The Royal Society

Genetic diversity in *Leavenworthia* populations with different inbreeding levels

F. Liu, L. Zhang and D. Charlesworth*†

Department of Ecology and Evolution, University of Chicago, 1101 E 57th Street, Chicago, IL 60637, USA

Levels of neutral genetic diversity within and between populations were compared between outcrossing (self-incompatible) and inbreeding populations in the annual plant genus *Leavenworthia*. Two taxonomically independent comparisons are possible, since self-incompatibility has been lost twice in the group of species studied. Within inbred populations of *L. uniflora* and *L. crassa*, no DNA sequence variants were seen among the alleles sampled, but high diversity was seen in alleles from populations of the outcrosser *L. stylosa*, and in self-incompatible *L. crassa* populations. Diversity between populations was seen in all species. Although total diversity values were lower in the sets of inbreeding populations, between-population values were as high, or higher, than those in the outcrossing taxa. Possible reasons for these diversity patterns are discussed. As the effect of inbreeding appears to be a greater than twofold reduction in diversity, we argue that some process such as selection for advantageous mutations, or against deleterious mutations, or bottlenecks occurring predominantly in the inbreeders, appears necessary to account for the findings. If selection for advantageous mutations is responsible, it appears that it must be some form of local adaptive selection, rather than substitution of alleles that are advantageous throughout the species. This is consistent with the finding of high between-population diversity in the inbreeding taxa.

Keywords: genetic diversity; *Leavenworthia*; selfing; alcohol dehydrogenase

1. INTRODUCTION

A central problem in population genetics is the maintenance of genetic variability within and between natural and other populations (e.g. Kimura 1983; Kreitman 1996). Situations in which differences in diversity patterns can be discerned can illuminate factors that play important roles in determining genetic variability. For instance, the finding of lowered sequence diversity at loci in regions of low versus high recombination regions of *Drosophila* species, and in species of the plant *Lycopersicon* (C. H. Langley and W. Stephan, personal communication) has been interpreted in terms of selective sweeps (or 'hitch-hiking', see Maynard Smith & Haigh 1974) when advantageous mutations occur within a region where recombination rarely occurs (Aguadé, *et al.* 1989; Begun & Aquadro 1992; Langley 1990; Stephan & Langley 1989). As a genotype with an advantageous mutation spreads in a non-recombining population, the entire population will come to have this genotype, causing loss of diversity in the population.

A different possibility is that diversity is reduced by hitch-hiking effects as natural selection eliminates deleterious mutations ('background selection': Charlesworth *et al.* 1993). If these mutations are frequent enough in a genomic region with low recombination rates, a high frequency of individuals will be subject to selective elimination because they carry mutations, and so effective population sizes (N_e values), which determine equilibrium levels of neutral genetic variability in populations (Kimura 1983), are reduced. This is similar to what happens in populations selected by humans for desirable characteristics; human selection means that some proportion of the plants or animals are condemned to be non-breeders, i.e. the effective size of the herd or stock is reduced. As with selective sweeps, this process should occur particularly strongly in inbreeding populations, due to their lack of effective recombination (i.e. the rarity of production of recombinant genotypes, see Narain 1966). Additionally, increased homozygosity due to inbreeding reduces N_e. Inbred populations are homozygous, i.e. individuals often carry two identical genomes. Complete inbreeding thus halves N_e (Pollak 1987) and a very inbred population behaves similarly to one of half its actual size, at least as far as molecular variation is concerned. Finally, bottlenecks in population size may be more extreme in inbreeders (Schoen & Brown 1991), in which a single seed can found a new population (Baker 1955).

Given these theoretical expectations, it is important to obtain estimates of genetic diversity from populations differing in their levels of inbreeding, to test the theoretical prediction that inbreeding populations will have reduced neutral diversity, and estimate the magnitude of any effect, and to try and distinguish which of these factors may have caused any differences in diversity that are found. An effect of breeding system differences on allozyme diversity is quite well documented (Hamrick & Godt 1996; Jarne & Staedler 1995), but such variants may

*Author for correspondence.
†Present address: Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JT.

Received 8 October 1997 *Accepted* 3 November 1997

be selectively maintained (e.g. Hudson *et al.* 1994; Karl & Avise 1992; Kreitman 1996; Pogson & Zouros 1994), so theories for neutral differences may not apply to them. Even assuming selective neutrality, differences between genotypes for allozymes probably do not increase linearly with genealogical times (Marshall & Brown 1975), and cannot therefore be used in tests that may reveal the causes of any differences found. It is therefore important to study DNA sequence diversity, to provide data on silent variants, which can be quantified in terms of standard measures, are probably close to neutral, and can be used in tests for selection.

At present, no good quantitative data are available on within-population diversity at the DNA level for plants, despite an abundance of data on differences between populations of cultivated and wild taxa (e.g. Clegg 1990; Schaal *et al.* 1991; Henry & Damerval 1997). Diversity has been documented in the chloroplast genome, between populations (Lavin *et al.* 1991; Soltis *et al.* 1989) and within populations (Banks & Birky 1985; Dumolin *et al.* 1995; McCauley 1994; Milligan 1991), but different breeding systems have not been compared, and no published sequence data exist. The few sequence diversity data available for plants are all from *A. thaliana* (Hanfstingl *et al.* 1994; Innan *et al.* 1996; Miyashita *et al.* 1996), but no comparison with outcrossing species currently exists. In *Mimulus*, restriction fragment length polymorphism (RFLP) variation in chloroplast DNA was lower in inbreeding *M. micranthus* than in the more outcrossing *M. guttatus*, but no data on within-population diversity were obtained (Fenster & Ritland 1992). The sole studies documenting effects of breeding system on nuclear DNA diversity employed RFLP variants (Miller & Tanksley 1990), minisatellite (Wolff *et al.* 1994) or microsatellite variants (Awadalla & Ritland 1997). In these studies, diversity was not quantified in terms of standard measures, but was clearly less in inbred than outcrossing taxa even though the species compared by Awadalla & Ritland (1997) differed only moderately in selfing rate. The inbreeding plant *Arabidopsis thaliana* also shows very low within-population microsatellite diversity (Todokoro *et al.* 1995). However the selective significance of these length differences, and the mutation process (e.g.Di Rienzo *et al.* 1994; Estoup *et al.* 1995; Michalakis & Veuille 1996), are uncertain. Similar uncertainties exist for random amplified polymorphic DNA (RAPD) marker data, and such diversity is also difficult to relate quantitatively to measures using other kinds of markers. We are unaware of any studies of RAPD diversity within natural populations that compares taxa with different breeding systems.

Ideally, the effect of breeding system on diversity should be studied in a set of related taxa differing only in the breeding system differences that are the focus of interest. Plants offer opportunities to make such comparisons. We here report some data from plants of the mustard genus, *Leavenworthia*, which is a classic example of breeding system evolution (Lloyd 1965; Lyons & Antonovics 1991; Rollins 1963; Solbrig & Rollins 1977). The general aim of our work was to obtain estimates of genetic diversity from populations with different levels of inbreeding, to test whether inbreeding populations have reduced DNA sequence diversity compared with outbreeding ones, and to estimate the magnitude of any such difference.

2. MATERIALS AND METHODS

(a) *The genus* Leavenworthia

Leavenworthia is a small genus of eight diploid annual species in the Brassicaceae, the same family as the very well studied plant *A. thaliana*, and has breeding systems ranging from self-incompatible populations to highly selfing. *L. stylosa* and some populations of *L. crassa* are self-incompatible, while most *L. crassa* populations are self-compatible, with varying frequencies of self-fertilization, and *L. uniflora* is uniformly highly inbreeding (Lloyd 1965; Rollins 1963; Solbrig & Rollins 1977; Lyons & Antonovics 1991). At least two independent comparisons between more outcrossing and more inbreeding populations are thus possible. One is between the sister species *L. stylosa* and *L. uniflora*, in the group with $n=15$ chromosomes (Christiansen 1993), and the other involves comparisons between different populations of *L. crassa*, in a different section of the genus having $n=11$ chromosomes. All species are restricted to the southeastern USA and are confined to 'Cedar glade' habitats, so that populations are restricted in size, patchily distributed and isolated from one another by unsuitable habitat, though *L. uniflora* has a wider distribution than the other species (Quarterman 1950; Rollins 1963).

(b) *Population samples*

Population samples were collected from three populations of each of *L. stylosa* and *L. uniflora*, and eight populations of *L. crassa* (see table 1). Plants were grown from seeds from the different populations (see Charlesworth *et al.* 1994), and breeding systems for populations for which this information was not already available were estimated by measuring self-incompatibility or autogamous fruit set in the greenhouse. These data were consistent with previous studies of the species (above), and will be described in detail elsewhere. They are summarized in table 1.

(c) *Cloning, characterization of the loci and sequence analysis*

To study the *Leavenworthia* alcohol dehydrogenases, sequences of plant alcohol dehydrogenases were obtained from GenBank. The amino acid sequences from the dicotyledon species: *A. thaliana*, strawberry (*Fragaria ananassae*) and potato (*Solanum tuberosum*) were aligned using SeqEd version 1.0.3 (Applied Biosystems). The aligned base sequences were examined to find conserved regions of at least 20 bases from which two internal primers were designed (S1: 5′GATGTT(C, A)TACTTCTGGGAAGC3′ and S2: 5′ATCG(C)TGGACACATTC-AAATGC3′). Genomic DNA was prepared from *Leavenworthia* leaves of individual plants by a plant miniprep method using the CTAB method (Rogers & Bendich 1985) and amplified by polymerase chain reaction (PCR) as described by Charlesworth *et al.* (1998). The products of the reactions were run on 1% agarose gels with 1 ×TAE buffer and stained with ethidium bromide. Three bands were seen with all *Leavenworthia* species, and shown to correspond to three different loci (Charlesworth *et al.* 1998). The bands from the individual loci were cloned and sequences of both strands of the PCR products were obtained. Here, we present results for the *Adh-1* locus, which has six introns in the same positions in the coding sequence as in the *Adh* locus of *A. thaliana* and codes for 379 amino acids, the same as in *A. thaliana* (Chang & Meyerowitz 1986); Charlesworth *et al.* 1998).

To survey variation at loci of interest, the sequence information obtained for the *Adh-1* locus was used to design further primers for this locus (described in more detail in Charlesworth *et al.* 1998). The *Adh* loci from the three *Leavenworthia* species were

Table 1. *Breeding systems of plants from* Leavenworthia *populations*

population number	self-compatability	autogamous fruit set in the greenhouse and 95% confidence interval	estimated selfing rate
L. stylosa			
95007	self-incompatible	very low	≈ 0
Hem 1[a]	self-incompatible	very low	≈ 0
9113[b]	self-incompatible	very low	≈ 0
L. uniflora			
9108[b]	self-compatible	not tested	high
95008	self-compatible	0.663 ± 0.036	high
95011[c]	self-compatible	0.947 ± 0.026	high
L. crassa			
95014[c]	self-incompatible	not tested[d]	low
8919, 8707[b]	self-incompatible	not tested[d]	0
95005	some plants self-compatible	0.315 ± 0.043	intermediate
95010	some plants self-compatible	0.321 ± 0.074	intermediate
8921[b]	self-compatible	not tested[d]	0.6
95002	self-compatible	0.363 ± 0.145	high
95003	self-compatible	0.546 ± 0.103	high
95004	self-compatible	0.915 ± 0.053	high
9107[b]	self-compatible	not tested[d]	0.95

[a]Seeds supplied by Dr T. E. Hemmerly.
[b]Seeds supplied by Dr E. E. Lyons, together with selfing rates estimates, using allozyme data from progeny arrays of individual maternal plants.
[c]Seeds supplied by Dr G. Hilton.
[d]Seeds did not germinate, so plants could not be tested.

aligned using the Clustal W multiple alignment program, after manually adding gaps in the position of introns missing from the *Adh-2* and *Adh-3* loci. Using primers specific for *Adh-1*, PCR amplification products of portions of the locus, mainly from a large coding sequence, exon four, were used to analyse genomic DNA samples from different individuals from the populations. For most populations studied, the genomic DNA used for these analyses was extracted from leaves of growing plants, but for self-incompatible *L. crassa*, we used DNA extracted from seeds, because plants failed to grow from seed.

First, 'cold SSCP' analysis was done. This single-strand conformation polymorphism detection method is expected to be capable of detecting single base differences in PCR products up to about 350 bases (Hongyo *et al.* 1993). After allelic types were classified in this manner, both strands of two or more alleles of each SSCP phenotype were sequenced by cycle sequencing on an ABI sequencer, using either the same primers or a pair of primers that yielded a longer sequence. Since the true level of diversity is underestimated if one assumes that each SSCP type represents an unique sequence, it is necessary to evaluate the magnitude of this effect. For this locus, very good agreement was found between the two methods (table 2), enabling us to include alleles classified according to their SSCP patterns in our calculations of diversity measures without significant underestimation. In addition, SSCP analysis can detect heterozygosity at the locus; the two alleles of heterozygotes were cloned for sequencing separately.

(d) *Estimation of sequence diversity*

The sequences obtained were aligned after removal of the primer sequences. Sequences were aligned using SeqEd version 1.03 followed by manual adjustment for regions with indels to minimize the number of substitutions or indels. Using a Fortran program written for this purpose, we calculated for each sample of alleles the per base estimates of silent nucleotide diversity, π (see Nei 1987), and mean number of segregating sites, *Sn*, which was used to estimate the scaled neutral mutation rate $\theta = 4N_e\mu$ (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. This underestimates the diversity per base, which is preferable to the alternative of reducing the number of bases by the length of each such region, which would overestimate diversity. Calculations 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 sub-populations was measured as $\pi_T - \pi_S$ rather than F_{st} as it is less affected by total species diversity (π_T) which appears in the denominator when calculating F_{st} (Charlesworth *et al.* 1997). Tests for neutrality were calculated according to Tajima (1989).

3. RESULTS

Table 3 and figures 1 and 2 summarize the sequence diversity results for a region from exon four to intron five (total of 665 bases) in the *Adh-1* locus, analysed separately for silent and intron sites (114 and about 150 sites, respectively). The diversity values in outcrossing populations of *L. stylosa* are remarkably high compared with those found for *Drosophila* loci (reviewed by Moriyama & Powell 1995), for which the mean π value for silent sites averages 0.0135 in *D. melanogaster* (based on 15 autosomal loci) and 0.0420 in *D. pseudoobscura* (five loci). These are mostly species-wide diversity values, and within-population values are generally similar, when data are available to make this comparison (Moriyama & Powell 1995). In the outcrosser *L. stylosa*, π_S values are almost as high as the diversity

Table 2. *Comparisons of SSCP phenotypes and sequences of alleles from three* Leavenworthia *species*

(The table shows only those SSCP phenotypes for which two or more sequences were obtained.)

species and population	SSCP phenotype	number of alleles sequenced	number of sequences		number of base differences between the different sequences
			440 bases sequenced	665 bases sequenced	
L. stylosa					
9113	1	3	2	2	3
9113	2	4	1	1	0
L. uniflora					
95008	1	5	1	1	0
95011, 9108	2	8	1	1	0
L. crassa					
95014	5	2	1	1	0
95005, 9107	3	8	1	1	0
95010	4	3	1	1	0
8921, 8919, 95002, 95003, 95005	1	16	1	1	0
95004	2	5	1	1	0

Table 3. *DNA sequence diversity results, and Tajima's test results*

(The sequence variability estimates shown are θ, derived from the number of segregating sites per site, and the diversity (the probability of a base difference per site π). Tajima's test results are expressed as standardized values, for all populations in which there was any diversity. All values were non-significant.)

population	selfing rate	no. alleles studied	silent sites				introns			
			seg. sites/ total sites	θ	diversity π	Tajima's D	seg. sites/ total sites	θ	diversity π	Tajima's D
L. stylosa										
95007	0	11	9/115	0.0267	0.0256	−0.055	14/149	0.0321	0.0432	0.503
Hem 1	0	7	13/115	0.0461	0.0381	−0.370	20/145	0.0563	0.0581	0.076
9113	0	9	8/116	0.0254	0.0283	−0.183	13/145	0.0330	0.0330	−0.002
all populations		27	16/115	0.0361	0.0307	−0.194	30/149	0.0522	0.0552	0.088
L. uniflora										
9108	≈ 1	5	0/115	0	0	—	0/142	0	0	—
95008	≈ 1	12	0/115	0	0	—	0/146	0	0	—
95011	≈ 1	10	0/115	0	0	—	0/142	0	0	—
all populations		27	8/115	0.0181	0.0357	0.891	6/146	0.0107	0.0211	0.686
L. crassa										
95014	≈ 0	5	4/116	0.0166	0.0207	0.425	7/145	0.0232	0.0290	0.502
8919	0	6	5/116	0.0189	0.0144	−0.364	8/145	0.0242	0.0184	−0.408
all SI populations		11	5/116	0.0147	0.0191	0.295	8/145	0.0188	0.0283	0.562
95005 intermediate		14	4/116	0.0108	0.0152	0.315	8/145	0.0174	0.0222	0.280
8921	0.6	9	0/116	0	0	—	0/136	0	0	—
95010 intermediate		6	0/116	0	0	—	0/136	0	0	—
all intermediate populations		29	4/116	0.0088	0.0085	−0.0207	8/145	0.0141	0.0124	−0.0957
95002	high	4	0/116	0	0	—	0/136	0	0	—
95003	high	5	0/116	0	0	—	0/136	0	0	—
95004	high	5	0/116	0	0	—	0/140	0	0	—
9107	0.95	5	0/116	0	0	—	0/145	0	0	—
all selfing populations		19	8/116	0.0197	0.0282	0.431	12/149	0.0230	0.0379	0.697

estimated for *D. pseudoobscura*, with a mean of 0.0421 for intron sites and 0.0277 for silent sites in the coding regions of the locus. As in *Drosophila*, diversity for intron sites is consistently higher than for silent sites in the coding regions of the gene, but the variances are equal to roughly half the means, so these differences, though repeatable, may not be statistically significant. Whenever it was possible to compute Tajima's test, the results were non-significant.

As expected from the theory outlined above, we find that sequence diversity within outcrossing populations is higher than in the inbred populations that are most closely related to them, for both silent sites and intron

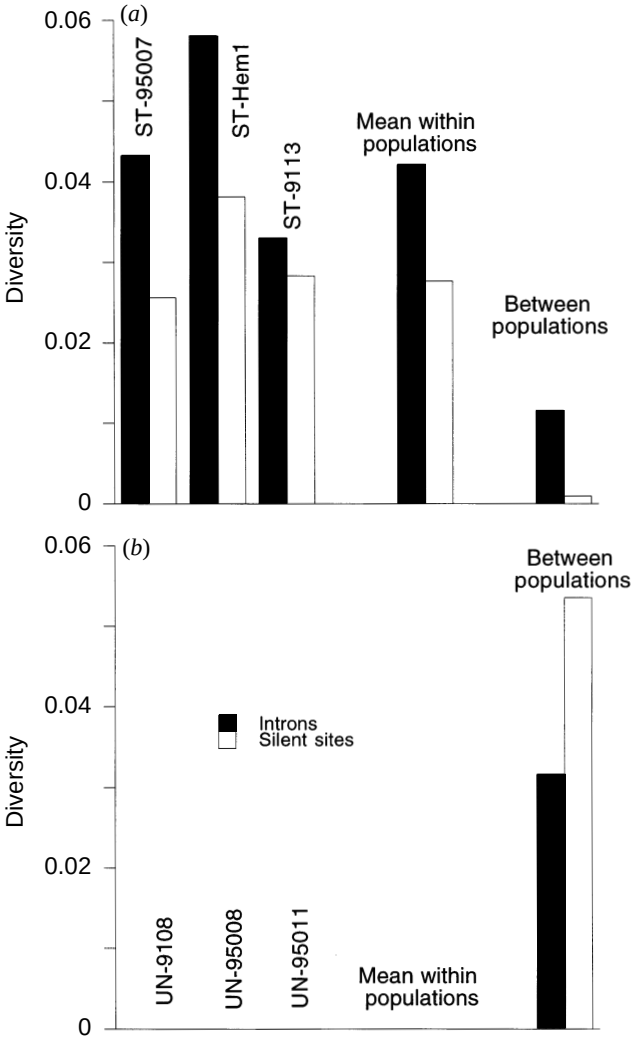

Figure 1. Comparison of the sequence diversity within and between populations of (*a*) *L. stylosa* and (*b*) *L. uniflora* for silent and intron sites in a portion of the *Adh-1* locus.

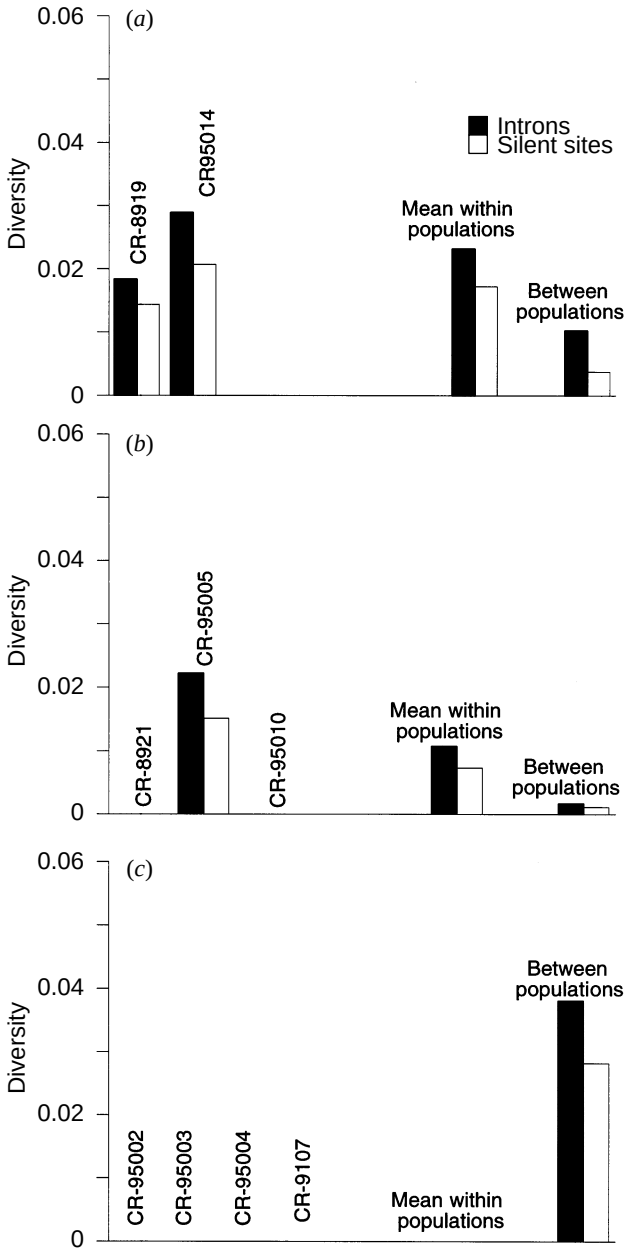

Figure 2. Comparison of the sequence diversity within and between populations of *L. crassa* with different selfing rates. (*a*) Self-incompatible populations; (*b*) populations with intermediate selfing rates; (*c*) highly selfing populations.

positions. In *L. crassa*, all samples from inbreeding populations have very low diversity, as do two of the three samples from populations with intermediate selfing rates. The standard errors on these diversity values are high (Tajima 1993) and 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, the differences are consistent across different populations in both independent comparisons (between *L. stylosa* and *uniflora*, and between self-incompatible and self-compatible *L. crassa* populations).

In the outcrossing taxa, diversity between populations is similar to that within populations, while for the inbreeding populations diversity between populations greatly exceeds the within-population values; the allele sequences differ between inbreeding populations but, for most such populations, no within-population diversity was found in the samples of alleles studied. Thus, $F_{st}=1$ for *L. uniflora* and for the set of inbreeding *L. crassa* populations, compared with 0.235 and 0.087 for intron and silent variants, repectively, in *L. stylosa*, and 0.181 and 0.098 for outcrossing *L. crassa* populations.

In addition to the presumably neutral changes just described, several amino acid polymorphisms were found in each species, but only in *L. stylosa* did we find any such variants within populations; in the other species, they were found only as fixed differences between populations. The polymorphisms in *L. stylosa* are listed in table 4, and it can be seen that in our sequences a charge change at position 68 (histidine/asparagine) is polymorphic in all three populations. There were a number of other charge changes among the sequences, several of which were polymorphic in one or two of the three *L. stylosa* populations (table 4).

Table 4. *Amino acid polymorphisms within and between three* Leavenworthia stylosa *populations*

(The presence (✓) or absence (−) of each polymorphism within each population is tabulated in the right hand three columns.)

position	amino acids	charge change	population		
			95007	Hem 1	9113
10	ED	no	−	−	−
55	KN	yes	✓	−	−
68	HN	1/2	✓	✓	✓
141	PR	yes	−	✓	−
142	PR	yes	−	−	✓
445	KN	yes	−	✓	−
494	VL	no	−	−	✓
514	ED	no	✓	−	−
540	SN	no	−	✓	−
545	EK	yes	−	−	✓
547	EK	yes	−	−	✓

4. CONCLUSIONS AND DISCUSSION

In this first set of DNA sequence diversity data from the genus *Leavenworthia*, the expected pattern is apparent of low DNA sequence diversity within inbreeding populations, but with some diversity between populations. Despite the lack of a statistical test for diversity differences, the consistent differences between inbreeding and outcrossing populations are striking and apparently robust and repeatable.

The effect we have found considerably exceeds a twofold difference between the mean diversity measures within outcrossing and inbreeding populations. This large effect argues for an effect of inbreeding on neutral diversity additional to the direct effect of homozygosity on effective population size (see above). It is unlikely that the effect is caused by lower population size of the inbreeding taxa, as it is most extreme in the comparison between *L. stylosa* and the much more widely distributed *L. uniflora*, which must have a considerably larger total population. Nor are inbreeding populations of *L. crassa* smaller than the self-incompatible populations (though these may of course have been larger in the past). The finding that inbreeding populations have much lower neutral diversity than outbreeders is thus consistent with a possible role for background selection. The plausibility of this hypothesis depends largely on deleterious mutation rates. Current evidence, mostly from mutation accumulation experiments in *Drosophila*, suggests rates sufficient to cause large background selection effects (e.g. Crow 1993; Simmons & Crow 1977). Estimates of genomic mutation rates from inbreeding depression in six highly inbreeding plant species, including *Leavenworthia* (Ågren & Schemske 1993; Charlesworth *et al.* 1994; Johnston & Schoen 1995), are consistent with those from *Drosophila*. Since such populations probably harbour few, if any, polymorphisms maintained by overdominance (because the maintenance of such polymorphisms requires highly implausible symmetrical selection coefficients, see Kimura & Ohta 1971), these estimates are free from complications from overdominant loci (Charlesworth *et al.* 1990).

The background selection hypothesis is, however, only one possibility (see above). Can we rule out any of the others? Our finding of between-population diversity in inbreeding populations rules out a simple selective sweep model, in which unconditionally advantageous alleles have become fixed in species. Such events would eliminate variability both between and within populations. Recent selective sweeps due to local selection could lead to diversity between, but not within, inbred populations, but it seems implausible that such events could be frequent enough. The pattern of diversity observed could equally well be caused by local selection maintaining allele frequency differences at selected loci, together with background selection (Charlesworth *et al.* 1997).

In cases such as *L. crassa*, where some populations have lost self-incompatibility and evolved selfing, spread of alleles causing selfing must have occurred. Yet in *L. crassa* there is considerable allozyme diversity within some of the inbreeding populations (D. Charlesworth & Z. Yang, unpublished data), and sequence diversity between them. The allozyme diversity in inbred populations may be explicable in terms of recent evolution of selfing, since equilibrium levels of genetic diversity will not be reached until many generations after a change in breeding system. It does, however, contrast with the complete absence of sequence diversity at the *Adh-1* locus in all the inbreeding populations. As in the case of differences between variability estimates from different kinds of markers (Karl & Avise 1992; Mitton & Latta 1997; Pogson & Zouros 1994; Raybould *et al.* 1996), it is therefore possible that some form of balancing selection maintains allozyme diversity. It will therefore be important to study further loci to see whether discordance between allozyme and DNA sequence diversity is general in comparisons of in- and outbreeding populations.

If selective sweeps have been important in causing the low diversity we have observed in inbreeding populations, their effect should be detectable in sequence data from populations. As diversity recovers due to new neutral mutations after such an event, there will be an excess of rare variants. One test based on this is Tajima's test (Tajima 1989). Given some diversity, this has good power to detect recent selective sweeps (Braverman *et al.* 1995; Simonsen *et al.* 1995), whereas background selection does not skew the frequency distribution of variant sites to the same extent and rarely produces significant results (Charlesworth *et al.* 1995). It should therefore be possible to distinguish between selective sweeps and background selection, provided that variants are detected. The total absence of within-population diversity in most of the inbred populations we have studied prevents our using such tests, so we cannot test this possibility without obtaining a greater length of sequence. We have suggested above that *L. crassa* must have undergone such an event, but the one intermediate selfing *L. crassa* population that contained some diversity yielded no significant Tajima's test results, so there is no evidence for a recent selective sweep.

It may be difficult to distinguish between background selection and population bottlenecks. If bottlenecks were the sole cause of low diversity, only those populations affected will have low variability. Levels of allozyme diversity (and N_e values estimated from them) vary more from

one population to another in inbreeders than in outcrossing taxa, consistent with the effects of bottlenecks having occurred in the ancestry of the populations that have low diversity values (Schoen & Brown 1991). However, this is not conclusive evidence for bottlenecks, as there are several possible reasons for variable diversity in inbreeders. Inbreeders may have low average within-population diversity, due to selective sweeps or background selection, with occasional populations having high diversity (e.g. Bonnin *et al*. 1996) because of population mixture, local adaptation in populations between which subdivision is not apparent (Charlesworth *et al*. 1997), or recent evolution of selfing (see above). Since we find no sequence diversity in any of the three *L. uniflora* populations studied, nor in five inbreeding populations of *L. crassa*, our results do not suggest occasional bottlenecks for either of these species. Repeated extinction and recolonization events could probably reduce diversity more generally, even in populations of outcrossers (Kimura & Maruyama 1971; Slatkin 1977; Wade & McCauley 1988; Whitlock 1993). More theoretical work is needed, particularly to incorporate the effect of breeding system differences, but in inbreeders extinction and recolonization probably reduces species-wide diversity (Ingvarsson 1997), whereas background selection reduces π_T only slightly (Charlesworth *et al*. 1997). Unless direct evidence for bottlenecks is obtained, or the alternative interpretations are ruled out, the issue cannot be resolved.

To know whether the pattern we have found is general, it will be important to have data from several further loci. The estimates of genetic diversity from just a single region are subject to considerable error (Tajima 1993). There is also the possibility that diversity at any given locus might be affected by balancing selection at a linked locus. Loci maintained polymorphic for long times are expected to have higher neutral diversity at closely linked sites (Hudson & Kaplan 1988; Takano *et al*. 1993), but this should not extend to loosely linked loci, even in highly inbreeding populations (Charlesworth *et al*. 1997; Nordborg *et al*. 1996). If balancing selection is important at loci in the genomes studied, it could obscure the expected pattern of lower values of diversity measures in inbreeding populations, but this is not likely to be a problem in our study, since a large difference between inbreeding and outcrossing populations was detected, in the expected direction. The very high diversity in the outcrossing populations might, however, be explicable partly in this way, as alcohol dehydrogenase shows allozyme polymorphisms in many species and in some there is evidence that it is selectively maintained (Kreitman 1991). We have not examined the populations for allozyme variants of alcohol dehydrogenase, because this enzyme is expressed at low levels in plant tissue under unstressed conditions (Dolferus *et al*. 1997), and we were unable to obtain strong bands on gels, suing extracts from *Leavenworthia* leaves or flowers under several conditions tested. As we find an amino acid polymorphism involving a charge change within all three populations of *L. stylosa*, this is a possibility. There is, however, no sign that the other polymorphic sites in the locus cluster in the vicinity of this difference, as has been found in *Drosophila* (Hudson *et al*. 1987; Takano *et al*. 1993). Finally, any given locus studied might be located in a region of low recombination, and therefore be subject to forces lowering its diversity (see above). To see if patterns are consistent, we have cloned four further loci from *Leavenworthia*, and preliminary diversity data from one (*PgiC*) are consistent with the results presented here.

We thank Dr T. E. Hemmerly, Dr E. E. Lyons and Dr G. Hilton for seed collections, M. Kreitman and D. S. Guttman for assistance with molecular methods, Zhe Yang for laboratory help, the staff of the University of Chicago greenhouses for growing plants, and B. Charlesworth, N. H. Barton, P. D. Keightley and P. Awadalla for discussions. This work was supported by grants DEB-9317683 from the US National Science Foundation and GM50355 from the National Institutes of Health.

REFERENCES

Ågren, J. & Schemske, D. F. 1993 Outcrossing and inbreeding depression in two annual monoecious herbs *Begonia hirsuta* and *B. semiovata*. *Evolution* **47**, 125–135.

Aguadé, M., Miyashita, N. & Langley, C. H. 1989 Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. *Genetics* **122**, 607–615.

Awadalla, P. & Ritland, K. 1997 Microsatellite variation and evolution in the *Mimulus guttatus* species of contrasting mating systems. *Molec. Biol. Evol.* **14**, 1023–1034.

Baker, H. G. 1955 Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**, 347–348.

Banks, J. A. & Birky, C. W. 1985 Chloroplast DNA diversity is low in a wild plant, *Lupinus texensis*. *Proc. Natn. Acad. Sci. USA* **82**, 6950–6954.

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

Bonnin, I., Huguet, T., Gherardi, M., Prosperi, J.-M. & Olivieri, I. 1996 High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *Am. J. Bot.* **83**, 843–855.

Braverman, J. M., Hudson, R. N., Kaplan, N. L., Langley, C. H. & Stephan, W. 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796.

Chang, C. & Meyerowitz, E. 1986 Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natn. Acad. Sci. USA* **83**, 1408–1412.

Charlesworth, B., Charlesworth, D. & Morgan, M. T. 1990 Genetic loads and estimates of mutation rates in very inbred plant populations. *Nature* **347**, 380–382.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

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

Charlesworth, D., Charlesworth, B. & Morgan, M. T. 1995 The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632.

Charlesworth, D., Lyons, E. E. & Litchfield, L. B. 1994 Inbreeding depression in two highly inbreeding populations of *Leavenworthia*. *Proc. R. Soc. Lond.* B **258**, 209–214.

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

Clegg, M. T. 1990 Molecular diversity in plant populations. In *Plant population genetics, breeding and genetic resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir), pp. 98–1156. Sunderland, Mass.: Sinauer.

Crow, J. F. 1993 Mutation, mean fitness, and genetic load. *Oxf. Surv. Evol. Biol.* **9**, 3–42.

DiRienzo, A. A., Peterson, A. C., Garza, J. C., Valdes, A. M. & Slatkin, M. 1994 Mutational processes of simple sequence repeat loci in human populations. *Proc. Natn. Acad. Sci. USA* **91**, 3166–3170.

Dolferus, R., Ellis, M., Bruxelles, G. d., Trevaskis, B., Hoeren, F., Dennis, E. S. & Peackock, W. J. 1997 Strategies of gene action in *Arabidopsis* during hypoxia. *Ann. Bot.* (Suppl. A) **79** 21–31.

Dumolin, S., Demesure, B. & Petit, R. J. 1995 Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoret. Appl. Genet.* **91**, 1253–1256.

Estoup, A., Garnery, L., Solignac, M. & Cornuet, J. M. 1995 Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and a test of the infinite allele and stepwise mutation model. *Genetics* **140**, 679–695.

Fenster, C. B. & Ritland, K. 1992 Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *Am. J. Bot.* **79**, 1440–1447.

Hamrick, J. L. & Godt, M. J. W. 1996 Effects of life history traits on genetic diversity in plant species. *Phil. Trans. R. Soc. Lond.* B **351**, 1291–1298.

Hanfstingl, U., Berry, A., Kellogg, E. A., Costa, J. T., Rüdiger, W. & Ausubel, F. M. 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.

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

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

Hudson, R. R. & Kaplan, N. L. 1988 The coalescent process in models with selection and recombination. *Genetics* **120**, 831–840.

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

Hudson, R. R., Bailey, K., Skerecky, D., Kwiatowski, J. & Ayala, F. J. 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340.

Ingvarsson, P. K. 1997 The effect of delayed population growth on the genetic differentiation of local populations subject to frequent extinctions and recolonizations. *Evolution* **51**, 29–35.

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

Jarne, P. & Staedler, T. 1995 Population genetic structure and mating system in freshwater pulmonates. *Experientia* **51**, 482–497.

Johnston, M. O. & Schoen, D. J. 1995 Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science* **267**, 226–229.

Karl, S. A. & Avise, J. C. 1992 Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* **256**, 100–102.

Kimura, M. 1983 *The neutral theory of molecular evolution*. Cambridge University Press.

Kimura, M. & Maruyama, T. 1971 Pattern of neutral polymorphism in a geographically structured population. *Genet. Res. Camb.* **18**, 125–131.

Kimura, M. & Ohta, T. 1971 *Theoretical topics in population genetics*. Princeton University Press.

Kreitman, M. 1991 Detecting selection at the level of DNA. In *Evolution at the molecular level* (ed. R. K. Selander, A. G. Clark & T. S. Whitham), pp. 204–221. Sunderland, Mass.: Sinauer.

Kreitman, M. 1996 The neutral theory is dead. Long live the neutral theory. *BioEssays* **18**, 678–683.

Langley, C. H. 1990 The molecular genetics of *Drosophila*. In *Population biology of genes and molecules* (ed. N. Takahata & J. F. Crow), pp. 75–91. Tokyo: Baifukan.

Lavin, M., Matthews, S. & Hughes, C. 1991 Chloroplast DNA variation in *Gliricidia sepium* (Leguminosae): intraspecific phylogeny and tokogeny. *Amer. J. Bot.* **78**, 1576–1585.

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

Lyons, E. E. & Antonovics, J. 1991 Breeding system evolution in *Leavenworthia*: breeding system variation and reproductive success in natural populations of *Leavenworthia crassa* (Cruciferae). *Amer. J. Bot.* **78**, 270–287.

Marshall, D. R. & Brown, A. H. D. 1975 The charge state model of protein polymorphisms in natural populations. *J. Molec. Evol.* **6**, 149–163.

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

Maynard Smith, J. & Haigh, J. 1974 The hitch-hiking effect of a favorable gene. *Genet. Res. Camb.* **219**, 1114–1116.

McCauley, D. E. 1994 Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proc. Natn. Acad. Sci. USA* **91**, 8127–8131.

Michalakis, Y. & Veuille, M. 1996 Length variation of CAG/CAA trinucleotide repeats in natural populations of *Drosophila melanogatser* and its relation to the recombination rate. *Genetics* **143**, 1713–1725.

Miller, J. C. & Tanksley, S. D. 1990 RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theoret. Appl. Genet.* **80**, 437–448.

Milligan, B. G. 1991 Chloroplast DNA diversity within and among populations of *Trifolium pratense*. *Curr. Genet.* **19**, 411–416.

Mitton, J. B. & Latta, R. 1997 A comparison of population differentiation across four classes of marker in limber pine (*Pinus flexuosa* James). *Genetics* **146**, 1153–1163.

Miyashita, N. T., Innan, H. & Terauchi, R. 1996 Intra- and interspecific variation of the alcohiol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. *Mol. Biol. Evol.* **13**, 433–436.

Moriyama, E. N. & Powell, J. R. 1995 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**, 261–277.

Narain, P. 1966 Effect of linkage on homozygosity of a population under mixed selfing and random mating. *Genetics* **54**, 303–314.

Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.

Nordborg, M., Charlesworth, B. & Charlesworth, D. 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc. R. Soc. Lond.* B **163**, 1033–1039.

Pogson, G. H. & Zouros, E. 1994 Allozyme and RFLP heterozygosities as corrrelates of growth rate in the scallop *Placopecten magellanicus*: a test of the associative overdominance hypothesis disequilibrium. *Genetics* **137**, 221–231.

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

Quarterman, E. 1950 Major plant communities of Tennessee cedar glades. *J. Ecol.* **31**, 234–254.

Raybould, A. F., Mogg, R. J. & Clarke, R. T. 1996 The genetic structure of *Beta vulgaris* ssp. *maritima* (sea beet) populations: RFLPs and isozymes show different patterns of gene flow. *Heredity* **77**, 245–250.

Rogers, S. O. & Bendich, A. J. 1985 Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Pl. Molec. Biol.* **5**, 69–76.

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

Schaal, B. A., O'Kane, S. L. & Rogstad, S. H. 1991 DNA variation in plant populations. *Trends Ecol. Evol.* **6**, 329–333.

Schoen, D. J. & Brown, A. H. D. 1991 Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natn. Acad. Sci. USA* **88**, 4494–4497.

Simmons, M. J. & Crow, J. F. 1977 Mutations affecting fitness in *Drosophila* populations. *A. Rev. Genet.* **11**, 49–78.

Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. 1995 Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Slatkin, M. 1977 Gene flow and genetic drift in a species subject to local extinctions. *Theoret. Popul. Biol.* **12**, 253–262.

Solbrig, O. T. & Rollins, R. C. 1977 The evolution of autogamy in species of the mustard genus *Leavenworthia*. *Evolution* **31**, 265–281.

Soltis, D. E., Soltis, P. S. & Ness, B. D. 1989 Chloroplast DNA variation and mutliple origins of autopolyploidy in *Heuchera micrantha* (Saxifagaceae). *Evolution* **43**, 650–656.

Stephan, W. & Langley, C. H. 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the vermilion and forked loci. *Genetics* **121**, 89–99.

Tajima, F. 1989 Statistical method for testing the neutral mutation hypothesis. *Genetics* **123**, 585–595.

Tajima, F. 1993 Measurement of DNA polymorphism. In *Mechanisms of molecular evolution* (ed. N. Takahata & A. G. Clark), pp. 37–59. Sunderland, MA: Sinauer.

Takano, T. S., Kukasabe, S. & Mukai, T. 1993 DNA polymorphism and the origin of protein polymorphism at the *Gpdh* locus of *Drosophila melanogaster*. In *Mechanisms of molecular evolution* (ed. N. Takahata & A. G. Clark), pp. 37–59. Sunderland, MA: Sinauer.

Todokoro, S., Terauchi, R. & Kawano, S. 1995 Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. *Japanese J. Genetics* **70**, 543–554.

Wade, M. J. & McCauley, D. E. 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**, 995–1005.

Wakeley, J. 1996 The variance of pairwise nucleotide differences in two populations with migration. *Theoret. Popul. Biol.* **49**, 39–57.

Whitlock, M. 1993 Lack of correlation between heterozygosity and fitness in forked fungus beetles. *Heredity* **70**, 574–581.

Wolff, K., Rogstad, S. H. & Schaal, B. A. 1994 Population and species variation of minisatellite DNA in *Plantago*. *Theoret. Appl. Genet.* **87**, 733–740.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute towards production costs.