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 1Section 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. chasmianii, 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 (Tankley 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 F2; population of the interspecific cross L. esculentum × L. pennellii. It contains more than 1000 markers that are distributed over 1276 cM (Tankley 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 chromosome 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 (Augadé et al. 1989; Stephan and Langley 1989), while the same regions exhibit normal amounts...
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) (Chet 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) RFLP 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 pimpinelliflorum), 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, Ï†s) 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),

\[ Ï†s = \sum_{l} \sum_{j} \sum_{i} \sum_{r} \frac{d_{ijr}}{d_{ljr}}, \]

where \( d_{ijr} = (2/w) \ln(G_{ijr}) \). \( G_{ijr} \) solves the equation \( G_{ijr} = 4F_{ir} (3 - 2G_{ijr}) \), where \( F_{ir} = m_{ir}(m_{ir} + m_{sr}); m_{ir} \) and \( m_{sr} \) are the number of fragments generated by restriction enzyme \( r \) in individuals \( i \) and \( j \) of species \( s \) at locus \( l \), while \( m_{ir} \) is the number of shared fragments. \( T_{rl} \) is the total number of comparisons for enzyme \( r \), species \( s \), and locus \( l \); \( w \) is the number of nucleotides in the recognition sequence of the restriction enzyme \( [w = 6 \text{ for the restriction enzymes in}\ Miller \text{ and Tanksley (1990):} \text{DraI, EcoRI, HindIII, EcoRI, and XbaI}]. \)

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, \( d_{ps} \), for each locus \( l \) over three apparently independent, evolutionary paths: esculentum to pimpinelliflorum, 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,

\[ d_{ps} = \frac{1}{E_{ljps}} \sum_{j} \sum_{i} \sum_{r} \frac{d_{ijr}}{d_{ljr}}, \]

where \( d_{ljps} \) is estimated as above, but now both individual \( i \) in one species and individual \( j \) in the other species of the phylogenetic path \( p \). \( E \) is the number of enzymes used, \( l_{p} \) 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,

\[ \text{Rap}_{ps} = \frac{\sum_{p} \sum_{l} \sum_{j} d_{ljps}}{(1/L) \sum_{l} \sum_{j} d_{ljps}}, \]

and average relative divergence,

\[ \text{Rap}_{ps} = \frac{1}{P} \sum_{p} \frac{\sum_{l} \sum_{j} d_{ljps}}{(1/L) \sum_{l} \sum_{j} d_{ljps}}. \]

The rescaled nucleotide diversities \( \Pi = \hat{\tau}_{sl} \hat{c}_{sl} \) are used in the analysis of covariance (analyses using the other rescaling factors, \( \hat{c}_{sl} \), yield similar results, not shown).

Analysis of covariance: The model of analysis of covariance for the crossing-over per physical length and species effects
This map is also useful for other Lycopersicon species 3 of which are polymorphic within the regions immediately adjacent to the telomeres. A total of 64 restriction sites over this length of DNA, as little or no structural variation among their karyotypes. One of the 3 polymorphic sites is a replacement reduced relative to that in the distal portions. Crossing-over per physical length, R, is the density of RN/μm estimated for locus i (i = 1, . . . , 36), and εi is an error effect. We conduct two types of analyses with this model: a “parametric analysis” of IIi in terms of estimated R and a “ranked analysis” (same model with ranked observations and ranked R values).

**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, \( \hat{a}_{sl} \) and \( \hat{a}_{d} \). Finally, Table 1 shows the estimates of IIi (the average number of differences per site within species, \( \hat{p}_{20} \), rescaled by the relative average divergence, \( \hat{a}_{sl} \)). Estimates of the unscaled nucleotide diversity (and those rescaled by \( \hat{a}_{d} \)) can be obtained by appropriate multiplication.

Table 2 presents the results of the analyses of covariance IIi in terms of species and R. In both the analysis of IIi and the analysis of ranked IIi in terms of ranked R, there is strong support for a species effect (P < 0.0001). A positive R effect on IIi is also strongly supported in both analyses. There is no support for an interaction effect, R × species, i.e., for separate slopes. The estimated slopes (of the species-specific regression of IIi on R) vary considerably among species, predicting between 9 and 246% (estimated mean, 54%) differences in nucleotide diversity (relative to the species average, \( \bar{II} \)) over the total range of crossing-over values (i.e., between 0 and 0.22 RN/μm). DNA sequence polymorphism is substantially lower (average IIi, \( \bar{II} = 0.0042 \)) in selfing species (L. parviflorum, chersanii, and esculentum) than in the partially outcrossing species, L. chmielewski and pimpinellifolium (\( \bar{II} = 0.0069 \)) and lower still than in those species with mating self-incompatibility, L. hirsutum, peruvianum, LA2150, and pennellii (\( \bar{II} = 0.0175 \)). This conclusion corroborates previous allozyme studies (Rick 1983; Doebley 1989; Breto 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 L. peruvianum population. The sucr gene is located in the centromeric region of chromosome 3. Based on its position (genetic position = 55.6 (Ch et al. 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 $R$ ($RN/\mu m$) 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 $\pi$ (Nei 1987) and $\theta$ (Waterson 1975), we estimate the number, $L$, of silent sites surveyed as

$$L = 2\sum_{i} l_i,$$

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

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>RN/μm</th>
<th>α^0δ</th>
<th>α^0δ</th>
<th>L. cheesmani</th>
<th>L. chilenense</th>
<th>L. esculentum</th>
<th>L. hirsutum</th>
<th>L. peruvianum</th>
<th>L. pimpinellifolium</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG78:</td>
<td>1</td>
<td>0.064</td>
<td>1.924</td>
<td>2.252</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
<td>0.000</td>
</tr>
<tr>
<td>TG24:</td>
<td>1</td>
<td>0.192</td>
<td>1.702</td>
<td>1.899</td>
<td>0.000</td>
<td>0.002</td>
<td>0.003</td>
<td>0.020</td>
<td>0.007</td>
<td>0.004</td>
</tr>
<tr>
<td>CD10:</td>
<td>1</td>
<td>0.192</td>
<td>0.494</td>
<td>0.476</td>
<td>0.003</td>
<td>0.003</td>
<td>0.004</td>
<td>0.020</td>
<td>0.007</td>
<td>0.004</td>
</tr>
<tr>
<td>TG21:</td>
<td>1</td>
<td>0.101</td>
<td>0.654</td>
<td>0.866</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.022</td>
<td>0.044</td>
<td>0.001</td>
</tr>
<tr>
<td>TG77:</td>
<td>1</td>
<td>0.108</td>
<td>1.326</td>
<td>1.695</td>
<td>0.003</td>
<td>0.003</td>
<td>0.004</td>
<td>0.000</td>
<td>0.013</td>
<td>0.002</td>
</tr>
<tr>
<td>TG17:</td>
<td>1</td>
<td>0.101</td>
<td>1.341</td>
<td>1.182</td>
<td>0.000</td>
<td>0.001</td>
<td>0.004</td>
<td>0.014</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>TG27:</td>
<td>1</td>
<td>0.000</td>
<td>1.085</td>
<td>1.522</td>
<td>0.000</td>
<td>0.018</td>
<td>0.000</td>
<td>0.004</td>
<td>0.014</td>
<td>0.010</td>
</tr>
<tr>
<td>TG31:</td>
<td>2</td>
<td>0.038</td>
<td>0.466</td>
<td>0.606</td>
<td>0.000</td>
<td>0.000</td>
<td>0.051</td>
<td>0.000</td>
<td>0.025</td>
<td>0.018</td>
</tr>
<tr>
<td>CD53:</td>
<td>2</td>
<td>0.138</td>
<td>0.739</td>
<td>0.928</td>
<td>0.004</td>
<td>0.000</td>
<td>0.021</td>
<td>0.000</td>
<td>0.039</td>
<td>0.008</td>
</tr>
<tr>
<td>TG14:</td>
<td>2</td>
<td>0.147</td>
<td>1.103</td>
<td>1.064</td>
<td>0.002</td>
<td>0.003</td>
<td>0.017</td>
<td>0.000</td>
<td>0.011</td>
<td>0.005</td>
</tr>
<tr>
<td>TG40:</td>
<td>3</td>
<td>0.000</td>
<td>1.724</td>
<td>1.454</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
<td>0.009</td>
<td>0.002</td>
<td>0.006</td>
</tr>
<tr>
<td>CD59:</td>
<td>4</td>
<td>0.116</td>
<td>1.062</td>
<td>0.959</td>
<td>0.001</td>
<td>0.007</td>
<td>0.005</td>
<td>0.027</td>
<td>0.035</td>
<td>0.002</td>
</tr>
<tr>
<td>CD70:</td>
<td>4</td>
<td>0.100</td>
<td>1.115</td>
<td>1.287</td>
<td>0.000</td>
<td>0.010</td>
<td>0.005</td>
<td>0.009</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>CG2:</td>
<td>4</td>
<td>0.124</td>
<td>0.332</td>
<td>0.396</td>
<td>0.014</td>
<td>0.005</td>
<td>0.051</td>
<td>0.000</td>
<td>0.037</td>
<td>0.002</td>
</tr>
<tr>
<td>CG39:</td>
<td>4</td>
<td>0.158</td>
<td>1.415</td>
<td>1.194</td>
<td>0.001</td>
<td>0.007</td>
<td>0.029</td>
<td>0.000</td>
<td>0.041</td>
<td>0.003</td>
</tr>
<tr>
<td>CD64:</td>
<td>5</td>
<td>0.120</td>
<td>0.714</td>
<td>0.742</td>
<td>0.006</td>
<td>0.000</td>
<td>0.000</td>
<td>0.014</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td>TG69:</td>
<td>5</td>
<td>0.188</td>
<td>0.788</td>
<td>0.977</td>
<td>0.005</td>
<td>0.004</td>
<td>0.027</td>
<td>0.000</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>CD67:</td>
<td>6</td>
<td>0.022</td>
<td>1.207</td>
<td>1.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.005</td>
<td>0.021</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>CD54:</td>
<td>6</td>
<td>0.183</td>
<td>0.681</td>
<td>0.860</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.036</td>
<td>0.016</td>
<td>0.004</td>
</tr>
<tr>
<td>CD57:</td>
<td>7</td>
<td>0.156</td>
<td>0.729</td>
<td>0.780</td>
<td>0.002</td>
<td>0.011</td>
<td>0.002</td>
<td>0.022</td>
<td>0.007</td>
<td>0.010</td>
</tr>
<tr>
<td>TG61:</td>
<td>7</td>
<td>0.196</td>
<td>0.580</td>
<td>0.724</td>
<td>0.006</td>
<td>0.001</td>
<td>0.029</td>
<td>0.000</td>
<td>0.043</td>
<td>0.012</td>
</tr>
<tr>
<td>CD60:</td>
<td>8</td>
<td>0.165</td>
<td>1.140</td>
<td>0.854</td>
<td>0.000</td>
<td>0.000</td>
<td>0.028</td>
<td>0.004</td>
<td>0.025</td>
<td>0.003</td>
</tr>
<tr>
<td>TG45:</td>
<td>8</td>
<td>0.081</td>
<td>0.392</td>
<td>0.340</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.015</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>TG18:</td>
<td>9</td>
<td>0.169</td>
<td>0.684</td>
<td>0.526</td>
<td>0.000</td>
<td>0.000</td>
<td>0.011</td>
<td>0.032</td>
<td>0.014</td>
<td>0.016</td>
</tr>
<tr>
<td>CD77:</td>
<td>10</td>
<td>0.216</td>
<td>1.700</td>
<td>1.571</td>
<td>0.000</td>
<td>0.002</td>
<td>0.014</td>
<td>0.000</td>
<td>0.016</td>
<td>0.008</td>
</tr>
<tr>
<td>TG12:</td>
<td>10</td>
<td>0.058</td>
<td>0.744</td>
<td>0.563</td>
<td>0.000</td>
<td>No data</td>
<td>No data</td>
<td>No data</td>
<td>No data</td>
<td>0.009</td>
</tr>
<tr>
<td>CD72:</td>
<td>10</td>
<td>0.192</td>
<td>0.895</td>
<td>0.792</td>
<td>0.010</td>
<td>0.018</td>
<td>0.007</td>
<td>0.021</td>
<td>0.031</td>
<td>0.005</td>
</tr>
<tr>
<td>CD05:</td>
<td>12</td>
<td>0.124</td>
<td>0.596</td>
<td>0.730</td>
<td>0.004</td>
<td>0.000</td>
<td>0.001</td>
<td>0.009</td>
<td>0.034</td>
<td>0.007</td>
</tr>
<tr>
<td>TG63:</td>
<td>10</td>
<td>0.151</td>
<td>2.257</td>
<td>1.547</td>
<td>0.000</td>
<td>0.000</td>
<td>0.007</td>
<td>0.006</td>
<td>0.028</td>
<td>0.008</td>
</tr>
<tr>
<td>CD17:</td>
<td>11</td>
<td>0.207</td>
<td>1.454</td>
<td>1.157</td>
<td>0.001</td>
<td>0.004</td>
<td>0.022</td>
<td>0.025</td>
<td>0.023</td>
<td>0.008</td>
</tr>
<tr>
<td>TG46:</td>
<td>11</td>
<td>0.166</td>
<td>0.336</td>
<td>0.334</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>0.049</td>
<td>0.044</td>
<td>0.029</td>
</tr>
<tr>
<td>TG26:</td>
<td>11</td>
<td>0.146</td>
<td>1.123</td>
<td>1.206</td>
<td>0.012</td>
<td>0.006</td>
<td>0.011</td>
<td>0.020</td>
<td>0.014</td>
<td>0.008</td>
</tr>
<tr>
<td>TG30:</td>
<td>11</td>
<td>0.208</td>
<td>1.199</td>
<td>1.250</td>
<td>0.001</td>
<td>0.003</td>
<td>0.004</td>
<td>0.021</td>
<td>0.031</td>
<td>0.008</td>
</tr>
<tr>
<td>CD19:</td>
<td>12</td>
<td>0.192</td>
<td>1.187</td>
<td>1.383</td>
<td>0.000</td>
<td>0.007</td>
<td>0.004</td>
<td>0.023</td>
<td>0.018</td>
<td>0.020</td>
</tr>
</tbody>
</table>

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, α^0δ and α^δ, for the loci. The remaining columns contain Psl, the estimate of π̂ rescaled by α^0δ.
TABLE 2
Analysis of covariance of the average number of differences per site within species, \( \bar{\pi}_n \), rescaled by the relative average divergence, \( \bar{e} \).

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>F ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>8</td>
<td>0.00293</td>
<td>4.82</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>R</td>
<td>1</td>
<td>0.00086</td>
<td>11.35</td>
<td>0.0009</td>
</tr>
<tr>
<td>Species ( \times ) R</td>
<td>8</td>
<td>0.00059</td>
<td>0.98</td>
<td>0.45</td>
</tr>
</tbody>
</table>

The results for the parametric analysis on each line are presented, while below and italicized are the results for the ranked \( \Pi_s \) and ranked \( R \) analysis.

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

The \( \text{Sod}-2 \) 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/\( \mu \text{m} \) (see Figure 1). Our method allows the survey of variation in a region of roughly 3300 bp of the 3.5-kb \( \text{Sod}-2 \) transcriptional unit. Due to the high AT-content of the introns [including that all restriction site polymorphisms in exons are present in genomic regions of low crossing-over was 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 \( \Pi_s \) for each species (as a function of \( \Pi_s \)). 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 \( \text{su}c\text{r} \) and \( \text{Sod}-2 \) genes in \( L. \text{peruvianum} \) 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 \( D. \text{ananasae} \) (Stephan and Langley 1989), \( D. \text{melanogaster} \) (Aguadé et al. 1989; Begun and Aquadro 1992), \( D. \text{simulans} \) (Begun and Aquadro 1991; Berry et al. 1991), \( D. \text{maurantina} \), and \( D. \text{schellia} \) (Hilton et al. 1994)] and in mice (Nachman 1997), has been confirmed in a relatively distant relative, \( L. \text{acer} \). 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 \( \Pi_s \), corrected for the species average (over loci), i.e., \( \Pi_s + (\Pi - \Pi_s) \). 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 \( L. \text{acer} \) 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 \( \Pi_s \) on \( R \) in each species from the same analysis of covariance model. As can be seen in Figure 3 those species with intermediate \( \Pi_s \). \( L. \text{hirsutum} \) and \( \text{pennellii} \) have the largest slopes; in contrast, completely selfing species (with the lowest \( \Pi_s \)) and the
DNA Polymorphism in Lycopersicon

most consistent outbreeder (L. peruvianum) have shallower slopes (with the highest $\Pi$). 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 nonlinearity of the relationship.

![Figure 2](image_url)

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

![Figure 3](image_url)

**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, $\bar{\Pi}$, (in units of corrected nucleotide differences per site). The open bars represent the selfers (in order, L. parviflorum, chesmanii, 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 $\Pi$, while the two species (L. pennellii and hirsutum) with intermediate $\Pi$ 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 $\Pi$, 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 T\(T_i\) 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 T\(T_i\) and the slope of the species-specific regression of X 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


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


DNA Polymorphism in Lycopersicon


Communicating editor: G. B. Golding