Copyright © 1998 by the Genetics Society of America

DNA Polymorphism in Lycopersicon and Crossing-Over per Physical Length

Wolfgang Stephan* and Charles H. Langley†

**Department of Biology, University of Rochester, Rochester, New York 14627 and †Section of Evolution and Ecology, University of California, Davis, California 95616*

Manuscript received April 2, 1998
Accepted for publication July 21, 1998

ABSTRACT

Surveys in Drosophila have consistently found reduced levels of DNA sequence polymorphism in genomic regions experiencing low crossing-over per physical length, while these same regions exhibit normal amounts of interspecific divergence. Here we show that for 36 loci across the genomes of eight Lycopersicon species, naturally occurring DNA polymorphism (scaled by locus-specific divergence between species) is positively correlated with the density of crossing-over per physical length. Large between-species differences in the amount of DNA sequence polymorphism reflect breeding systems: selfing species show much less within-species polymorphism than outcrossing species. The strongest association of expected heterozygosity with crossing-over is found in species with intermediate levels of average nucleotide diversity. All of these observations appear to be in qualitative agreement with the hitchhiking effects caused by the fixation of advantageous mutations and/or "background selection" against deleterious mutations.

THE genus Lycopersicon consists of nine species, of which the only cultivated species is *L. esculentum* (tomato), represented in the wild by var. *cerasiforme* (Rick 1983). Lycopersicon species are crossable with one another in all combinations, though with varying degrees of difficulties (Soost 1958). The karyotypes of the 12 chromosome pairs are very similar with little or no structural differences among species (Rick 1983).

Despite this uniformity of karyotypes and the small number of species, Lycopersicon encompasses a great diversity of mating systems. *L. cheesmanii*, endemic to the Galapagos Islands, has an autogamous mating system, which is typical of many other endemic flowering plants of the archipelago (Rick 1966). Another species exhibiting virtually complete autogamy is *L. parviflorum* (Rick 1983). Self-fertilization prevails among natural populations as well as cultivated varieties of *L. esculentum.* In contrast, *L. pimpinellifolium* shows regional differences in relative levels of outcrossing *vs.* selfing. Autogamy predominantly occurs in peripheral populations of southern Peru and Equador while allogamy prevails in the central parts of the species distribution (Rick 1983). *L. chmielewskii* is another species with a facultative mating system. It has a limited distribution and has not been as extensively studied as other Lycopersicon species. The remaining species (*L. chilense*, *L. hirsutum*, *L. pennellii*, and *L. peruvianum*) differ from these two facultative outcrossers by the presence of a self-incompatibility system (Rick 1987). Self-incompatibility occurs in these four species to a varying degree and is probably most widely distributed in *L. chilense*, *L. pennellii*, and *L. peruvianum* (marginal populations of *L. pennellii*, and *L. peruvianum* are self-compatible). The self-incompatibility system in Lycopersicon is gametophytic and controlled by a single, multiallelic S locus (Tanksley and Loaiza-Figueroa 1985).

Genetic linkage maps have been established in tomato since the beginning of classical genetics (Jones 1911; Butler 1952). Due to the low level of genetic variation among cultivars of *L. esculentum*, the current map was constructed using an F_2 population of the interspecific cross *L. esculentum* × *L. pennellii.* It contains more than 1000 markers that are distributed over 1276 cM (Tanksley *et al.* 1992; Pillen *et al.* 1996; Fulton *et al.* 1997). The centromeres have been localized on these maps. In addition, a quantitative cytogenetic map of the distribution of recombination nodules (RNs) is available for comparison with the linkage map (Sherman and Stack 1995). This cytogenetic map [based on spreads of chromosomal synaptonemal complexes (SCs)] describes the frequency and distribution of RNs at a per 0.1-μm resolution for each of the 12 chromosomes in *L. esculentum.* The distribution of RNs is thought to reflect the distributions of subsequent chiasmata and crossovers.

Our major goals in this study are to investigate the relationship between crossing-over and the level of DNA polymorphism in Lycopersicon, using information from these sources, and to analyze the impact of mating system on DNA polymorphism. This work has been stimulated by data from surveys of DNA polymorphism in natural populations of Drosophila, which consistently show that genetic variation is lower for loci in regions where crossing-over per physical length is relatively infrequent (Aguadé *et al.* 1989; Stephan and Langley 1989), while the same regions exhibit normal amounts

Corresponding author: Wolfgang Stephan, Department of Biology, University of Rochester, Rochester, NY 14627-0211.
E-mail: stephan@troi.cc.rochester.edu

of interspecific divergence (Begun and Aquadro 1991; Berry *et al.* 1991). Interest in Lycopersicon was motivated by the interspecific variation in outcrossing associated with differences in patterns of allozyme variation resulting from it (Rick *et al.* 1979; Rick and Tanksley 1981) and the clear evidence for large differences among chromosomal regions in the level of crossing-over per physical length (Sherman and Stack 1995, and references therein). We approach these goals in three steps: (1) We align the RN-cytogenetic maps and linkage maps to estimate the local density of crossing-over per physical length. (2) We reanalyze Miller and Tanksley's (1990) RFLP data obtained from eight Lycopersicon species (*L. chilense* is absent) and 41 loci distributed across all 12 chromosomes. (3) We conduct a four-cutter survey of DNA sequence variation at the *sucrose accumulator* gene (*sucr*) (Chetelat *et al.* 1995) and the cystolic superoxide dismutase gene, *Sod-2* (Perl-Treves *et al.* 1990), using a sample from a *L. peruvianum* population; *sucr* is located in the centromere region of chromosome *3* in a region of reduced crossing-over per physical length, whereas *Sod-2* is on the long arm of chromosome *1* in a region of normal crossing-over (Figure 1).

MATERIALS AND METHODS

Construction of a crossing-over per physical length map: We construct a map to estimate the density of crossing-over per physical length based on the quantitative cytogenetic map for the cultivated tomato, *L. esculentum* (Sherman and Stack 1995), which shows the frequency of RNs in each 0.1-μm segment of the SCs of the 12 chromosomes of *L. esculentum* (>400 observed SCs per chromosome). We apply the "lowess procedure" (Chambers *et al.* 1983; weighting parameter is 5%) to smooth the local variation along the chromosomes thereby emphasizing the regional characteristics of the map (for instance, extended segments of low or high recombination rates) over local variation [much of which reflects the finite sampling of the original observations (Sherman and Stack 1995)]. In a second step, we align the updated genetic maps (Pillen *et al.* 1996; Fulton *et al.* 1997) and these (smoothed) RN maps in a linear fashion such that the centromeres and telomeres of the chromosomes' cytogenetic maps correspond to ends of the genetic maps of each chromosome arm. In those cases where the genetic location of the centromere covered several adjacent intervals (Fulton *et al.* 1997), the centromere is assumed to be in the midpoint of these intervals. The density of RNs per micrometer (RN/μm) for each of the mapped loci can be assigned by interpolation.

RFLP data source and analysis: Thirty-six loci of the data set of Miller and Tanksley (1990) that could be localized unambiguously on the recent genetic linkage map (Pillen *et al.* 1996) were used in this analysis. The raw data were given as sets of restriction fragment lengths for each locus, each plant, and each restriction enzyme (Miller 1989). These RFLP data (southern blots of digests with five six-cutter restriction enzymes) were obtained from a total of 156 plants representing nine taxa [eight species and one sample from an isolated population identified as *L. peruvianum* var. *humifusum*, *LA2150*; following Miller and Tanksley (1990), *LA2150* is considered a separate taxon]. As mentioned above, the nine taxa can be partitioned into three groups based on their mating systems (Rick 1987): self-compatible and typically self-fertilizing (*L. cheesmanii*, *esculentum*, and *parviflorum*), self-compatible with intermediate levels of outcrossing (*L. chmielewskii* and *pimpinellifolium*), and typically self-incompatible and consequently outcrossing (*L. hirsutum*, *pennellii*, *LA2150*, and *peruvianum*).

We estimate genetic variation within each taxon s (expected number of pairwise differences per nucleotide site, $\hat{\pi}_{sl}$) for each locus l on the basis of proportion of shared restriction fragments (same lengths), using Equations 5.52 through 5.55 from Nei (1987),

$$\hat{\pi}_{sl} = \sum_{i} \sum_{j<i} \sum_{r} \frac{\hat{d}_{ijrsl}}{T_{rsl}},$$

where $\hat{d}_{ijrsl} = -(2/w_r)\ln(\hat{G}_{ijrsl})$. $\hat{G}_{ijrsl}$ solves the equation $\hat{G}_{ijrsl} = \sqrt[4]{\hat{F}_{ijrsl}(3 - 2\hat{G}_{ijrsl})}$, where $\hat{F}_{ijrsl} = m_{ijrsl}/(m_{irsl} + m_{jrsl})$; m_{irsl} and m_{jrsl} are the numbers of fragments generated by restriction enzyme r in individuals i and j of species s at locus l, while m_{ijrsl} is the number of shared fragments. T_{rsl} is the total number of comparisons for enzyme r, species s, and locus l; w_r is the number of nucleotides in the recognition sequence of the restriction enzyme [$w_r = 6$ for the restriction enzymes in Miller and Tanksley (1990): *Dra*I, *Eco*RI, *Hin*dIII, *Eco*RI, and *Xba*I]. Estimates of genetic differences derived from the proportion of shared restriction fragments are necessarily based on simplifying assumptions (exclusively single nucleotide differences, haploidy, complete detectability of all fragments, etc.). As Kaplan (1983) points out, when divergence is small, estimates based on alternative sets of assumptions do not differ greatly, while at high levels of divergence estimates are strongly model dependent. Despite the generally low levels of polymorphism at these loci (see Table 1) the interpretation of the quantitative results must be tempered by the acknowledgment of the indirect nature of the estimation procedure.

Recognizing that systematic differences among loci in the levels of variation exist because of differences in probe size and inherent mutation rate, we estimate interspecific divergence, $\hat{d}_{pl}$, for each locus l over three apparently independent, evolutionary paths p: *esculentum* to *pimpinellifolium*, *hirsutum* to *pennellii*, and *cheesmanii* to *peruvianum* (Miller and Tanksley 1990), using the same method, *i.e.*, on the basis of the proportion of shared fragments,

$$\hat{d}_{pl} = \frac{1}{EI_pJ_p} \sum_{r}\sum_{i}\sum_{j} \hat{d}_{ijrpl},$$

where $\hat{d}_{ijrpl}$ is estimated as above, but now between individual i in one species and individual j in the other species of the phylogenetic path p. E is the number of enzymes used, I_p is the number of individuals of one species of path p, and J_p is the number of individuals of the other species. Two rescaling factors for each of the L loci are calculated by averaging over the P evolutionary paths in two different ways: the relative average divergence,

$${}^{\mathrm{rad}}\delta_l = \frac{\sum_p \hat{d}_{pl}}{(1/L)\sum_l \sum_p \hat{d}_{pl}},$$

and average relative divergence,

$${}^{\mathrm{ard}}\delta_l = \frac{1}{P}\sum_p \frac{\hat{d}_{pl}}{(1/L)\sum_l \hat{d}_{pl}}.$$

The rescaled nucleotide diversities $\Pi_{sl} = \hat{\pi}_{sl}/{}^{\mathrm{rad}}\delta_l$ are used in the analysis of covariance (analyses using the other rescaling factor, ${}^{\mathrm{ard}}\delta_l$, yield similar results, not shown).

Analysis of covariance: The model of analysis of covariance for the crossing-over per physical length and species effects

on genetic variation is

$$\Pi_{sl} = S_s + C_s R_l + \varepsilon_{sl},$$

where S_s ($s = 1, \ldots, 9$) are the species-specific "elevations" or intercepts, C_s are the species-specific slopes of Π_{sl} on crossing-over per physical length, R_l is the density of RN/μm estimated for locus l ($l = 1, \ldots, 36$), and ε_{sl} is an error effect. We conduct two types of analyses with this model: a "parametric analysis" of Π_{sl} in terms of estimated R_l and a "ranked analysis" (same model with ranked observations and ranked R_l values).

Experimental procedures: The determination of four-cutter restriction site variation at the sucrose accumulator locus (*sucr*) and the cystolic Cu/Zn superoxide dismutase locus (*Sod-2*) is based on the survey of five plants from the 1995 maintenance *L. peruvianum* population of the Tomato Genetics Resource Center of the University of California at Davis (accession *LA2744*). The founding seeds of this accession were originally collected at Sobraya (Azapa), Tarapacá, Chile in 1986. The plants used in this study (kindly provided by C. M. Rick) were from the second generation of mass sib-pollination of a greenhouse population (48 individuals). The protocol for the preparation of genomic DNA is adapted from Chetelat *et al.* (1995). The choices of PCR primers are based on the published *L. esculentum* sequences (accession numbers Z12027 and X87372 for *sucr* and *Sod-2*, respectively). Primers are placed in coding sequences and spaced approximately every 500 bp. Longer PCR fragments were also examined, but the interpretation of the banding pattern was ambiguous, particularly when restriction site heterozygosities occurred. PCR reaction mixture (Long *et al.* 1998) included tricine buffer and *Taq* Extender (Stratagene, La Jolla, CA). The PCR products are cut directly in the PCR buffer with eight four-cutter enzymes (*Alu*I, *Cfo*I, *Dde*I, *Hae*III, *Hin*I, *Hpa*II, *Rsa*I, *Scr*FI) and run on a gel made of Synergel (Research Product International Corp.). The gels are stained with ethidium bromide and photographed.

RESULTS

Maps of the density of RN/μm and linkage: To analyze the effects of crossing-over on DNA polymorphism in Lycopersicon, we constructed a combined physical and recombination map using a quantitative RN-cytogenetic map of *L. esculentum* (Sherman and Stack 1995) and the high-resolution linkage map from the cross of *L. esculentum* and *L. pennellii* (Pillen *et al.* 1996; Fulton *et al.* 1997). Figure 1 shows the estimated density of recombination nodules, *R*, along the SC for chromosomes *1* and *3*. RNs are thought to be the earliest cytogenetic manifestations of the process yielding meiotic chromosome exchange (Carpenter 1979a,b). As has been recognized for decades (reviewed in Sherman and Stack 1995), all euchromatic chromosome arms show an extended centromeric proximal region in which the frequency of exchange per physical length is severely reduced relative to that in the distal portions. Crossing-over is suppressed in the centric heterochromatin and in the regions immediately adjacent to the telomeres. This map is also useful for other Lycopersicon species as little or no structural variation among their karyotypes is observed (Khush and Rick 1968), and the genetic maps are comparable among species (van Ooijen *et al.* 1994).

Results of analysis of covariance: Shown in Table 1 are the 36 loci from the survey of Miller and Tanksley (1990) and estimates of average numbers of differences per site within species for each locus and each of the eight species. One locus, TG12, lacked sufficient data in several species. Five additional loci in the original study of Miller and Tanksley were excluded because of paucity of observations or ambiguity in interpretation of the original observations, *e.g.*, multiple loci per probe. Table 1 also shows the rescaling factors, $^{rad}\delta_l$ and $^{ard}\delta_l$. Finally, Table 1 shows the estimates of Π_{sl} (the average number of differences per site within species, $\hat{\pi}_{sl}$, rescaled by the relative average divergence, $^{rad}\delta_l$). Estimates of the unscaled nucleotide diversity (and those rescaled by $^{ard}\delta_l$) can be obtained by appropriate multiplication.

Table 2 presents the results of the analyses of covariance Π_{sl} in terms of *species* and *R*. In both the analysis of Π_{sl} and the analysis of ranked Π_{sl} in terms of ranked *R*, there is strong support for a species effect ($P < 0.0001$). A positive *R* effect on Π_{sl} is also strongly supported in both analyses. There is no support for an interaction effect, $R \times$ species, *i.e.*, for separate slopes. The estimated slopes (of the species-specific regression of Π_{sl} on *R*) vary considerably among species, predicting between 9 and 246% (estimated mean, 54%) differences in nucleotide diversity (relative to the species average, $\overline{\Pi}_s$) over the total range of crossing-over values (*i.e.*, between 0 and 0.22 RN/μm). DNA sequence polymorphism is substantially lower (average Π_{sl}, $\overline{\Pi} = 0.0042$) in selfing species (*L. parviflorum*, *cheesmanii*, and *esculentum*) than in the partially outcrossing species, *L. chmielewski* and *pimpinellifolium* ($\overline{\Pi} = 0.0069$) and lower still than in those species with mating self-incompatibility, *L. hirsutum*, *peruvianum*, *LA2150*, and *pennellii* ($\overline{\Pi} = 0.0175$). This conclusion corroborates previous allozyme studies (Rick 1983; Doebley 1989; Bretó *et al.* 1993; but see also Hamrick and Godt 1989) and the earlier interpretations of these data (Miller and Tanksley 1990).

Restriction site variation at *sucr* and *Sod-2* in *L. peruvianum*: The positive correlation between crossing-over per physical length and DNA polymorphism is supported by a RFLP analysis of two gene regions in a survey of a *L. peruvianum* population. The *sucr* gene is located in the centromeric region of chromosome *3*. Based on its position [genetic position = 55.6 (Chetelat *et al.* 1995)] on the genetic map (Pillen *et al.* 1996; Fulton *et al.* 1997), we estimate a rate of crossing-over of 0.00 RN/μm (Figure 1). Variation is surveyed in a region of ∼3750 bp (of the 4-kb *sucr* transcriptional unit). Our four-cutter method (eight enzymes) allows us to identify a total of 64 restriction sites over this length of DNA, 3 of which are polymorphic within the *L. peruvianum* sample. One of the 3 polymorphic sites is a replacement polymorphism. Six fixed differences are found between

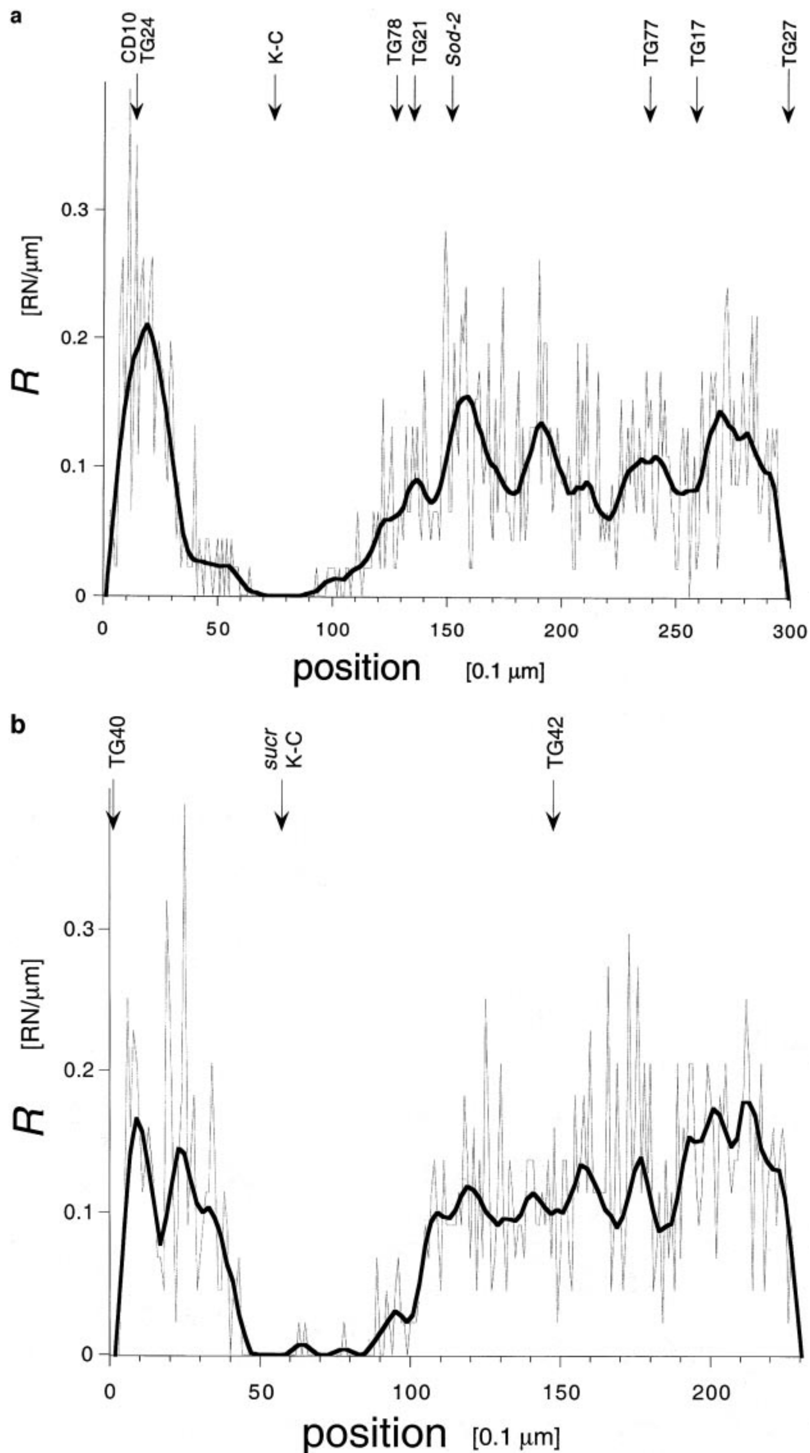

Figure 1.—Maps of the densities of recombination nodules R (RN/μm) from Sherman and Stack (1995). The gray lines are their original data and the black lines are the smoothed estimates. These "smoothed" maps are aligned with the genetic maps of each chromosome arm so that the R values for each of the surveyed loci can be interpolated. (a) The map for chromosome *1.* The positions of various loci are indicated as is the position of the kinetochore/centromere (K-C). (b) A similar map for chromosome 3. Also shown on these maps are the positions of the two loci surveyed for four-cutter restriction map variation in *L. peruvianum*, *Sod-2* and *sucr.*

L. esculentum and *L. peruvianum.* The relatively low number of observed restriction sites is largely due to the AT-rich composition of the introns that make up 51.3% of the total sequence. To obtain estimates for the standard nucleotide diversity statistics π (Nei 1987) and θ (Watterson 1975), we estimate the number, $\hat{L}$, of silent sites surveyed as

$$\hat{L} = 2\sum_i l_i,$$

where the sum is over all restriction sites observed in *L. esculentum*, and l_i is the number of silent positions associated with restriction site *i.* Factor 2 takes into account that sequences that are one off the recognition sequence of a restriction enzyme are included in the screen for DNA polymorphism. We resort to this procedure because the *sucr* and *Sod-2* DNA sequences of *L. peruvianum* are unknown. Assuming that all restriction site polymorphisms are due to changes of single nucleo-

TABLE 1

Summary of our analyses of Miller and Tanksley's (1990) RFLP data

Locus	Chromosome	RN/μm	$^{ard}\delta_l$	$^{rad}\delta_l$	*L. cheesmanii*	*L. chmielewskii*	*L. esculentum*	*L. hirsutum*	*L. parviflorum*	*L. pennellii*	*LA2150*	*L. peruvianum*	*L. pimpinellifolium*
TG78:	*1*	0.064	1.924	2.252	0.000	0.000	0.003	0.000	0.000	0.004	0.000	0.009	0.004
TG24:	*1*	0.192	1.702	1.899	0.000	0.002	0.003	0.020	0.001	0.007	0.004	0.015	0.009
CD10:	*1*	0.192	0.494	0.476	0.003	0.003	0.004	0.060	0.003	0.024	0.005	0.025	0.014
TG21:	*1*	0.101	0.654	0.866	0.000	0.000	0.002	0.044	0.001	0.010	0.003	0.027	0.000
TG77:	*1*	0.108	1.326	1.695	0.003	0.003	0.000	0.004	0.000	0.013	0.002	0.016	0.001
TG17:	*1*	0.101	1.341	1.182	0.002	0.000	0.001	0.014	0.000	0.009	0.006	0.025	0.002
TG27:	*1*	0.000	1.085	1.522	0.000	0.018	0.000	0.004	0.000	0.014	0.010	0.039	0.002
TG31:	*2*	0.038	0.466	0.606	0.000	0.000	0.000	0.051	0.000	0.025	0.004	0.018	0.000
CD53:	*2*	0.138	0.739	0.928	0.004	0.000	0.001	0.021	0.000	0.039	0.008	0.036	0.000
TG14:	*2*	0.147	1.103	1.064	0.002	0.000	0.003	0.017	0.000	0.011	0.005	0.023	0.003
TG40:	*3*	0.000	1.724	1.454	0.000	0.000	0.004	0.009	0.000	0.002	0.006	0.023	0.008
TG42:	*3*	0.113	0.619	0.619	0.008	0.001	0.006	0.011	0.000	0.023	0.009	0.016	0.011
TG49:	*4*	0.113	0.409	0.483	0.003	0.001	0.000	0.024	0.000	0.042	0.006	0.037	0.007
CD59:	*4*	0.116	1.062	0.959	0.001	0.007	0.005	0.027	0.000	0.035	0.007	0.037	0.021
CD70:	*4*	0.100	1.115	1.287	0.000	0.000	0.010	0.005	0.000	0.003	0.002	0.017	0.003
TG62:	*4*	0.124	0.332	0.396	0.014	0.005	0.000	0.051	0.000	0.037	0.002	0.036	0.007
CD39:	*4*	0.158	1.415	1.194	0.001	0.002	0.007	0.020	0.000	0.041	0.003	0.020	0.023
CD64:	*5*	0.120	0.714	0.742	0.000	0.006	0.000	0.000	0.000	0.014	0.011	0.027	0.005
TG69:	*5*	0.188	0.788	0.977	0.000	0.005	0.004	0.027	0.000	0.010	0.006	0.015	0.001
CD67:	*6*	0.022	1.207	1.250	0.000	0.000	0.005	0.021	0.000	0.008	0.013	0.029	0.010
TG54:	*6*	0.183	0.681	0.860	0.000	0.000	0.003	0.036	0.000	0.016	0.004	0.058	0.001
CD57:	*7*	0.156	0.729	0.780	0.002	0.011	0.002	0.002	0.000	0.007	0.010	0.034	0.010
TG61:	*7*	0.196	0.580	0.724	0.000	0.006	0.001	0.029	0.000	0.043	0.012	0.041	0.004
CD60:	*8*	0.165	1.140	0.854	0.000	0.000	0.028	0.004	0.000	0.025	0.003	0.052	0.007
TG45:	*8*	0.081	0.392	0.340	0.001	0.000	0.002	0.015	0.000	0.010	0.006	0.027	0.014
TG18:	*9*	0.169	0.684	0.526	0.000	0.000	0.011	0.032	0.000	0.014	0.016	0.034	0.013
CD77:	*10*	0.216	1.700	1.571	0.000	0.002	0.008	0.014	0.000	0.016	0.008	0.018	0.014
TG12:	*10*	0.058	0.744	0.563	0.000	No data	0.020	No data	No data	No data	0.009	0.020	0.004
CD72:	*10*	0.192	0.895	0.792	0.010	0.018	0.007	0.021	0.000	0.031	0.005	0.021	0.012
CD05:	*10*	0.124	0.596	0.730	0.004	0.000	0.001	0.009	0.000	0.034	0.007	0.050	0.007
TG63:	*10*	0.151	2.257	1.547	0.000	0.000	0.007	0.006	0.000	0.028	0.008	0.017	0.006
CD17:	*11*	0.207	1.454	1.157	0.001	0.004	0.022	0.025	0.001	0.023	0.008	0.029	0.021
TG46:	*11*	0.166	0.336	0.334	0.000	0.000	0.008	0.049	0.000	0.044	0.029	0.060	0.014
TG26:	*11*	0.146	1.123	1.206	0.012	0.006	0.011	0.020	0.000	0.014	0.008	0.030	0.012
TG30:	*11*	0.208	1.199	1.250	0.001	0.003	0.004	0.021	0.000	0.031	0.008	0.032	0.006
CD19:	*12*	0.192	1.187	1.138	0.000	0.007	0.004	0.023	0.000	0.018	0.020	0.023	0.023

The first two columns show the 36 loci and their corresponding chromosomes. Column 3 contains for each locus the estimated number of RN/μm, R. The next two columns contain the estimates of the relative rate of divergence, $^{ard}\delta_l$ and $^{rad}\delta_l$, for the loci. The remaining columns contain Π_{sl}, the estimate of $\hat{\pi}_{sl}$ rescaled by $^{rad}\delta_l$.

TABLE 2

Analysis of covariance of the average number of differences per site within species, $\hat{\pi}_{sl}$, rescaled by the relative average divergence, $^{rad}\delta_l$

Source	d.f.	Sum of squares	F ratio	P
Species	8	0.00293	4.82	<0.0001
		376638	*16.21*	*<0.0001*
R	1	0.00086	11.35	0.0009
		74111	*25.52*	*<0.0001*
Species × R	8	0.00059	0.98	0.45
		24217	*1.04*	*0.40*

The results for the parametric analysis on each line are presented, while below and italicized are the results for the ranked Π_{sl} and ranked R_l analysis.

tides and considering only silent site variation, the estimates of nucleotide diversity for *sucr* are (with $\hat{L} = 222$ and sample size $n = 10$; five diploid genomes) $\hat{\theta} = 0.0032$ and $\hat{\pi} = 0.0038$.

The *Sod-2* gene is located on the long arm of chromosome *1.* On the basis of its position [genetic position = 45.8 (Pillen *et al.* 1996)] on the genetic map, we estimated a recombination rate of 0.137 RN/μm (see Figure 1). Our method allows the survey of variation in a region of roughly 3300 bp of the 3.5-kb *Sod-2* transcriptional unit. Due to the high AT-content of the introns that make up 86.7% of the total of exon and intron sequences of *Sod-2*, we identified only 34 restriction sites, 8 of which were polymorphic within and among the surveyed *L. peruvianum* lines. Five fixed differences are found between *L. esculentum* and *L. peruvianum.* Assuming that all restriction site polymorphisms in exons are at synonymous positions, the estimates of nucleotide diversity are (with $\hat{L} = 210$) $\hat{\theta} = 0.0135$ and $\hat{\pi} = 0.0145$. Thus, we find higher levels of nucleotide diversity in *Sod-2* than in *sucr*, which is consistent with the estimates from the analysis of the anonymous regions (above).

DISCUSSION

Recombination map: By overlaying a high-resolution physical map (Sherman and Stack 1995) and an updated set of linkage data (Pillen *et al.* 1996; Fulton *et al.* 1997), we are able to construct a map for *L. esculentum* that gives the rate of crossing-over in units of RN/μm. This map captures many of the properties of the genetic linkage in Lycopersicon that have been reported since the dawn of classical genetics (Jones 1911). However, given the experimental errors of both the physical and the linkage maps and the assumptions of our construction, it must still be considered quantitatively crude. Another potential problem is that the maps vary between species. For instance, the comparison of a *L. peruvianum* intraspecific linkage map with the *L. esculentum* map that is based on *L. esculentum* × *L. pennellii* crosses revealed on average a 10% increase in chromosome length for the intraspecific map (van Ooijen *et al.* 1994). In agreement with earlier reports (Sherman and Stack 1995 and references therein), our map shows that crossing-over per physical length is suppressed over a substantial fraction of the euchromatic regions of each chromosome, in particular in those regions proximal to the centromeres and telomeres.

Species effects on levels of polymorphism: As expected from allozyme studies (Rick 1983) and from the observations by Miller and Tanksley (1990) in their original publication of these RFLP data, the selfing species show much lower average levels of variation than those with high degrees of outcrossing. Analysis of covariance reveals that there are highly significant differences in levels of variation between species (see Table 2). Figure 2a depicts the observed distribution of Π_{sl} for each species (as a function of $\overline{\Pi}_s$). A similar observation of reduced DNA sequence polymorphism was recently reported for selfing populations of Leavenworthia (Liu *et al.* 1998).

Recombination and species effects on levels of variation: Both the analysis of covariance and our survey of the *sucr* and *Sod-2* genes in *L. peruvianum* support the hypothesis that DNA polymorphism correlates with rates of crossing-over per physical length. Thus, this effect, which has been observed in several Drosophila species [including *D. ananassae* (Stephan and Langley 1989), *D. melanogaster* (Aguadé *et al.* 1989; Begun and Aquadro 1992), *D. simulans* (Begun and Aquadro 1991; Berry *et al.* 1991), *D. mauritiana*, and *D. sechellia* (Hilton *et al.* 1994)] and in mice (Nachman 1997), has been confirmed in a relatively distant relative, Lycopersicon. And very recently levels of RFLP were measured in selfing and outcrossing species of Aegilops (Dvorák *et al.* 1998). An association between allelic diversity and presence in genomic regions of low crossing-over was found. However there was no attempt to correct levels of polymorphism in Aegilops for locus-specific rates of divergence or to measure variation in terms of nucleotide diversity.

Figure 2b shows the normalized distribution of Π_{sl}, corrected for the species average (over loci), *i.e.*, $\Pi_{sl} + (\overline{\Pi} - \overline{\Pi}_s)$. Despite the considerable scatter the regression of these values on R for each locus yields a positive slope consistent with hypotheses tests in the analysis of covariance. A quantitative interpretation of this relationship in Lycopersicon in terms of theoretical models must await more extensive and detailed observations (see below).

In a separate analysis and despite the lack of support for heterogeneity among species, we examined the slopes of the regression of Π_{sl} on R in each species from the same analysis of covariance model. As can be seen in Figure 3 those species with intermediate $\overline{\Pi}_s$, *L. hirsutum* and *pennellii* have the largest slopes; in contrast, completely selfing species (with the lowest $\overline{\Pi}_s$) and the

most consistent outbreeder (*L. peruvianum*) have shallower slopes (with the highest $\bar{\Pi}_s$). Thus, the strongest association of expected heterozygosity with crossing-over per physical length occurs in species with intermediate levels of DNA polymorphism.

Population genetics theory: Perhaps the simplest explanation for a correlation between levels of crossing-over per physical length and levels of polymorphism would be that recombination itself contributes directly by increasing the input of new fragment lengths. This hypothesis would also predict a correlation of divergence with crossing-over per physical length. We examined this relationship for both measures of divergence and R (Table 1). In neither case is there any suggestion of a positive association between divergence and crossing-over.

Two models have been proposed to explain the reduction of DNA sequence polymorphism in regions of low rates of crossing-over: 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 loci 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. Qualitatively, both models can explain the observed

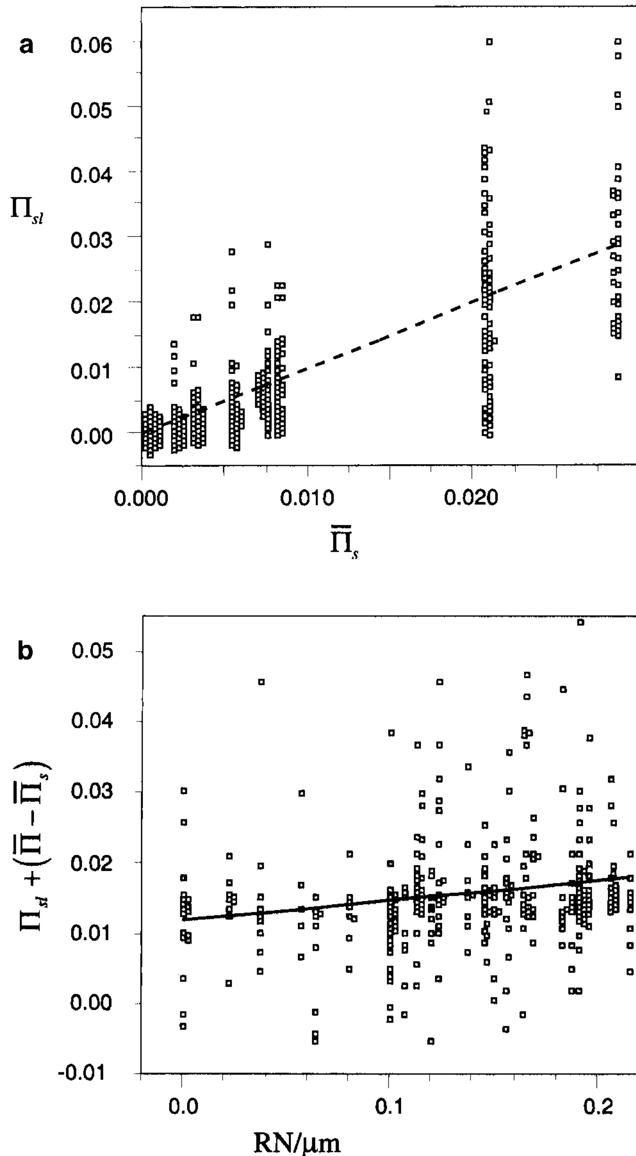

Figure 2.—The distribution of the Π_{sl}. (a) The estimates of $\hat{\pi}_{sl}$ rescaled by $^{rad}\delta_l$ are plotted against the species average, $\bar{\Pi}_s$. The columns of points are from left to right, *L. parviflorum*, *cheesmanii*, *esculentum*, *chmielewskii*, *LA2150*, *pimpinellifolium*, *pennellii*, *hirsutum*, and *peruvianum.* (b) The Π_{sl}, corrected for the species average (over loci), *i.e.*, $\Pi_{sl} + (\bar{\Pi} - \bar{\Pi}_s)$ are plotted against R, the estimated density of RN/µm. The line depicts the overall slope estimate from the analysis of covariance.

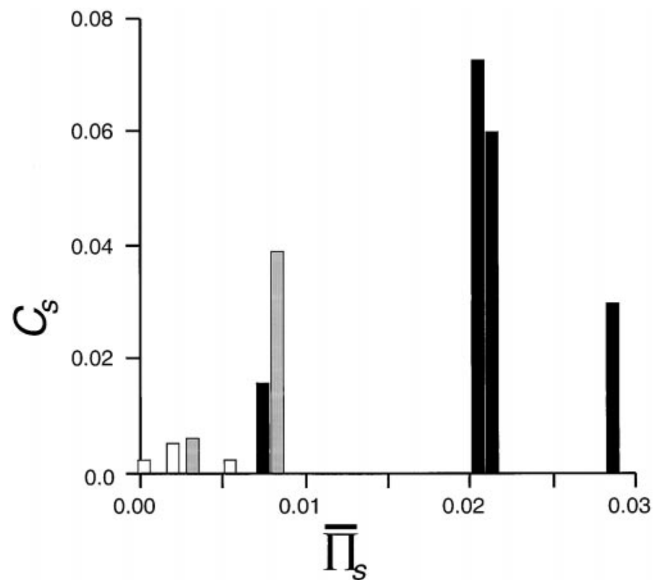

Figure 3.—The estimated values of C_s (in units of corrected nucleotide differences per site per RN/µm) from the parametric model analysis of covariance are plotted against the estimated average, $\bar{\Pi}_s$ (in units of corrected nucleotide differences per site). The open bars represent the selfers (in order, *L. parviflorum*, *cheesmanii*, and *esculentum*), the shaded bars the self-compatible species with intermediate levels of outcrossing (*L. chmielewskii* and *pimpinellifolium*), and the solid bars the species with self-incompatibility alleles (*LA2150*, *L. pennellii*, *hirsutum*, and *peruvianum*). The slope of the increase in average number of differences per site with increasing R is low for species with a low $\bar{\Pi}_s$, while the two species (*L. pennellii* and *hirsutum*) with intermediate $\bar{\Pi}_s$ show the strongest response with crossing-over per physical length. On the right is plotted the slope for the outcrossing species, *L. peruvianum*, which has the highest overall level of variation but a shallow slope with increasing R more typical of selfing species. A least-squares fit of C_s to a quadratic model in $\bar{\Pi}_s$ yields a good fit ($r^2 = 0.78$, $P < 0.01$; both the linear and quadratic coefficients are significantly different from zero), which supports the suggested nonlinearity of the relationship.

positive correlation between crossing-over per physical length and DNA sequence diversity within species. The large difference (greater than twofold) in $\overline{\Pi}$ between selfing species and self-incompatible species can also be attributed to hitchhiking of either kind (Hedrick 1980; Charlesworth *et al.* 1993; Nordborg 1997).

The apparently nonlinear relationship between $\overline{\Pi}_s$ and the slope of the species-specific regression of Π_{sl} on R (Figure 3) may also be attributable to either of the hitchhiking effects. The strength of the effects of hitchhiking depends on the density of selected sites and the intensity of selection relative to recombination rates. If the rate of outcrossing and (thus directly) the rate of crossing-over are sufficiently large, the hitchhiking effect of favorable or deleterious mutations will be limited (to small genomic regions). On the other hand, if outcrossing is rare (and thus also the impact of recombination and independent chromosome segregation), such hitchhiking effects may stretch across the whole genome, reducing the impact of regional genomic differences in crossing-over per physical length. A quantitative understanding of this nonlinear relationship between breeding structure and the hitchhiking effects requires more data on DNA sequence variation and a more quantitative elaboration of the predictions of these hitchhiking effect models.

We thank C. Rick for plant material and advice, R. Chetelat for DNAs and protocols, and A. Long for statistical advice. C. Aquadro, B. Charlesworth, and M. Nordborg made helpful suggestions that improved the presentation of this paper. This research was supported in part by National Science Foundation grants DEB-9407226 to W.S. and DEB-9509548 to C.H.L.

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.

Begun, D. J., and C. F. Aquadro, 1991 Molecular population genetics of the distal portion of the *X* chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129:** 1147–1158.

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

Berry, A. J., J. W. Ajioka and M. Kreitman, 1991 Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

Bretó, M. P., M. J. Asins and E. A. Carbonell, 1993 Genetic variability in *Lycopersicon* species and their genetic relationships. Theor. Appl. Genet. **86:** 113–120.

Butler, L., 1952 The linkage map of tomato. J. Hered. **43:** 25.

Carpenter, A. T. C., 1979a Synaptonemal complex and recombination nodules in wild-type *Drosophila melanogaster* females. Genetics **92:** 511–541.

Carpenter, A. T. C., 1979b Recombination nodules and synaptonemal complex in recombination-defective females of *Drosophila melanogaster.* Chromosoma **75:** 259–292.

Chambers, J., W. Cleveland, B. Kleiner and P. Tukey, 1983 *Graphical Methods for Data Analysis.* Duxbury Press, Boston.

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.

Chetelat, R. T., J. W. DeVerna and A. B. Bennett, 1995 Introgression into tomato (*Lycopersicon esculentum*) of *L. chmielewskii* sucrose accumulator gene (*sucr*) controlling fruit sugar composition. Theor. Appl. Genet. **91:** 327–333.

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.

Dvorák, 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.

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

Hamrick, J. L., and M. J. W. Godt, 1989 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 Inc., Sunderland, MA.

Hedrick, P. W., 1980 Hitchhiking: a comparison of linkage and pertial selfing. Genetics **94:** 791–808.

Hilton, H., R. M. Kliman and J. Hey, 1994 Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48:** 1900–1913.

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

Jones, D. F., 1911 Linkage in *Lycopersicon.* Am. Nat. **51:** 608–6212.

Kaplan, N. L., 1983 Statistical analysis of restriction enzyme data and nucleotide sequence data, pp. 75–106 in *Statistical Analysis of DNA Sequence Data*, edited by B. S. Weir. Marcel Dekker Inc., New York.

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

Khush, G. S., and C. M. Rick, 1968 Cytogenetic analysis of the tomato genome by means of induced deficiencies. Chromosoma **23:** 452–484.

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

Long, A. D., R. F. Lyman, C. H. Langley and T. F. C. Mackay, 1998 Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster.* Genetics **149:** 999–1017.

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

Miller, J. C., 1989 RFLP analysis of phylogenetic relationships, genetic variation and factors affecting the detection of polymorphism in the genus *Lycopersicon.* Dissertation, Cornell University, Ithaca, NY.

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.

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

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

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

Perl-Treves, R., M. Abu-Abied, N. Magal, E. Galun and D. Zamir, 1990 Genetic mapping of tomato cDNA clones encoding the chloroplastic and the cytosolic isozymes of superoxide dismutase. Biochem. Genet. **28:** 543–552.

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. Paterson. R. G. Landes, Austin, TX.

Rick, C. M., 1966 Some plant-animal relations on the Galapagos Islands, pp. 215–224 in *The Galapagos*, edited by R. I. Bowman. University of California Press, Berkeley.

Rick, C. M., 1983 Evolution of mating systems: evidence from allozyme variation, pp. 215–221 in *Genetics: New Frontiers* (Proc. XV Inter. Congress Genetics, New Delhi, December 1983). Oxford & IBH Publishing, New Delhi.

Rick, C. M., 1987 Genetic resources in *Lycopersicon*, pp. 17–26 in *Plant Biology*, Vol. 4, *Tomato Biotechnology*, edited by D. J. Nevins and R. A. Jones. Alan R. Liss Inc., New York.

Rick, C. M., and S. D. Tanksley, 1981 Genetic variation in *Solanum pennellii*: comparison with two other sympatric species. Plant Syst. Evol. **139:** 11–45.

Rick, C. M., J. F. Fobes and S. D. Tanksley, 1979 Evolution of mating systems in *Lycopersicon hirsutum* as deduced from genetic variation in electrophoretic and morphological characters. Plant Syst. Evol. **132:** 279–298.

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.

Soost, R. W., 1958 Progenies from sesquidiploid F_1 hybrids of *Lycopersicon esculentum* and *L. peruvanium.* J. Hered. **49:** 208–213.

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., 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. Pop. Biol. **41:** 237–254.

Tanksley, S. D., and F. Loaiza-Figueroa, 1985 Gametophytic self-incompatibility is controlled by a single locus on chromosome 1 in *Lycopersicon peruvanium.* Proc. Natl. Acad. Sci. USA **82:** 5093–5096.

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

van Ooijen, W., J. M. Sandbrink, M. Vrielink, R. Verkerk, P. Zabel and P. Lindhout, 1994 An RFLP map of *Lycopersicon peruvanium.* Theor. Appl. Genet. **89:** 1007–1013.

Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Pop. Biol. **7:** 256–276.

Communicating editor: G. B. Golding