgroup, the *PgiC* data showed no significant deviations from the neutral model, with any of several Leavenworthia reference loci (Table 4). The results were similar, using the *L. torulosa* S haplotype or the sequences from self-incompatible *L. crassa* populations as outgroups (data not shown). However, as Table 4 shows, divergence between *L. stylosa* and the outgroup species is low compared with the polymorphism levels within *L. stylosa*, so the statistical power of the HKA test to detect selection is low (Hey 1991; Ford and Aquadro 1996). Tajima's tests (Tajima 1989), both for individual populations and at the whole species level, and Fu and Li's tests (Fu and Li 1993) of the data from each species as a whole were also all nonsignificant.

The most striking feature of the data from *L. stylosa* is the haplotype structure and linkage disequilibrium. We therefore performed further tests, more specifically aimed at testing these aspects of the data. The haplotype diversity and number tests of Depaulis and Veuille (1998) for the whole set of sequences contained significantly more allelic types than expected, given the number of segregating sites, even taking into account the estimated recombination frequency (nonsignificant results were obtained for all four individual populations, but our evidence discussed below suggests that they

TABLE 3

Locality of origin of the *L. uniflora* and *L. torulosa* populations and sequence differences between these populations and *L. stylosa* based on data from other loci

		Number of taxon-specific nucleotide fixations compared with *L. stylosa*					
Population	Locality of origin	*Adh1*	*Adh2*	*Adh3*	*GapC2*	*Nir1*	*PgiC*
95008	Highway 157, Bullitt County, KY	2	2	2	5	0	1
95011	Speake, Morgan County, AL	5	6	2	17	2	6
9108	Decatur, Morgan County, AL	5	6	2	17	2	6

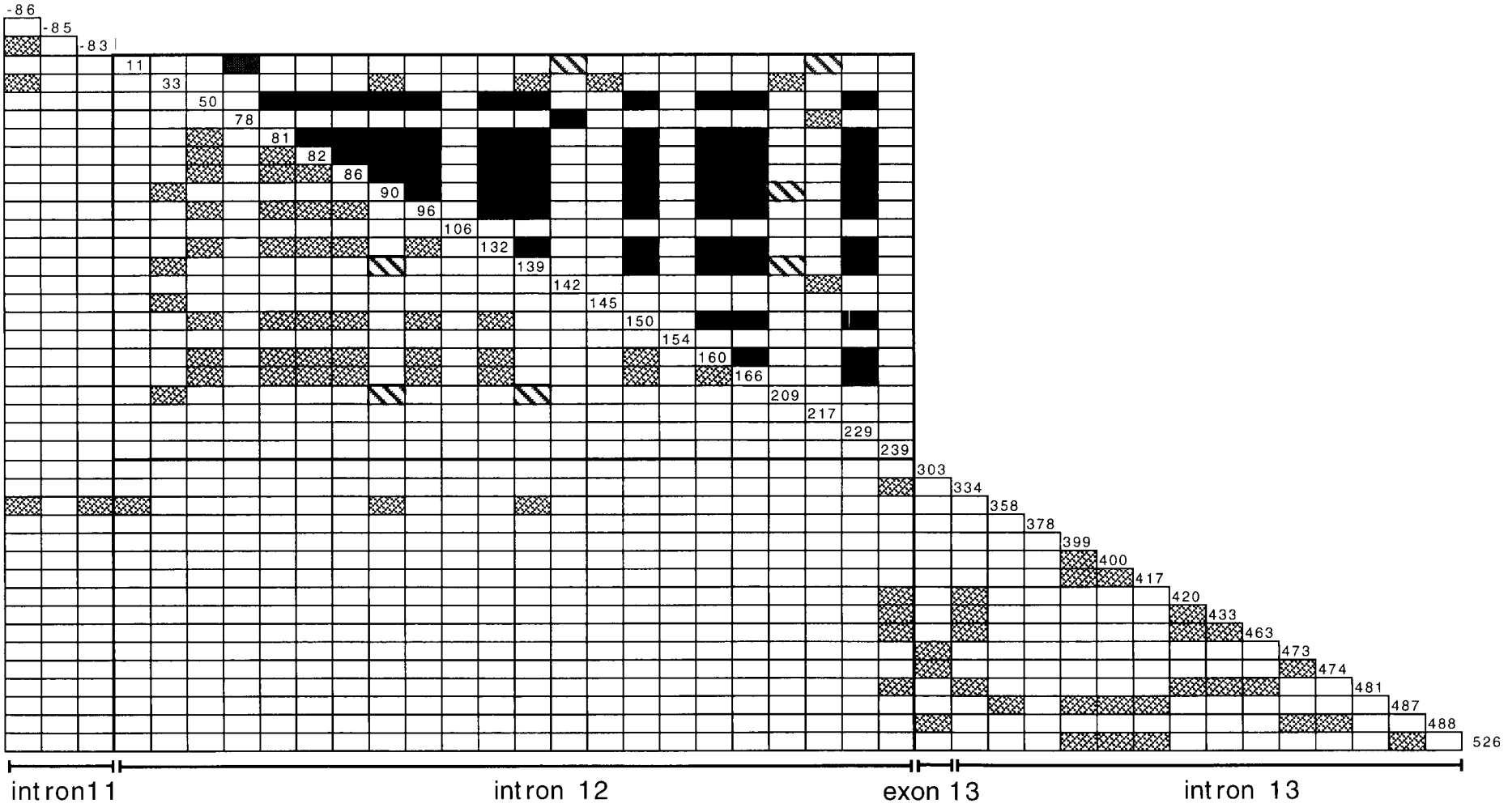


Figure 5.—Pairwise linkage disequilibria between polymorphic sites within *L. stylosa*. All segregating sites are included, excluding singletons and indel regions. The results for the intron 12 region are based on 26 alleles, while those for the complete region are based on only 9 alleles. Comparisons significant at the 0.1% level (which remained significant after Bonferroni correction) are shown as black cells, cross-hatched cells denote significance at the 1% level, and hatched cells the 5% level.

TABLE 4

Comparisons of polymorphism in *L. stylosa* with mean numbers of pairwise differences and numbers of fixed differences between *L. stylosa* and *L. uniflora* (outgroup) for six loci

Locus and type of sequence	Number of alleles	Number of sites analyzed	Number of polymorphic sites	Mean number of differences	*P* value of HKA tests[a]
PgiC introns	26	176	29	7.83	—
Adh1 introns	27	126	26	7.67	0.829
Adh1 exons	27	507	34	10.9	0.874
Adh2 exons	15	350	5	10.5	0.125
Adh3 exons	18	599	25	7.67	0.79
GapC2 introns	7	173	4	14.0	0.11
GapC2 exons	7	239	4	6.71	0.381
Nir1 introns	23	56	7	2.87	0.991
Nir1 exons	23	177	20	6.87	0.831

[a] The results of the HKA tests of the *L. stylosa PgiC* intron 12 region, using the stated gene regions as reference loci.

are not differentiated from one another, and so it is appropriate to test the entire sample of alleles). This excess of alleles is opposite to the intuitive impression created by the existence of the three haplotypes, and is clearly due to the diversity within the haplotypes, which is consistent with their having been maintained for a long time period. It is possible, however, that recombination is more frequent than our estimated value. The test becomes nonsignificant only when the recombination frequency is roughly double the estimated value. Kelly's (1997) test for whether linkage disequilibrium exceeds that expected under neutrality yields values for the four populations of 0.44, 0.30, 0.47, and 0.53 (on the basis of five or eight sequences per population). None of these is statistically significant, based on Kelly's simulations assuming nonrecombining sequences (Kelly 1997). As far as we are aware, no comparable test incorporating recombination is available, so it is uncertain whether such high disequilibrium is compatible with frequent recombination.

Within-population polymorphism levels in outcrossing and inbreeding Leavenworthia species: Figures 6 and 7 summarize the sequence diversity comparisons between populations with different selfing rates for the intron 12 region. The inbreeding populations show the expected pattern of low within-population diversity. As is evident from the results already described, the self-incompatible species *L. stylosa* has very high within-population diversity and low divergence between populations (Figure 6), while the highly selfing species *L. uniflora* and *L. torulosa* have no within-population variation. Comparing the three groups of *L. crassa* populations with different outcrossing rates, we observed a similar pattern: the higher the outcrossing rate, the higher the within-population diversity (Figure 7). Both the self-incompatible and the intermediate selfing populations of *L. crassa* show much lower among-population divergence than the highly selfing populations.

DISCUSSION

Within-population diversity in inbreeding and outcrossing populations of Leavenworthia, and possible causes of low diversity in inbreeding populations: The original purpose of this work was to test the theoretical prediction that highly selfing populations would have a more than twofold reduction in within-population neutral variation compared with closely related outcrossing species. Within-population diversities in all our comparisons (between *L. stylosa* and its inbreeding sister species *L. uniflora* and *L. torulosa*, or between the diversity measures within the three groups of *L. crassa*) clearly show the expected correlation with their outcrossing rates. Within *L. crassa* alone, the self-incompatible popu-

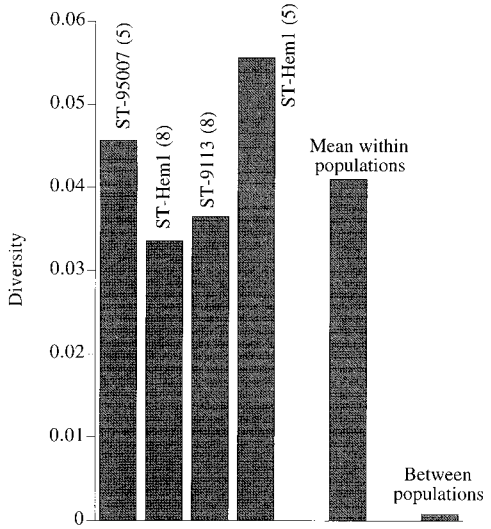


Figure 6.—Sequence diversity within and between populations of *L. stylosa* in a portion of the *PgiC* locus. Nucleotide diversity (π) values are based on ~260 bp of intron sequences from the intron 12 region of the locus. Numbers of alleles sequenced are given in parentheses after the population identification numbers.

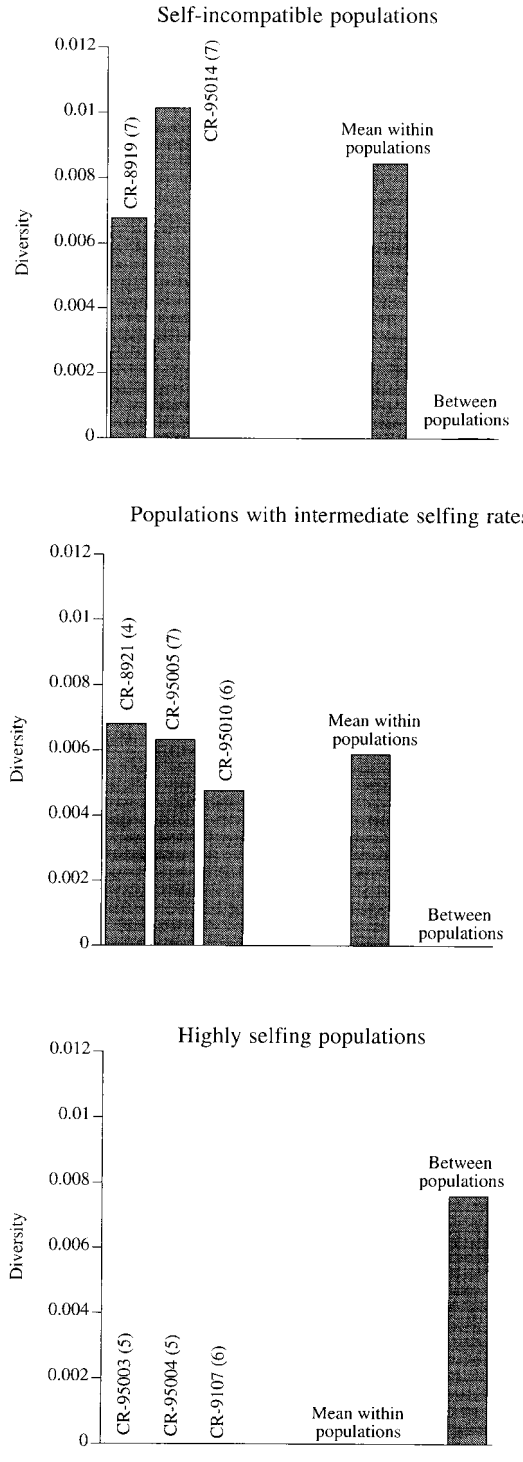


Figure 7.—Comparison of the sequence diversity (π) within and between populations of *L. crassa* based on the same *PgiC* locus region as in Figure 6.

lations show the highest diversity, the intermediate selfing populations have less, while the highly selfing populations have <10% of the values of the self-incompatible populations. There is as yet no general method for computing standard errors for within-population diversities in subdivided populations (Wakeley 1996). One cannot, therefore, test whether the samples from populations with contrasting outcrossing rates could derive from independently replicated similar evolutionary histories (*i.e.*, test the null hypothesis that they do not differ significantly). However, as for our previous results from a region of the *Adh1* including both exon and intron sequences, where no signs of selection were found (Liu *et al.* 1998), the differences are consistent across different populations in each independent comparison that can be made. The observation of dramatic reduction of within-population diversity in highly selfing species agrees with the theoretical predictions. However, it remains difficult to distinguish between various possibilities that could cause low diversity in the inbreeding populations.

We previously concluded that selective sweeps could not account for the reduced diversity in *L. uniflora* populations (Liu *et al.* 1998), on the basis of the finding of different alleles in different populations. It is now clear, however, that the populations studied represent two different species, *L. uniflora* and *L. torulosa*. Low diversity, without between-population differences, as in the two *L. uniflora* populations studied, could be explained by either hitchhiking or bottlenecks. In either interpretation, many loci in the genome should be similarly affected and have low diversity. This is, in fact, the case for several loci (Liu 1998).

Selective sweeps should certainly be considered a possibility, because the evolutionary loss of self-incompatibility in the selfing taxa must have involved hitchhiking events while the gene causing selfing was spreading through the populations. In an outcrosser, or in partially selfing populations such as *L. crassa*, a hitchhiking event at one locus would almost certainly not affect a randomly chosen locus. It is unlikely *a priori* that the gene causing the loss of incompatibility would be tightly linked to *PgiC*, but in the situation where an allele for selfing is spreading there might be little opportunity for recombination to separate variants at the two loci. Selfing would, however, have to be quite extreme, as even rare outcrossing would allow recombination and prevent effects on unlinked or loosely linked loci (Hedrick 1980). Selective sweeps are therefore quite unlikely to explain our data.

Magnitude and structure of the diversity in *L. stylosa*: An unexpected difficulty in ascertaining what has led to low diversity in the *PgiC* gene in inbreeding Leavenworthia species arises from the fact that two of the selfing species are closely related to *L. stylosa*, and the haplotype structure of the sequence variation in *PgiC* in this species suggests that the variation may be maintained by balancing selection, although the reason for this is not known. It is therefore worth discussing the diversity results from this species in some detail. In *L. stylosa*, our estimates of within-population diversity are high, compared with those from Drosophila species (reviewed by Moriyama and Powell 1996) and compared with other published

plant data, although there are currently few comparable data. Many of the available studies (*e.g.*, Gaut and Clegg 1993b; Cummings and Clegg 1998) used cultivated strains, and even the high diversity found in outcrossing plants such as maize may represent only a subset of the diversity present in wild species (see, *e.g.*, Cui *et al.* 1995). Other studies are of the highly inbreeding plant *A. thaliana* (Miyashita *et al.* 1993; Hanfstingl *et al.* 1994; Innan *et al.* 1996; Bergelson *et al.* 1998; Purugganan and Suddith 1998), which may be expected to have low genetic diversity (see above).

Only one extensive study of DNA sequence diversity of a *PgiC* gene with an allozyme polymorphism from natural plant populations of an outcrosser is available, from the dioecious species *D. tokoro* (Terauchi *et al.* 1997). In this case, the diversity estimates for intron regions averaged 0.028, lower than those reported here, though diversity increased in the 3′ direction within the gene and was highest in intron 10, the furthest extent sequenced. Although the regions sequenced contained the putative replacement sites causing the allozyme polymorphism, no haplotype structure was detected in *D. tokoro* (Terauchi *et al.* 1997). Within *L. stylosa* populations, the high overall diversity is clearly partly due to the presence of the different haplotypic classes in intron 12 but, except for low diversity within the L1 type, even the within-haplotype diversity values (Figure 3) are quite high, consistent with estimates for synonymous sites and noncoding regions in five further loci in this species (Charlesworth *et al.* 1998; Liu 1998; Liu *et al.* 1998). They are also similar to estimates from maize (Shattuck-Eidens *et al.* 1990; Gaut and Clegg 1993a; Henry and Damerval 1997).

Evidence for long-term maintenance of haplotypic classes in *L. stylosa*: The diversity values reported here are based mainly on a single intron region of the *PgiC* locus, chosen for study because it was expected that changes in intron regions would be unlikely to be under selection. We found no evidence for a balanced polymorphism in the flanking exons that might explain the linkage disequilibrium and divergence of the three major haplotypic types in intron 12. The *PgiC* locus has an allozyme polymorphism in both *L. stylosa* and *L. crassa*, but the only replacement polymorphism in the region we have sequenced (in exon 13) was seen only within the *L. stylosa* S haplotype. Furthermore, on the basis of 18 plants typed for both *PgiC* allozymes and intron 12 haplotypes, no correspondence was seen, implying that the allozyme variants are not in linkage disequilibrium with the variants in this region and suggesting that the amino acid replacements responsible for the allozyme variation are elsewhere in the protein. This is consistent with the interpretation of Terauchi *et al.* (1997) that the allozyme variants in *D. tokoro* are in the more N-terminal region of the protein than the regions we sequenced.

Selective maintenance of the diversity is, however, suggested by the remarkably high diversity between different haplotypic classes, including multiple fixed differences, which imply that the different haplotypes have been present for long periods of evolutionary time (see Figure 3). The similarity of the S-type sequences among the four Leavenworthia species (Figures 2 and 3) may reflect recent origins of these species, consistent with similar data based on other genes, including an alcohol dehydrogenase locus (Charlesworth *et al.* 1998; Liu *et al.* 1998). The finding that between-haplotype diversity is no higher when the alleles compared are from different species than when they come from a single species (Figure 3) further suggests that the two different haplotypes in the two highly selfing taxa in the $n = 15$ group of species, *L. uniflora* and *L. torulosa* (L1 and S, respectively), derive from a polymorphic progenitor population species; as both these haplotypes are present at high frequencies in contemporary *L. stylosa* populations, this is quite possible. Because *L. torulosa* is the more recently evolved of the two selfers (see Table 3), it is not possible that its S haplotype simply represents the ancestral condition, and that the haplotype diversity in *L. stylosa* arose since the time when the selfing taxa became isolated from their progenitors. The implication is thus that the polymorphism has persisted during the time that the two speciation events occurred to give rise to the selfers, *i.e.*, that it represents a "transspecific polymorphism," such as is seen when alleles are maintained by balancing selection, for instance at MHC (*e.g.*, Edwards *et al.* 1997) and self-incompatibility loci (*e.g.*, Ioerger *et al.* 1990; Dwyer *et al.* 1991).

If the allelic types of intron 12 have indeed persisted for large amounts of time, this suggests that this region is under some form of balancing selection. This, in turn, implies that low diversity in the related inbreeding populations (*L. uniflora* and *L. torulosa*) could be caused by failure to maintain the allelic diversity under high inbreeding, as is expected to occur for overdominant selection (Kimura and Ohta 1971; Charlesworth and Charlesworth 1995). If allozyme polymorphisms are indeed maintained by balancing selection, such loci may be unsuitable for studies of the effects of hitchhiking or background selection. It is thus important to ask whether the allelic structure we find for this region of *PgiC* could have arisen under neutrality.

Could the linkage disequilibria in *L. stylosa* have arisen under neutrality? Population subdivision (Kimura and Ohta 1971; Nei 1987) can be ruled out as an explanation for the linkage disequilibrium, because the haplotype polymorphism is present in all the *L. stylosa* populations studied (see Figure 2). Furthermore, analyses of polymorphisms at other loci in the same populations give no evidence for subdivision of *L. stylosa* populations. For an alcohol dehydrogenase locus, *Adh1*, diversity between populations of this species was very low, compared with that within populations (Charlesworth *et al.* 1998; Liu *et al.* 1998), and similar results have been

obtained for other loci (Liu 1998). The alternative possibility, that populations of *L. stylosa* have gone through bottlenecks and/or population expansions, which can induce linkage disequilibrium (Tachida 1994; Kirby and Stephan 1995), is inconsistent with our finding of high within-species diversity. Furthermore, if this were the cause, linkage disequilibria should also be found for other gene loci. However, no other cases were found in *L. stylosa*, although diversity was also high in this species for several other loci studied (Liu 1998).

Any explanation for the haplotype structure in *L. stylosa* must be consistent with these data. It must also take into account the evidence for recombination. The estimated ratio of recombination rate to mutation rate per base across intron 12 and its neighboring exons was roughly estimated, assuming neutrality, to be 2.85 (Hudson and Kaplan 1985). The indels in intron 12 also suggest that the haplotypes have recombined in the ancestry of *L. stylosa*, because the deleted condition of nucleotides 146–149 is shared by haplotypes L1 and S, the deletion at 159 is shared by L2 and S, and the large insertion at 162–216 is shared by L1 and L2 (see Figure 2). In addition, diversity levels in the *L. stylosa* populations in this gene region are high, even within haplotypic classes, suggesting a large effective population size, and ruling out a very small *Nr* value. This makes it unlikely that this gene is located in a genomic region with low recombination; in all other systems studied, in both animals and plants, genes in such regions have low diversity (Begun and Aquadro 1992; Dvorak *et al.* 1998; Stephan and Langley 1998).

All our findings therefore support the view that this gene region recombines, and that it is unlikely that *PgiC* is in a chromosomal region with an inversion. At present, this limits our ability to test whether the data are consistent with neutrality, because the test currently available to assess whether linkage disequilibrium is greater than that likely to be produced under neutrality assumes that sequences do not recombine (Kelly 1997); in the presence of recombination, this is a highly conservative test for selection, so the interpretation of our data remains uncertain. Because balancing selection can lead to high linkage disequilibria among polymorphic sites and possibly to haplotype blocks (Kimura 1956; Lewontin 1974) and has been invoked to explain linkage disequilibrium and unusually high synonymous and nonsynonymous diversity in HLA genes (*e.g.*, Markow *et al.* 1993; Trachtenberg *et al.* 1995), it is clearly desirable to develop tests that are sensitive to this deviation from neutral expectations. Even though no test is currently available, we should be cautious in interpreting the lower diversity in the inbreeding species related to *L. stylosa* in terms of selective sweeps or background selection, given the possibility that the difference may be caused by loss of a selectively maintained balanced polymorphism.

Conclusions: Because of the difficulty of distinguishing between different possible causes of the effect of selfing rates on sequence diversity, loci with allozyme polymorphisms may be unsuitable for studies of the effect of breeding systems on diversity. The *PgiC* study presented here thus yields only one set of results, from *L. crassa*, that can be used to estimate the magnitude of any effect of selfing on sequence; the results support our previous conclusion based on an alcohol dehydrogenase locus that a more than twofold reduction occurs (Liu *et al.* 1998).

The present results, however, have the interesting implication that the locus studied here appears in *L. stylosa* to be under balancing selection of a kind that does not maintain the variants under high inbreeding. Overdominance is one such form of selection (including mechanisms with similar properties, such as temporally varying environments; see Nagylaki 1994). Frequency-dependent selection seems unlikely for the maintenance of the diversity, because it should not be lost when populations evolve inbreeding. Rather, one would expect that inbreeding would generate homozygotes and that, by analogy with what occurs in outcrossing populations, this would produce associative overdominance at other loci, which would tend to prolong the time during which the variants are retained (Ohta 1971; Sved 1972). Although background selection also causes loss of diversity in populations that have evolved high selfing (Charlesworth *et al.* 1993), it would not be expected to cause loss of polymorphisms maintained by frequency-dependent selection unless this were very weak, which seems inconsistent with the apparent long time that the variants in Leavenworthia have been maintained. However, quantitative theoretical predictions for frequency-dependent selection in finite populations under inbreeding are not yet available. At present, therefore, some form of overdominant selection in *L. stylosa* seems likely. It must, however, be reiterated that the diversity reported here is within an intron, and is apparently not closely correlated with the *PgiC* allozyme variability.

We thank Li Zhang and Zhe Yang for genomic DNA and for help with technical aspects of molecular methods, F. Depaulis for performing his test of neutrality, and J. Comeron and B. Charlesworth for discussions. We also thank the greenhouse staff of the University of Chicago greenhouses for excellent plant care and Drs. E. E. Lyons, G. Hilton, and T. E. Hemmerly for plant material. This research was supported by National Institutes of Health grant P016M5035504, National Science Foundation Dissertation Improvement Grant DEB 9532071, and by the Natural Environment Research Council of Great Britain.

LITERATURE CITED

Barrett, S. C. H., L. D. Harder and A. C. Worley, 1996 Comparative biology of plant reproductive traits. Philos. Trans. R. Soc. Lond. B Biol. Sci. **351:** 1272–1280.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Bergelson, J., E. Stahl, S. Dudek and M. Kreitman, 1998 Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics **148:** 1311–1323.

Brown, A. H. D., 1975 Sample sizes required to detect linkage disequilibrium between two and three loci. Theor. Pop. Biol. **8:** 184–201.

Brown, A. H. D., 1979 Enzyme polymorphism in plant populations. Theor. Appl. Genet. **15:** 1–42.

Charlesworth, B., M. Nordborg and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70:** 155–174.

Charlesworth, D., and B. Charlesworth, 1995 Quantitative genetics in plants: the effect of breeding system on genetic variability. Evolution **49:** 911–920.

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

Charlesworth, D., M. T. Morgan and B. Charlesworth, 1993 Mutation accumulation in finite outbreeding and inbreeding populations. Genet. Res. **61:** 39–56.

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

Chomczynski, P., and N. Sacchi, 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156–159.

Christiansen, C. S., 1993 A phylogenetic approach to floral evolution in the mustard genus, Leavenworthia. B.A. Thesis, Amherst College.

Cui, Y. X., G. W. Wu, C. W. Magill and K. F. Schertz, 1995 RFLP-based assay of *Sorghum bicolor* (L.) Moench diversity. Theor. Appl. Genet. **90:** 787–796.

Cummings, M. P., and M. T. Clegg, 1998 Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): An evaluation of the background selection hypothesis. Proc. Natl. Acad. Sci. USA **95:** 5637–5642.

Depaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite sites model. Mol. Biol. Evol. (in press).

Dvorak, J., C. M. Luo and Z. L. Yang, 1998 Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics **148:** 423–434.

Dwyer, K. G., M. A. Balent, J. B. Nasrallah and M. E. Nasrallah, 1991 DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation. Plant Mol. Biol. **16:** 481–486.

Edwards, S. V., K. Chesnut, Y. Satta and E. K. Wakeland, 1997 Ancestral polymorphism of *Mhc* class II genes in mice: implications for balancing selection and the mammalian molecular clock. Genetics **146:** 655–668.

Ford, M. J., and C. F. Aquadro, 1996 Selection on X-linked genes during speciation in the *Drosophila athabasca* complex. Genetics **144:** 689–703.

Ford, V. S., B. R. Thomas and L. D. Gottlieb, 1995 The same duplication accounts for the PgiC genes in *Clarkia xantiana* and *C. lewisii* (Onagraceae). Syst. Bot. **20:** 147–160.

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

Gaut, B. S., and M. T. Clegg, 1993a Molecular evolution of the *Adh1* locus in the genus Zea. Proc. Natl. Acad. Sci. USA **90:** 5095–5099.

Gaut, B. S., and M. T. Clegg, 1993b Nucleotide polymorphism in the adh1 locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135:** 1091–1097.

Gottlieb, L. D., 1982 Conservation and duplication of isozymes in plants. Science **216:** 373–380.

Gottlieb, L. D., and V. S. Ford, 1996 Phylogenetic relationships among the sections of Clarkia (Onagraceae) inferred from the nucleotide sequences of *PgiC*. Syst. Bot. **21:** 1–18.

Gottlieb, L. D., and L. C. Greve, 1981 Biochemical properties of duplicated isozymes of phosphoglucose isomerase in the plant *Clarkia xantiana*. Biochem. Genet. **19:** 155–172.

Gottlieb, L. D., and N. F. Weeden, 1979 Gene duplication and phylogeny in Clarkia. Evolution **33:** 1024–1039.

Hamrick, J. L., and M. J. Godt, 1990 Allozyme diversity in plant species, pp. 43–63 in *Plant Population Genetics, Breeding, and Genetic Resources*, edited by A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir. Sinauer, Sunderland, MA.

Hamrick, J. L., and M. J. Godt, 1996 Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B Biol. Sci. **351:** 1291–1298.

Hanfstingl, U., A. Berry, E. A. Kellogg, J. T. Costa, W. Rüdiger *et al.*, 1994 Haplotype divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics **138:** 811–828.

Hebert, P. D. N., and M. J. Beaton, 1989 *A Practical Handbook of Cellulose Acetate Electrophoresis*. Helena Laboratories, Beaumont, TX.

Hedrick, P. W., 1980 Hitch-hiking: a comparison of linkage and partial selfing. Genetics **94:** 791–808.

Henry, A. M., and C. Damerval, 1997 High rates of polymorphism and recombination at the *Opaque-2* locus in maize. Mol. Gen. Genet. **256:** 147–157.

Hey, J., 1991 The structure of genealogies and the distribution of fixed differences between DNA sequence samples from natural populations. Genetics **128:** 831–840.

Hongyo, T., G. S. Buzard, R. N. Calvert and C. M. Weghorst, 1993 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-strand conformation polymorphism analyses. Nucleic Acids Res. **21:** 3637–3642.

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.

Innan, H., F. Tajima, R. Terauchi and N. T. Miyashita, 1996 Intragenic recombination in the *Adh* locus of a wild plant *Arabidopsis thaliana*. Genetics **143:** 1761–1770.

Ioerger, T. R., A. G. Clark and T.-H. Kao, 1990 Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. USA **87:** 9732–9735.

Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics **146:** 1197–1206.

Kimura, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. Evolution **10:** 278–287.

Kimura, M., and T. Ohta, 1971 Theoretical Topics in Population Genetics. Princeton University Press, Princeton, NJ.

Kirby, D. A., and W. Stephan, 1995 Haplotype test reveals departure from neutrality in a segment of the white gene of Drosophila melanogaster. Genetics **141:** 1483–1490.

Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Lewontin, R. C., 1995 The detection of linkage disequilibrium in molecular sequence data. Genetics **140:** 377–388.

Liu, F., 1998 Genetic diversity in *Leavenworthia* populations with different inbreeding levels. The Effect of Breeding System on the Level and Pattern of Molecular Variation in Plant Populations. Ph.D. Thesis, University of Chicago.

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

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

Markow, T., P. W. Hedrick, K. Zuerlein, J. Danilovs, J. Martin *et al.*, 1993 HLA polymorphism in the Havasupai: evidence for balancing selection. Am. J. Hum. Genet. **53:** 943–952.

Maruyama, T., 1971 An invariant property of a structured population. Genet. Res. **18:** 81–84.

Miyashita, N. T., M. Aguadé and C. H. Langley, 1993 Linkage disequilibrium in the white locus region of *Drosophila melanogaster*. Genet. Res. **62:** 101–109.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Murray, M. G., and W. F. Thompson, 1980 Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. **8:** 4321–4325.

Nagylaki, T. (Editor), 1994 *Biomathematics*. Springer-Verlag, New York.
Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
Ohta, T., 1971 Associative overdominance caused by linked detrimental mutations. Genet. Res. **18:** 277–286.
Pollak, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.
Price, R. A., J. D. Palmer and I. A. Al-Shehbaz, 1994 Systematic relationships of *Arabidopsis*: a molecular and morphological approach, pp. 7–19 in *Arabidopsis*, edited by E. M. Meyerowitz and C. R. Somerville. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Purugganan, M. D., and J. I. Suddith, 1998 Molecular population genetics of the *Arabidopsis* CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95:** 8130–8134.
Rollins, R. C., 1963 The evolution and systematics of *Leavenworthia* (Cruciferae). Contrib. Gray Herb. Harv. Univ. **192:** 3–98.
Rozas, J., and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.
Schoen, D. J., and A. H. D. Brown, 1991 Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl. Acad. Sci. USA **88:** 4494–4497.
Schoen, D. J., A.-M. L'Heureux, J. Marsolais and M. O. Johnston, 1997 Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). Evolution **51:** 1090–1099.
Shattuck-Eidens, D. M., M. Russell, N. Bell, S. L. Neuhausen and T. Helentjaris, 1990 DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. Genetics **126:** 207–217.
Sokal, R. R., and F. J. Rohlf, 1981 *Biometry: The Principles and Practice of Statistics in Biological Research*, Ed. 2. W. H. Freeman, New York.
Stebbins, G. L., 1957 Self fertilization and population variation in the higher plants. Am. Nat. **91:** 337–354.
Stephan, W., and C. H. Langley, 1998 DNA polymorphism in *Lycopersicon* and crossing-over per physical length. Genetics **150:** 1585–1593.
Sved, J. A., 1972 Heterosis at the level of the chromosome and at the level of the gene. Theor. Popul. Biol. **3:** 491–506.
Swofford, D. L., 1991 PAUP: phylogenetic analysis using parsimony, version 3.1. computer program distributed by the Illinois Natural History Survey, Champaign, IL.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis. Genetics **123:** 585–595.
Tajima, F., 1993 Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer, Sunderland, MA.
Tachida, H., 1994 Decay of linkage disequilibrium in a finite island model. Genet. Res. **64:** 137–144.
Terauchi, R., T. Terachi and N. T. Miyashita, 1997 DNA polymorphism at the Pgi Locus of a Wild Yam, *Dioscorea tokoro*. Genetics **147:** 1899–1914.
Trachtenberg, E. A., H. A. Erlich, O. Rickards, G. F. Destefano and W. Klitz, 1995 HLA class II linkage disequilibrium and haplotype evolution in the Cayapa Indians of Ecuador. Am. J. Hum. Genet. **57:** 415–424.
Thomas, B. R., V. S. Ford, E. Pichersky and L. D. Gottlieb, 1993 Molecular characterization of duplicate cytosolic phosphoglucose isomerase genes in *Clarkia* and comparison to the single gene in Arabidopsis. Genetics **135:** 895–905.
Wakeley, J., 1996 The variance of pairwise nucleotide differences in two populations with migration. Theor. Popul. Biol. **49:** 39–57.
Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer, Sunderland, MA.
Wyatt, R., E. A. Evans and J. C. Sorenson, 1992 The evolution of self-pollination in granite outcrop species of *Arenaria* (Caryophyllaceae). VI. Electrophoretically detectable genetic variation. Syst. Bot. **17:** 201–209.

Communicating editor: W. Stephan