Species and Recombination Effects on DNA Variability in the Tomato Genus

Emmanuelle Baudry,* Carole Kerdelhué,† Hideki Innan* and Wolfgang Stephan‡

**Department of Biology, University of Rochester, Rochester, New York 14627, †Laboratoire de Zoologie Forestière, INRA Centre d'Orléans, 45166 Olivet cedex, France and ‡Department of Evolutionary Biology, University of Munich, 80333 Munich, Germany*

Manuscript received December 19, 2000
Accepted for publication May 11, 2001

ABSTRACT

Population genetics theory predicts that strong selection for rare, beneficial mutations or against frequent, deleterious mutations reduces polymorphism at linked neutral (or weakly selected) sites. The reduction of genetic variation is expected to be more severe when recombination rates are lower. In outbreeding species, low recombination rates are usually confined to certain chromosomal regions, such as centromeres and telomeres. In contrast, in predominantly selfing species, the rarity of double heterozygotes leads to a reduced effective recombination rate in the whole genome. We investigated the effects of restricted recombination on DNA polymorphism in these two cases, analyzing five Lycopersicon species with contrasting mating systems: *L. chilense*, *L. hirsutum*, *L. peruvianum*, *L. chmielewskii*, and *L. pimpinellifolium*, of which only the first three species have self-incompatibility alleles. In each species, we determined DNA sequence variation of five single-copy genes located in chromosomal regions with either high or low recombination rate. We found that the mating system has a highly significant effect on the level of polymorphism, whereas recombination has only a weak influence. The effect of recombination on levels of polymorphism in Lycopersicon is much weaker than in other well-studied species, including Drosophila. To explain these observations, we discuss a number of hypotheses, invoking selection, recombination, and demographic factors associated with the mating system. We also provide evidence that *L. peruvianum*, showing a level of polymorphism (almost 3%) that is comparable to the level of divergence in the whole genus, is the ancestral species from which the other species of the genus Lycopersicon have originated relatively recently.

THE neutral theory predicts a positive correlation between the levels of intraspecific nucleotidic variation and the amounts of interspecific divergence between closely related species. However, in natural populations of Drosophila, genes located in chromosomal regions with low recombination rates were shown to have reduced levels of DNA polymorphism whereas the amount of divergence between species is roughly independent of recombination rates (Aguadé *et al.* 1989; Stephan and Langley 1989; Begun and Aquadro 1992). Similar results have since been obtained for diverse other species, namely mouse (Nachman 1997), human (Nachman *et al.* 1998), tomato (Stephan and Langley 1998), wheat (Dvorak *et al.* 1998), and sea beets (Kraft *et al.* 1998). These patterns have been explained by two diametrically opposed population genetic models that invoke selection and linkage: the selective sweep model (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992) and the background selection model (Charlesworth *et al.* 1993; Hudson and Kaplan 1995; Charlesworth 1996). The first model assumes the hitchhiking of neutral (or nearly neutral) variants on chromosomes bearing rare, strongly selected, favorable mutations at closely linked sites that go rapidly to fixation. The second model involves the loss of neutral (or nearly neutral) variants as a result of steady elimination of linked deleterious mutations from the population. For both models, the reduction in genetic variation at linked neutral sites is stronger in genomic regions where recombination is restricted.

In outbreeding species, reduced recombination rates are observed in certain regions of the genome, especially around centromeres. On the contrary, in species with a high level of inbreeding, the rarity of double heterozygotes results in lowered effective recombination rates in the whole genome. It is thus expected that both hitchhiking and background selection will strongly affect genetic variability in inbreeding species.

Our main goal is to investigate the influence of selective sweeps and background selection on DNA polymorphism both in genomic regions with low recombination rates and in inbreeding species. We therefore compared the levels of DNA polymorphism between closely related inbreeding and outbreeding species and, within each species, between genes located in high and low recombination regions of the genome. We performed these com-

Corresponding author: Wolfgang Stephan, Department of Evolutionary Biology, University of Munich, Luisenstr. 14, 80333 Munich, Germany. E-mail: stephan@zi.biologie.uni-muenchen.de

parisons in the genus Lycopersicon on the basis of DNA sequence data. An earlier attempt using Lycopersicon species was made by Stephan and Langley (1998), but their analysis is based on data on shared restriction fragments (Nei 1987, chapter 5). DNA sequence data are more reliable and informative for at least two reasons: (i) They allow us to estimate nucleotide polymorphism more directly, and (ii) it is possible to distinguish between synonymous and nonsynonymous substitutions, making this approach more effective in detecting effects of selection. Sequence data are currently scarce in plants, because only a few studies used DNA sequences to investigate the effects of the mating system (Miyashita *et al.* 1996; Liu *et al.* 1998, 1999; Savolainen *et al.* 2000) and no sequence data have thus far been published to test the effect of recombination.

The genus Lycopersicon consists of nine species, originating in Central and South America. It has several characteristics that make it well suited for our purpose:

1. Despite the small number of species, the genus presents a great diversity of mating systems. Some species are self-compatible (thereafter called SC species) and are obligatorily or facultatively autogamous, while other species have a self-incompatible locus (SI species) and are obligatorily allogamous.
2. Detailed genetic maps of the tomato genome are available (Tanksley *et al.* 1992; Pillen *et al.* 1996; Fulton *et al.* 1997).
3. There are large differences among chromosomal regions in the level of recombination rates (Sherman and Stack 1995; Stephan and Langley 1998).

We chose two SC species with intermediate to high levels of inbreeding and three SI species that are obligate outcrossers. In each species, we analyzed DNA polymorphism and divergence at five genes located in chromosomal regions with either high or low recombination rates.

MATERIALS AND METHODS

Sampling: Five species that differ in their mating system were used in the present study, namely *L. chilense* (SI), *L. hirsutum* (SI), *L. peruvianum* (SI), *L. chmielewskii* (SC), and *L. pimpinellifolium* (SC). Among the accessions available in the Tomato Genetics Resource Center at UC Davis (CA), one population from the central part of each species' natural distribution was chosen. Five plants per species were sampled, except for *L. hirsutum* for which only three individuals were studied. Individuals were grown from seeds directly collected in the field from different fruiting plants and kindly provided by C. M. Rick. The populations of *L. chilense* and *L. peruvianum* are from Chile, namely Antofagasta (accession no. LA2884) and Tarapaca (LA2744), respectively. The other three species are from Peru. *L. chmielewskii* was collected in Apurimac (LA3653), *L. hirsutum* in Ancash (LA1775), and *L. pimpinellifolium* in Lambayeque (LA1583).

Estimation of recombination rate, DNA amplification, and sequencing: By aligning the linkage map from the cross of *L. esculentum* and *L. pennellii* (Pillen *et al.* 1996) and a quantitative cytogenetic map of the distribution of recombination nodules in *L. esculentum* (Sherman and Stack 1995), recombination rates in the tomato genome have been estimated (Stephan and Langley 1998). These estimates have been used for the five Lycopersicon species, as the karyotypes of the 12 chromosome pairs of each species are very similar, with little or no structural differences and apparently little difference in recombination rates (Rick 1983; Pillen *et al.* 1996). We chose five genes from regions with very different rates of recombination: three genes in centromeric regions where recombination rates are low (*sucr*, CT208, and CT251, thereafter called "low-recombination genes"), and two genes in chromosomal regions where recombination is high (CT143 and CT268, "high-recombination genes"). *sucr* is the sucrose accumulator gene described in Elliott *et al.* (1993). CT208, CT251, CT143, and CT268 are four anonymous, single-copy cDNA markers previously developed and mapped in Tanksley *et al.* (1992), for which partial or total sequences are available in the EMBL database (Ganal *et al.* 1998) and the Tomato Gene Index at the Institute for Genomic Research (http://www.tigr.org/tdb/lgi/). Recombination rates per site show a maximum value of 0.46×10^{-8} per nucleotide site per generation for the three genes located in centromeric regions, but they are greater than 2.33×10^{-8} for the two genes located in high-recombination regions. There is thus a more than fivefold difference in recombination rates between the two categories of genes.

Genomic DNA isolated from leaves of mature plants was kindly provided by L. Rose and C. Langley (UC Davis). PCR primers were designed using the published cDNA sequences. Diploid DNA was PCR amplified and sequenced on both strands with an ABI 377 automatic sequencer (Perkin-Elmer, Norwalk, CT), using primers spaced every ~500 bp. Each diploid sequencing trace was analyzed by procedures and software that can reliably resolve heterozygotes. Haplotype phases were then determined from the unphased genotype data according to the procedure described in Clark (1990), except in the *L. peruvianum* sample where a very high level of polymorphism (see RESULTS) makes the procedure unreliable. When the alleles from one individual were polymorphic for indels, fragments of the gene were cloned into a pCR2.1 vector using the TA cloning kit (Invitrogen, San Diego) and sequenced as described above. Up to 10 clones were sequenced in each case.

Estimation of levels of polymorphism and divergence: We used two measures of intraspecific polymorphism, namely θ_{syn} for synonymous sites in coding regions and θ_{sil} for silent sites, *i.e.*, introns and synonymous sites (Nei 1987). Indels were excluded from the analysis. θ was used instead of π as it has a smaller stochastic variance. Divergence between species was estimated by excluding sites polymorphic for indels within or between species. Interspecific divergence at synonymous and silent sites was estimated for each locus over two apparently independent evolutionary paths: *L. chilense-L. peruvianum* and *L. hirsutum-L. pimpinellifolium* (Miller and Tanksley 1990; Stephan and Langley 1998). Analyses were performed using the mean of these two estimates. Levels of nucleotide diversity and divergence were estimated using DnaSP v3.50 (Rozas and Rozas 1999). Statistical analyses were performed using Statistica software (StatSoft, Tulsa, OK) to determine the effects of mating system and recombination on intraspecific polymorphism.

Several tests have been developed to determine significant departures of sequence data from neutral evolution, including the standard Tajima's *D* and Fu and Li's *D* tests (Tajima 1989; Fu and Li 1993) and Fu's F_s (Fu 1997) test, which compares the observed number of haplotypes in a sample to the expected number under neutrality. We have applied these three tests to the tomato data (except that Fu's F_s statistic was not calculated in *L. peruvianum*, as the number of haplotypes could not

TABLE 1

Estimates of nucleotide diversity at silent (θ_{sil}) and synonymous sites (θ_{syn})

	Low-recombination genes[a] (%)			High-recombination genes[a] (%)	
	CT208	*sucr*	CT251	CT268	CT143
L. chmielewskii ($n = 10$)					
θ_{sil}	0.000 (1174.2)	0.000 (671.8)	0.000 (702.2)	0.000 (422.3)	0.048 (1478.0)
θ_{syn}	0.000 (168.2)	0.000 (267.8)	0.000 (319.2)	0.000 (422.3)	0.698 (101.3)
L. pimpinellifolium ($n = 10$)					
θ_{sil}	0.060 (1172.2)	0.103 (683.5)	0.000 (716.4)	0.168 (421.3)	0.120 (1478.0)
θ_{syn}	0.000 (168.2)	0.264 (267.5)	0.000 (320.4)	0.168 (421.3)	0.348 (101.5)
L. chilense ($n = 10$)					
θ_{sil}	0.119 (1193.2)	0.944 (662.5)	0.198 (715.3)	1.842 (422.3)	0.000 (1474.5)
θ_{syn}	0.210 (168.2)	0.688 (267.5)	0.221 (319.3)	1.842 (422.3)	0.000 (101.5)
L. hirsutum ($n = 6$)					
θ_{sil}	0.000 (1191.2)	0.407 (645.0)	0.441 (695.2)	1.036 (422.9)	0.149 (1474.5)
θ_{syn}	0.000 (168.2)	0.327 (268.0)	0.692 (316.2)	1.036 (422.9)	0.431 (101.5)
L. peruvianum ($n = 10$)					
θ_{sil}	1.291 (1177.8)	3.189 (587.4)	2.594 (667.7)	2.682 (421.8)	2.195 (1304.5)
θ_{syn}	0.843 (167.8)	3.040 (267.4)	3.582 (315.8)	2.682 (421.8)	2.090 (101.5)

The estimates of θ_{sil} and θ_{syn} are expressed in percentages. Parentheses indicate the number of sites sequenced. *n*, number of alleles sampled in the population.

[a] Estimates of recombination rates are given in Table 2.

be reliably estimated from our data in this species). In addition, two tests of neutrality that use both intra- and interspecific data were performed. The McDonald and Kreitmann (MK) test (McDonald and Kreitman 1991) compares synonymous and nonsynonymous variation within and between species. The Hudson-Kreitmann Aguadé (HKA) test (Hudson *et al.* 1987) compares the levels of intraspecific polymorphism to the level of divergence between species at two loci.

Test for gene flow between species: We used the method described in Wakeley and Hey (1997) and Wang *et al.* (1997) to test whether the pattern of polymorphism observed for the outcrossing Lycopersicon species differs significantly from the predictions of a model of speciation via isolation. This model assumes that an ancestral panmictic population splits into two descendant, randomly mating populations at a certain time point and that there is no gene flow between the descendant populations at later times. The procedures described in Wang *et al.* (1997) were performed using a program kindly provided by J. Hey. The parameters for population sizes and speciation time were estimated from the data assuming this simple isolation speciation model. A total of 1000 coalescent simulations were run using the estimated parameters. The population recombination parameter used in the simulations, $4N_e c$, was estimated for each gene and each species with the SITES program (Hey and Wakeley 1997), except for *L. peruvianum* for which haplotypes could not be determined (see above). Here *L. hirsutum* estimates were used, after correction for the difference in population size.

RESULTS

Silent polymorphism and divergence: Table 1 shows a summary of synonymous and silent (introns and synonymous sites) polymorphism in five Lycopersicon species for the three genes located in centromeric regions and the two genes located in regions of high recombination. An average total of 4996 bp was sequenced for the three genes in centromeric regions, representing between 2432.9 and 2571.0 silent sites (depending on species). In the two genes in high-recombination regions, an average total of 3619 bp was sequenced, representing between 1726.3 and 1900.3 silent sites. The observed values of θ_{sil} and θ_{syn} range from 0 for several genes in *L. chmielewskii* to >3% for *sucr* in *L. peruvianum*. Table 2 shows recombination rates and levels of divergence for the five loci studied. Divergence between species at silent sites ranges from 2.76% in CT143 to 3.64% in CT268. At synonymous sites, divergence values range from 2.02% in CT208 to 3.81% in CT143.

Most interestingly, nucleotide diversity in *L. peruvianum* is almost as high as divergence between species (Table 3). The average ratio of θ_{sil} in *L. peruvianum* to divergence between species at silent sites is 0.77, with values ranging from 0.46 at the CT208 locus to 1.19 at *sucr*. The average ratio of θ_{sil} to divergence in *L. chilense*, the second most polymorphic species, is only 0.24. Similar results are obtained when synonymous sites are con-

TABLE 2

Recombination rate and divergence

Gene	Recombination rate $\times 10^{-8}$	Divergence at silent sites (%)	Divergence at synonymous sites (%)
CT208	0.00	2.79	2.02
sucr	0.00	2.68	3.22
CT251	0.46	3.62	3.66
CT268	2.33	3.64	3.64
CT143	2.73	2.76	3.82

The recombination rate is given per nucleotide site per generation.

TABLE 3

Polymorphism observed in *L. peruvianum vs.* variation in the whole genus

Gene	$\theta_{sil\ peru}/D_{sil}$[a]	$\theta_{syn\ peru}/D_{syn}$[b]	Percentage of sites that show fixed differences between four species[c] and are polymorphic in *L. peruvianum*
CT208	0.46	0.42	14.1
sucr	1.19	0.95	45.5
CT251	0.72	0.98	44.0
CT268	0.74	0.74	35.3
CT143	0.80	0.55	33.3

[a] Ratio of θ_{sil} in *L. peruvianum* to divergence between species at silent sites.
[b] Ratio of θ_{syn} in *L. peruvianum* to divergence between species at synonymous sites.
[c] *L. chmielewskii*, *L. pimpinellifolium*, *L. hirsutum*, and *L. chilense*.

sidered or when divergence estimates between other species are used. The high value of polymorphism in *L. peruvianum* relative to divergence between species is due to two main factors. First, among the five Lycopersicon species, *L. peruvianum* is by far the most polymorphic species. The average level of polymorphism in *L. peruvianum* is almost four times higher than in *L. chilense*. Second, a high proportion of fixed differences between *L. chmielewskii*, *L. pimpinellifolium*, *L. hirsutum*, and *L. chilense* is also present as polymorphisms in *L. peruvianum* (Table 3). For four of the genes (*sucr*, CT251, CT268, and CT143), ~40% of the sites with fixed differences between these four species exhibit the same variants within *L. peruvianum*. This percentage is lower in CT208 (14.1%). Our observations therefore suggest that a high proportion of the variation found in the genus Lycopersicon originated in *L. peruvianum*.

To test whether there is gene flow between the extant outcrossing Lycopersicon species, we used the method of Wakeley and Hey (see MATERIALS AND METHODS). As *L. peruvianum* is presumably the ancestral species from which the other species derived, we applied the test to the pairs *L. chilense-L. peruvianum* and *L. hirsutum-L. peruvianum*. The estimates of effective population sizes and speciation times are shown in Table 4, as well as the tail probabilities of the test statistic. The isolation speciation model was rejected (at the 0.05 level) for none of these comparisons. Thus, our analyses revealed little evidence of gene flow between these species.

Effect of mating system on polymorphism: The effect of mating system and species on silent polymorphism is highly significant (Mann-Whitney U test, $P < 0.001$). The two SC species have drastically reduced levels of within-population polymorphism compared to the three SI species. This conclusion corroborates previous allozyme studies (RICK 1983; DOEBLEY 1989) and restriction fragment length polymorphism (RFLP) analyses (MILLER and TANKSLEY 1990; STEPHAN and LANGLEY 1998). According to the neutral theory, complete selfing is expected to lead to a twofold reduction in effective population size. However, the actual reduction observed in our data is more than fourfold, as average θ_{sil} is 0.0096% in *L. chmielewskii* and 0.090% in *L. pimpinellifolium* while the lowest value observed in the SI species (*L. hirsutum*) is 0.40%. Polymorphism is more reduced in *L. chmielewskii* than in *L. pimpinellifolium*. This may be due to a difference in outcrossing rates between the two species. This rate reaches 40% in the central range of *L. pimpinellifolium* (RICK *et al.* 1978b) while outcrossing is restricted in *L. chmielewskii* (RICK *et al.* 1976). In our *L. pimpinellifolium* sample, the stigma is clearly exerted (R. CHETELAT, personal communication). In SC species, this morphological characteristic shows that the degree of allogamy may be high (RICK *et al.* 1977).

Effect of recombination on polymorphism: Figure 1 presents a scatterplot of polymorphism *vs.* recombination rate in the five Lycopersicon species studied. In all five species, correlation coefficients between recombination rate and θ_{sil} or θ_{syn} are positive, but none of these correlations is significant. The correlation is weakest in *L. peruvianum*. For further analysis of the effect of recombination, θ_{syn} and θ_{sil} values were centered by species, because mating system and species act as confounding factors. Using this approach, we found a significant effect of recombination on θ_{sil} when we considered only the two SC species together ($P = 0.036$). However, no significant effect was detected when the three SI species or all the five species were considered together. The effect of recombination on θ_{syn} is marginally significant when all species except *L. peruvianum* are included ($P = 0.0505$). In each species, levels of polymorphism appear to be highly scattered among loci. One might expect that some of the scatter is due to differences in neutral mutation rate among loci. If so, a tighter correlation is expected when the ratio of polymorphism to divergence values is plotted. This is, however, not the case, suggesting that the unexplained variance is not due simply to differences in the neutral mutation rate among loci (further discussed in the next two sections).

Neutrality tests using intraspecific data: To understand the reasons for the heterogeneity in levels of polymorphism among loci, we performed several neutrality tests within and between species. Table 5 shows the results of Tajima's *D* (TAJIMA 1989), Fu and Li's *D* (FU

TABLE 4

Estimates of parameters of the isolation speciation model

Species 1	Species 2	θ_1	θ_2	θ_A	τ	T	P_{WWH}
chil	*peru*	6.14	38.92	130.18	15.14	0.39	0.46
hirs	*peru*	13.20	88.96	88.54	38.94	0.43	0.43

θ_1, θ_2, and θ_A are the population mutation rates of species 1, species 2, and the ancestral species, respectively. Species 1 is *L. chilense* (*chil*) or *L. hirsutum* (*hirs*); species 2 is *L. peruvianum* (*peru*). τ is the estimated time of the split between the two species scaled in mutational units (*i.e.*, $\tau = 2ut$, where u denotes the mutation rate per sequence per generation and t the time since the split in generations). T is the estimated time of the split scaled in units of twice the effective population size of *L. peruvanium* ($T = \tau/\theta_2$). P_{WWH} is the tail probability value of the test statistic of Wang *et al.* (1997), *i.e.*, the proportion of simulated values greater than or equal to the observed value.

and Li 1993), and Fu's F_s statistics (Fu 1997). All tests have been performed using silent sites only. For *L. chilense*, high positive values of Tajima's D are found at the four genes where polymorphism is observed. D is significantly different from zero in *sucr*. This indicates that there is an excess of variants at intermediate frequencies. In three genes (CT251, *sucr*, and CT268), this excess exists because variation is organized in two highly differentiated haplotypes. For *sucr* and CT268, Fu's F_s statistics produce positive values that are significantly different from zero ($P < 0.01$). This indicates that the number of observed haplotypes is lower than expected under neutrality. These findings and the observed lack of fixed differences between *L. chilense* and *L. peruvianum* at *sucr* and CT268 strongly suggest that the *L. chilense* population resulted from recent admixture or hybridization of two diverged populations, with at least one of these originating from *L. peruvianum*.

For the other four species, hardly any of these tests rejects neutrality, even at the 0.05 level. However, there are some general trends that depend more on the species than on the gene. For *L. hirsutum*, positive values of the D and F_s statistics are observed for three genes (out of four polymorphic loci), namely CT251, CT268, and *sucr*. Negative values are observed in CT143, but this last case is probably an exception (discussed below). In *L. peruvianum*, negative values of D are observed for all genes except CT143, which exhibits a slightly positive value. Negative values of D are expected after population size expansions or after hitchhiking events (Fu 1997). Whether any of these hypotheses is an adequate explanation is difficult to decide with the present data set. In *L. pimpinellifolium*, two genes (CT208 and *sucr*) show positive values of Tajima's D and Fu and Li's D while two have negative values (CT143 and CT268). The only gene polymorphic in *L. chmielewskii* (CT143) has a positive D value. But Fu's F_s statistics consistently show negative values in these two SC species at each polymorphic gene. Negative values of Fu's F_s statistic may be due to population size expansion or hitchhiking.

Neutrality tests using both intra- and interspecific data: MK tests were performed for the three genes in which replacement substitutions have been observed. Tests were performed both for every pair of species and for all five species together. None of the tests rejected the null model in CT251 and CT268 at the 0.05 level. In *sucr*, however, there are more fixed replacement substitutions than expected in every pairwise test, although neutrality is rejected in only one case (*L. peruvianum*-*L. pimpinellifolium* comparison, $G = 12.796$; $P < 0.0005$). The MK test performed simultaneously on the five species rejects neutrality in *sucr* ($G = 6.87$; $P < 0.005$), which suggests that this gene is undergoing positive selection.

HKA tests were performed for the two genes located in high recombination regions, namely CT143 and CT268, in each species (except *L. chilense* as it is presumably of recent hybrid origin). In each species, HKA tests were performed using polymorphism data from the species (Hudson *et al.* 1987) and divergence between *L. chilense* and *L. peruvianum* or *L. hirsutum* and *L. pimpinellifolium*. In all cases, results of the tests were identical for both divergence estimates. The HKA tests rejected neutrality at the 0.05 level only in *L. hirsutum*. In this species, variation at the CT143 locus is strongly reduced, suggesting that a recent hitchhiking event has occurred at this locus. This is consistent with the observation that *L. hirsutum* is the only species that is polymorphic and has a strongly negative (though not significant) value of Tajima's D at this locus ($D = -0.83$). Thus, the combined results of the MK and HKA tests suggest that some of the observed heterogeneity of variation among loci (in particular at *sucr* and CT143) is due to positive selection on individual sites at or near these genes.

DISCUSSION

***L. peruvianum*: the ancestral species of the genus?** A large proportion of the variation found within or between species of the genus Lycopersicon originated from *L. peruvianum*. This suggests that *L. peruvianum* is the ancestral species from which the other Lycopersicon species are derived. The differences between the other species would then be caused partly by lineage sorting of the polymorphism of *L. peruvianum* and partly by new

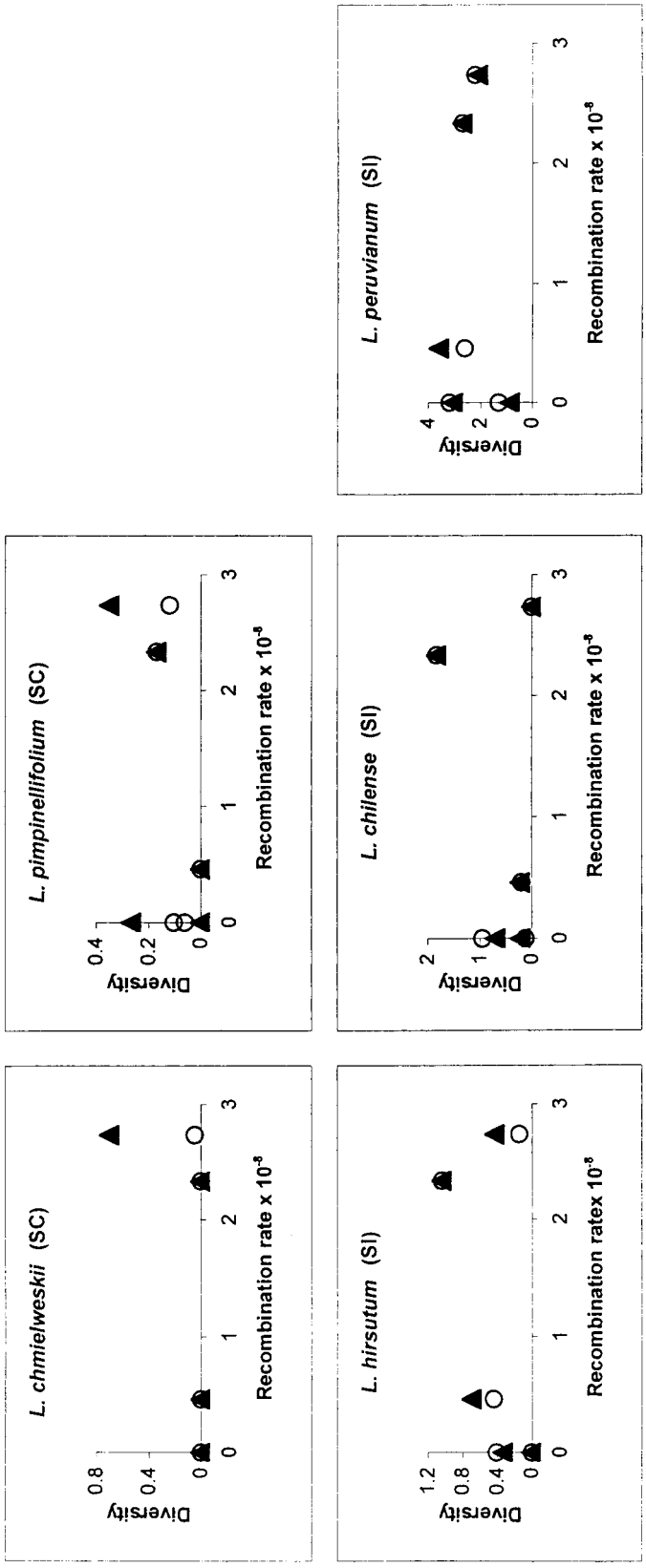

FIGURE 1.—Nucleotide diversity for silent sites, θ_{sil} (circles), and for synonymous sites, θ_{syn}, (triangles) against recombination rate per site per generation for five Lycopersicon species. Estimates of nucleotide diversity are expressed in percentages.

TABLE 5

Results of statistical tests based on intraspecific data

	Low recombination			High recombination	
	CT208	*sucr*	CT251	CT268	CT143
L. chmielewskii					
D	—	—	—	—	0.02
*D**	—	—	—	—	1.03
F_s	—	—	—	—	−1.60
L. pimpinellifolium					
D	0.02	0.63	—	−1.40	−0.18
*D**	1.03	1.03	—	−1.59	−0.02
F_s	−1.60	−1.12	—	−0.59	−2.29
L. chilense					
D	0.99	2.22*	1.23	1.51	—
*D**	1.24	1.33	1.24	1.57*	—
F_s	1.70	6.92**	3.79	11.60**	—
L. hirsutum					
D	—	1.25	1.27	0.49	−0.83
*D**	—	1.55*	1.57*	0.39	−0.79
F_s	—	4.18	2.28	1.08	−0.63
L. peruvianum					
D	−0.63	−0.50	−0.04	−0.58	0.27
*D**	−0.78	−0.25	−0.25	−0.24	0.25

In *L. peruvanium* F_s was not calculated because the number of haplotypes could not be reliably inferred. *D*, Tajima's *D*; *D**, Fu and Li's *D*; F_s, Fu's F_s statistic. *, significant at the 0.05 level; **, significant at the 0.01 level. —, Statistics were not calculated due to a lack of silent polymorphism data.

variants that have arisen after the speciation events. If divergence between species is estimated by taking into account only the variants that have arisen after speciation (*i.e.*, excluding variable sites that are also present in *L. peruvianum*), the estimates of divergence between the four species are much lower. This suggests that the other species, although morphologically differentiated, may have originated from *L. peruvianum* fairly recently. In particular, the origin of *L. chilense* is probably extremely recent, as there are no fixed differences between *L. chilense* and *L. peruvianum* at two of the five genes. This is reflected in the estimates of divergence times obtained by fitting the isolation model to the data (Table 4). Our hypothesis of relatively recent speciation events in the genus Lycopersicon is also consistent with the observations that all species are intercrossable (Hogenboom 1979; Rick 1979) and that their karyotypes are very similar (Rick 1983). In addition, it may explain the difficulties in determining the phylogenetic relationships between the nine species of the genus (Warnock 1988).

DNA polymorphism in SC and SI species: An important result of this study is that the predicted positive correlation between recombination rate and nucleotide diversity in the five Lycopersicon species is unexpectedly weak. This might seem contradictory to Stephan and Langley's (1998) results, in particular as they reported a highly significant positive correlation between recombination rate and DNA polymorphism in eight Lycopersicon species (*L. chilense* was not included). However, in the previous analysis, as in this study, genes located in low-recombination regions of the genome on average do not exhibit much reduced levels of polymorphism relative to the high-recombination loci. The difference between the two studies is largely due to the higher number of loci (36 used by Stephan and Langley *vs.* 5 here).

We also found that SC species have drastically reduced levels of within-population variation at the five loci. *L. pimpinellifolium* and *L. chmielewskii* on average have 4 and 40 times, respectively, lower levels of polymorphism than *L. hirsutum*, the least polymorphic SI species. Previous allozyme (Rick 1983; Doebley 1989) and RFLP studies (Miller and Tanksley 1990; Stephan and Langley 1998) also found very reduced levels of within-population polymorphism in SC species. Under the standard neutral model, DNA polymorphism in a population depends on mutation rate and the effective size of the population. If selection is acting, the frequency of selected mutations, beneficial or deleterious, and the recombination rate per physical unit will also influence the level of polymorphism. Next we discuss how these factors may explain our observations in Lycopersicon.

Standard neutral model: We first consider a situation where all populations have similar constant effective population sizes and evolve under neutrality. In this case, average levels of polymorphism should not depend on recombination rates. Our analyses, however, suggest a

weak effect of recombination on polymorphism. This effect is not significant, but found consistently for all species. Under the standard neutral model, the effect of (complete) selfing is to halve the effective population size and thus the expected genetic variability. As the SC species studied have significant levels of outcrossing, we expect a lower than twofold decrease in polymorphism in these species relative to the SI species. The observed reduction is much higher than that, suggesting that the standard neutral model that assumes approximately equal effective population sizes of the SC species cannot explain our data.

Population size and structure: The strong effect of mating system on polymorphism that we observed in Lycopersicon can be explained simply if we assume that SC species have lower effective population sizes than SI species. Effective population size depends on numerous factors. Among those likely to be important in plants are census population size, mating system, population size fluctuations, and population structure. As *L. pimpinellifolium* is thought to have larger populations than *L. hirsutum* (Rick *et al.* 1978a) but lower polymorphism, differences in census population size alone are unlikely to explain the reduced variation observed in the SC species. As mentioned above, selfing cannot be the sole force either. However, fluctuations in population size and population subdivision may be important factors.

In selfing species, populations can originate from a single individual. It is thus possible that in the SC species, most populations originated from a very reduced number of individuals and have expanded afterward. Such populations would have a very small effective population size relative to their census size and thus a reduced level of polymorphism. Populations that have undergone recent demographic expansions are expected to have negative values of Tajima's D and Fu's F_s statistics (Charlesworth *et al.* 1993; Fu 1997). In *L. pimpinellifolium* and *L. chmielewskii*, Tajima's D values are zero or negative at all loci, except *sucr*, but this locus is probably undergoing positive selection (see above). Fu's F_s statistics show negative values for all loci. It is thus possible that the SC populations that we have studied have been undergoing demographic expansion, but have relatively small effective population sizes.

In addition, population substructuring may be an important factor, as field observations indicate that wild Lycopersicon populations are highly fragmented (Rick 1986). Population subdivision is expected to influence effective population size in various ways. The effective size of a metapopulation can be substantially larger than the actual number of individuals in the entire population when the migration rate among subpopulations is small (Nei and Takahata 1993). On the contrary, frequent extinction and recolonization of subpopulations can lead to a greatly reduced effective metapopulation size (Maruyama and Kimura 1980; Whitlock and Barton 1997). If Lycopersicon SC species are existing in metapopulations with high rates of extinction, the low overall effective population size will lead to a reduced nucleotide polymorphism. It is also possible that the interaction of migration and selection in substructured populations plays a role. The effects of background selection and hitchhiking on DNA polymorphism in subdivided populations are currently not well understood (but see Charlesworth *et al.* 1997) and it is difficult to determine if they could explain the pattern of polymorphism observed in Lycopersicon.

Selection and linkage: Finally, we consider a situation where all populations have similar effective sizes but are experiencing selection for beneficial mutations (hitchhiking) or against deleterious mutations (background selection). At least two cases can be distinguished: weak selection and strong selection. In the case of weak selection, polymorphism is expected to reach the maximum value very rapidly because the recombination rate increases in both the hitchhiking and background selection models (Figure 2A). The shapes of the curves in Figure 2A are consistent with our observations of a weak effect of recombination on polymorphism in the SI and SC species. However, for weak selection, a less than twofold reduction of polymorphism is expected in the SC species compared to the SI species (as in the case of neutrality), which is incompatible with our data. On the other hand, when selection is strong, polymorphism increases slowly with recombination rate and remains very low in selfing species even when recombination rates are high (Figure 2B). This is in agreement with what we observed in the SC species. However, we did not observe a greatly reduced level of polymorphism in genes located in low-recombination regions in the SI species, which is incompatible with strong selection. Whatever the type of selection, it seems reasonable to assume that it is the same in the SI and SC species. As neither weak nor strong selection can explain both the weak effect of recombination and the strong effect of the mating system on polymorphism, it seems that a model in which populations of approximately the same effective size that experience only hitchhiking or background selection cannot explain our results.

Nonetheless, it is possible that the Lycopersicon populations that we studied were undergoing background selection or hitchhiking in the past, but local selective events (at individual loci) were also occurring. If one or more loci in low-recombination regions have undergone balancing selection, and thus have higher than expected levels of polymorphism, while one or more loci in high-recombination regions have experienced a recent hitchhiking event, the correlation created by background selection or hitchhiking between recombination and polymorphism would be much reduced. As discussed above, we suspect that the *sucr* locus, located in a region of low recombination, has undergone balancing selection. There is also evidence that a recent hitchhiking event has occurred at or near the CT143

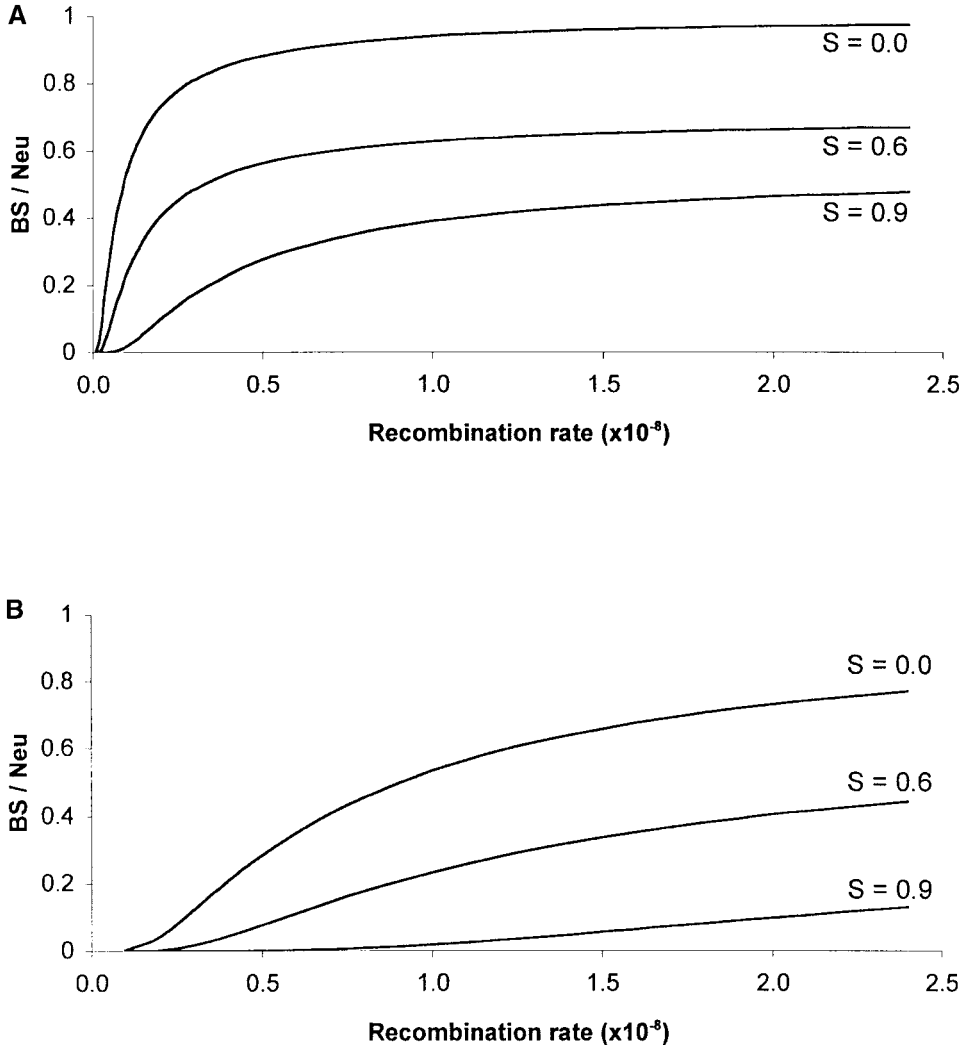

FIGURE 2.—Expected nucleotide diversity against recombination rate for (A) relatively weak selection (with a deleterious mutation rate, u, per site of 10^{-10}), and (B) strong selection (with $u = 10^{-8.5}$). BS/Neu is the expected amount of variation under background selection divided by that under neutrality. S is the selfing rate. For the background selection model Equation (60) in NORDBORG (1997) was used, assuming no dominance. NORDBORG (1997) obtained the equation by rescaling the effective population size, the recombination, and selection parameters. The curves for the hitchhiking model look very similar. In this model, the curves can be obtained from Equation (5) in WIEHE and STEPHAN (1993) by rescaling the three parameters following NORDBORG (1997). This equation gives a reasonable approximation when $S < 0.9$ (H. INNAN and W. STEPHAN, unpublished results).

locus, located in a high-recombination region. These findings can explain, at least partially, the weak effect of recombination on polymorphism found in our data.

Our observation that polymorphism in low-recombination regions is not much reduced could also be due to a low density of genes in these parts of the genome or a low density of genes in the tomato genome in general. If this is the case, the density of targets of natural selection is also low. This may be an important factor, as the rate of nonneutral mutations per map unit may strongly influence the effect of hitchhiking and background selection.

Another key parameter is the variance of the recombination rate across the genome, in particular the presence of local hot or cold spots of recombination. Accurate estimates of fine-scale recombination rates require a comparison of detailed genetic and physical maps. Physical maps of the tomato genome are not yet completed but the current partial maps have shown the existence of recombination hot spots located in several genes (BONNEMA *et al.* 1997; FRIDMAN *et al.* 2000). This suggests that the relationship between recombination rate and position along the chromosomes may be complex in Lycopersicon species. This may be partially responsible for the large scatter of nucleotide diversity among loci and possibly also for the weak effect of recombination on levels of polymorphism.

Conclusions: We have reviewed here several hypotheses, not mutually exclusive, that could explain the patterns of polymorphism in the genus Lycopersicon. More data will be needed to determine the relative importance of these hypotheses. So far, relationships between recombination rate and genetic variation have been investigated in two other plant genera with RFLP data, specifically Aegilops (DVORAK *et al.* 1998) and Beta (KRAFT *et al.* 1998). Preliminary results for maize have also been reported by GAUT *et al.* (2000). In these species, results seem to be similar to our observations in tomato. That is, nucleotide diversity is not extremely reduced in the genes located in low-recombination regions. This situation is in sharp contrast to the steep slope of the regression line of polymorphism *vs.* recombination rate observed in *Drosophila melanogaster* (AQUADRO *et al.* 1994) and *D. ananassae* (STEPHAN *et al.* 1998; CHEN *et al.* 2000),

although in the latter species only a limited number of loci has been surveyed so far.

Numerous studies have investigated the relationships between mating system and variability in plants (Hamrick and Godt 1990; Awadalla and Ritland 1997). However, most of these studies have been performed using allozyme data, and several lines of evidence suggest that allozyme variants may not be neutral (*e.g.*, Karl and Avise 1992; Hudson *et al.* 1994; Pogson and Zouros 1994). Comparisons of sequence diversity data in closely related selfing and outcrossing species are currently available only in the genera Arabidopsis (Miyashita *et al.* 1996; Savolainen *et al.* 2000) and Leavenworthia (Liu *et al.* 1998, 1999). The level of within-population variability in selfing species compared to outcrossing species is very reduced in these genera. The weak influence of recombination and the strong effect of the mating system on DNA polymorphism that we observed in Lycopersicon might therefore be common in plants.

We thank Charles Rick for supplying the plants used in this study, Laura Rose for the DNA extractions, Charles Langley for access to his lab facilities and discussions, Roger Chetelat for useful information on Lycopersicon biology, and Jody Hey for software. We also thank two reviewers and L. Rose for valuable comments on a previous version of this article. This work was supported by funds from the Universities of Rochester and Munich to W.S. and by a postdoctoral fellowship from the Japan Society for the Promotion of Science to H.I.

LITERATURE CITED

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

Aquadro, C. F., D. J. Begun and E. C. Kindahl, 1994 Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 46–55 in *Non-neutral Evolution: Theories and Molecular Data*, edited by G. B. Golding. Chapman & Hall, New York.

Awadalla, P., and K. Ritland, 1997 Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. Mol. Biol. Evol. **14:** 1023–1034.

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.

Bonnema, G., D. Schipper, S. van Heusden, P. Zabel and P. Lindhout, 1997 Tomato chromosome 1: high resolution genetic and physical mapping of the short arm in an interspecific *Lycopersicon esculentum* × *L. peruvianum* cross. Mol. Gen. Genet. **253:** 455–462.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. **68:** 131–149.

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

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

Chen, Y., B. J. Marsh and W. Stephan, 2000 Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. Genetics **155:** 1185–1194.

Clark, A. G., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

Doebley, J., 1989 Isozymic evidence and the evolution of crop plants, pp. 165–191 in *Isozymes in Plant Biology*, edited by D. E. Soltis and P. S. Soltis. Dioscorides Press, Hong Kong.

Dvorak, J., M.-C. 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.

Elliott, K. J., W. O. Butler, C. D. Dickinson, Y. Konno, T. S. Vedvick *et al.*, 1993 Isolation and characterization of fruit vacuolar invertase genes from two tomato species and temporal differences in mRNA levels during fruit ripening. Plant Mol. Biol. **21:** 515–524.

Fridman, E., T. Pleban and D. Zamir, 2000 A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. USA **97:** 4718–4723.

Fu, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

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

Fulton, T. M., J. C. Nelson and H. S. D. Tanksley, 1997 Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, in *Lycopersicon esculentum*, followed through three successive backcross generations. Theor. Appl. Genet. **95:** 881–894.

Ganal, M. W., R. Czihal, U. Hannappel, D.-U. Kloos, A. Polley *et al.*, 1998 Sequencing of cDNA clones from the genetic map of tomato (*Lycopersicon esculentum*). Genome Res. **8:** 842–847.

Gaut, B. S., M. Le Thierry d'Ennequin, A. S. Peek and M. C. Sawkins, 2000 Maize as a model for the evolution of plant nuclear genomes. Proc. Natl. Acad. Sci. USA **97:** 7008–7015.

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 Associates, Sunderland, MA.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hogenboom, N. G., 1979 Incompatibility and incongruity in *Lycopersicon*, pp. 435–444 in *The Biology and Taxonomy of Solanaceae*, edited by J. Hawkes, R. N. Lester and A. D. Skelding. Academic Press, London.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

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

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

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

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

Kraft, T., T. Säll, I. Magnusson-Rading, N.-O. Nilsson and C. Halldén, 1998 Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). Genetics **150:** 1239–1244.

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

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

Maruyama, T., and M. Kimura, 1980 Genetic variability and effective population size when local extinction and recolonization are frequent. Proc. Natl. Acad. Sci. USA **77:** 6710–6714.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

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

Miyashita, N. T., H. Innan and R. Terauchi, 1996 Intra- and interspecific variation in the alcohol dehydrogenase locus region

of wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. Mol. Biol. Evol. **13:** 433–436.

Nachman, M. W., 1997 Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147:** 1303–1316.

Nachman, M. W., V. L. Bauer, S. L. Crowell and C. F. Aquadro, 1998 DNA variability and recombination rates at X-linked loci in humans. Genetics **150:** 1133–1141.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and N. Takahata, 1993 Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evol. **37:** 240–244.

Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Pillen, K., O. Pineda, C. B. Lewis and S. D. Tanksley, 1996 Status of genome mapping tools in the taxon Solanaceae, pp. 281–308 in *Genome Mapping in Plants*, edited by A. H. Patterson and R. G. Landes. R. G. Landes Company, Austin, TX.

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

Rick, C. M., 1979 Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*, pp. 667–678 in *The Biology and Taxonomy of the Solanaceae*, edited by J. G. Hawkes, R. N. Lester and A. D. Skelding. Academic Press, London.

Rick, C. M., 1983 Evolution of mating systems: evidence from allozyme variation, pp. 215–221 in *XV International Congress of Genetics*. Oxford & IBH, New Delhi.

Rick, C. M., 1986 Reproductive isolation in the *Lycopersicon peruvianum* complex, pp. 477–495 in *Solanaceae: Biology and Systematics*, edited by W. G. D'Arcy. Columbia University Press, New York.

Rick, C. M., E. Kesicki, J. F. Fobes and M. Holle, 1976 Genetic and biosystematic studies on two new sibling species of *Lycopersicon* from Interandean Peru. Theor. Appl. Genet. **47:** 55–68.

Rick, C. M., J. F. Fobes and M. Holle, 1977 Genetic variation of *Lycopersicon pimpinellifolium*: evidence of evolutionary change in mating system. Plant Syst. Evol. **127:** 139–170.

Rick, C. M., J. F. Fobes and S. D. Tanksley, 1978a Evolution of mating systems in *Lycopersicon hirsutum* as deduced from genetic variation in electrophoretic and morphological characters. Plant Syst. Evol. **132:** 279–298.

Rick, C. M., M. Holle and R. W. Thorp, 1978b Rates of cross-pollination in *Lycopersicon pimpinellifolium*: impact of genetic variation in floral characters. Plant Syst. Evol. **129:** 31–44.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

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

Sherman, J. D., and S. M. Stack, 1995 Two-dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics **141:** 683–708.

Stephan, W., and C. H. Langley, 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.

Stephan, W., and C. H. Langley, 1998 DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics **150:** 1585–1593.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism—analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

Stephan, W., L. Xing, D. A. Kirby and J. M. Braverman, 1998 A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. Proc. Natl. Acad. Sci. USA **95:** 5649–5654.

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

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de Vincente, M. W. Bonierbale *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics **132:** 1141–1160.

Wakeley, J., and J. Hey, 1997 Estimating ancestral population parameters. Genetics **145:** 847–855.

Wang, R. L., J. Wakeley and J. Hey, 1997 Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics **147:** 1091–1106.

Warnock, S. J., 1988 A review of taxonomy and phylogeny of the genus *Lycopersicon*. Hort. Sci. **23:** 669–673.

Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics **146:** 427–441.

Wiehe, T. H. E., and W. Stephan, 1993 Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 842–854.

Communicating editor: O. Savolainen