Switch in Codon Bias and Increased Rates of Amino Acid Substitution in the Drosophila saltans Species Group

Francisco Rodríguez-Trelles, Rosa Tarrío and Francisco J. Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525

Manuscript received July 19, 1998
Accepted for publication June 1, 1999

ABSTRACT

We investigated the nucleotide composition of five genes, Xdh, Adh, Sod, Per, and 28S rRNA, in nine species of Drosophila (subgenus Sophophora) and one of Scaptodrosophila. The six species of the Drosophila saltans group markedly differ from the others in GC content and codon use bias. The GC content in the third codon position, and to a lesser extent in the first position and the introns, is higher in the D. melanogaster and D. obscura groups than in the D. saltans group (in Scaptodrosophila it is intermediate but closer to the melanogaster and obscura species). Differences are greater for Xdh than for Adh, Sod, Per, and 28S rRNA, which are functionally more constrained. We infer that rapid evolution of GC content in the saltans lineage is largely due to a shift in mutation pressure, which may have been associated with diminished natural selection due to smaller effective population numbers rather than reduced recombination rates. The rate of GC content evolution impacts the rate of protein evolution and may distort phylogenetic inferences. Previous observations suggesting that GC content evolution is very limited in Drosophila may have been distorted due to the restricted number of genes and species (mostly D. melanogaster) investigated.

SUEOKA (1962; see also Freese 1962) has postulated that if \( u \) is the rate of conversion \( A/T \rightarrow G/C \) (either \( A \) or \( T \) to either \( G \) or \( C \) and \( v \) is the reciprocal rate, the \( G + C \) composition of a genome will evolve until an equilibrium is reached, with the \( G + C \) frequency simply determined by \( P = u/(u+v) \). The rate of conversion \( A/T \leftrightarrow G/C \) is a joint consequence of selective constraints (which Sueoka often assumed to be small) and mutation pressure. One or the other of the values of \( P \) and \( u/v \) has been referred to as the GC pressure, mutational pressure, or mutation bias (e.g., Gillespie 1991; Li 1997, p. 401) and the observed frequency of \( G+C \) as the GC content. Sueoka (1962, 1988, 1992, 1993) pointed out that when two organisms differ appreciably in GC content, their proteins will differ in primary structure, even in the case of enzymes with identical function, with the exception of the active site that would be conserved owing to functional constraints.

The effect of GC mutation bias on changing GC content has been shown, for example, in a mutator strain, mutT, of Escherichia coli with an elevated mutation rate of \( A/T \rightarrow G/C \) (Cox and Yanofsky 1967). More generally, fluctuating mutation bias has been invoked as a major factor to explain properties of DNA base composition in bacteria and other microorganisms as well as in mitochondrial and nuclear genomes (Sueoka 1962, 1988; Muto and Osawa 1987; reviewed in Li 1997, chaps. 13 and 14). The significance of mutation relative to selection has been established (1) by comparing the regressions of total GC content on the GC content in the three different codon positions in bacteria (Jukes and Bhushan 1986; Muto and Osawa 1987; Sueoka 1988); (2) by the correlation between total GC content and the base composition of flanking gene regions, introns, and silent coding sites in mammals (Ikekura 1985; D’Onofrio et al. 1991); and (3) by the accumulation of AT in the coding and noncoding regions of insect mitochondrial DNA (e.g., Crozier and Crozier 1993). Variation in mutation bias has further been related to switches in codon usage patterns in bacteria (Shields 1990; Li 1997); to variation in the amino acid composition of bacterial proteins (Sueoka 1962; Li 1997; Gu et al. 1998); and to variation in insect mitochondrial (Jukes and Bhushan 1986; Jermini et al. 1994) and mammalian nuclear genomes (D’Onofrio et al. 1991; Collins and Jukes 1993). But it has also been argued that GC content variation may be a consequence of natural selection toward an optimal GC value (Gillespie 1991, p. 85; D’Onofrio et al. 1991).

Intraspecific variation in GC content along the nuclear genome is quite large in Drosophila (Carulli et al. 1993; Kliman and Hey 1994; Akashi et al. 1998; Kliman and Eyre-Walker 1998), but the mutational equilibrium of the genome is thought to have remained essentially constant during the diversification of the genus because (1) the base composition of introns is generally very low \( G+C \) (Shields et al. 1988; Moriyama and Hartl 1993); (2) the pattern of codon usage is fairly homogeneous across species except when differ-
ences can be accounted for by changes in the natural selection pressure (Akashi 1995, 1996; Akashi and Schaeffer 1997; but see Powell 1997, p. 376); and (3) estimates of the pattern of point mutation reflect considerable stability over evolutionary time (Petrov and Hartl 1999). Previous studies, however, have largely been restricted to two species, Drosophila melanogaster and D. pseudoobscura, of the Sophophora subgenus, and D. virilis for the subgenus Drosophila, all three of which have quite similar overall base composition (reviewed in Powell 1997). Studies that have included a larger taxonomic spectrum have focused on the coding region of Adh, Sod, and the 28S rRNA (reviews in Powell and DeSalle 1995; Powell 1997), regions dominated by strong functional constraints.

In this study, we investigate five gene regions under different degrees of functional constraint, namely Xdh, Adh, Sod, Per, and two domains (D1 and D2) of the 28S rRNA untranslated region in the Sophophora subgenus, including several species of the little-investigated saltans group as well as the obscura and melanogaster groups. Our results suggest that GC mutation pressure has had a major influence on the molecular evolution of Drosophila, with implications for theories about the evolution of codon bias.

MATERIALS AND METHODS

Species and sequences: The Xdh region was investigated in nine species of Drosophila and in Scaptodrosophila lebanonensis, which was used as an outgroup. Six species belong to the saltans group: D. saltans, D. prosaltans, D. necordata, D. marginata, D. sturtevanti, and D. subobscura. The Xdh coding sequence of D. subobscura is from a strain from Helsinki, Finland, kept in our laboratory, as is the strain of S. lebanonensis. Xdh sequences of D. melanogaster, D. pseudoobscura, and D. subobscura (only intron II) were available from the literature (GenBank accession nos. Y00307, M33977, and Y06237, respectively). The Xdh gene region investigated includes about half of exon II (371 codons), intron II (30-60 bp in most cases), and most of exon III (324 codons), or ~52% of the Xdh coding region. Details about the amplification and sequencing primers and strategy can be found in Tarrío et al. (1998).

The sequences of Adh, Sod, Per, and 28S rRNA were obtained from the literature. The Adh sequences consist of 135 codons of exon II, and include D. saltans (GenBank accession no. AF045113), D. prosaltans (AF045119), D. marginata (AF045124), D. necordata (AF045120), D. sturtevanti (AF045114), D. subsaltans (AF045117), D. melanogaster (X78384), D. pseudoobscura (U64560), D. subobscura (X55391), and S. lebanonensis (X54814). The Sod sequences include D. saltans, D. melanogaster, D. pseudoobscura, D. subobscura, and S. lebanonensis (Kwiatowski et al. 1994), and consist of 145 codons (only 114 codons in the two obscura species), plus 321-725 bp of the intron I (not available for S. lebanonensis). The Per sequences include D. saltans (L06336), D. melanogaster (M13653), and D. pseudoobscura (X13978), and stretch 51 codons of the Thr-Gly domain that can be unambiguously aligned. The 28S rRNA sequences of D. prosaltans, D. marginata, D. necordata, D. sturtevanti, D. melanogaster, and D. pseudoobscura (Pelándakis and Solignac 1993) consist of 541 bp corresponding to the two divergent domains D1 and D2.

Nucleotide composition and codon-usage bias: Sequences were aligned using the CLUSTAL W (v. 1.5) program (Thompson et al. 1994). Chi-square statistics were used to test for random use of codons within amino acid classes and for homogeneity of codon usage among species. Deviation from a uniform use of codons was measured with the effective number of codons (ENC) statistic (Wright 1990). ENC ranges from 20, when only one codon is used for each amino acid, to 61, when all synonymous codons are used equally. ENC is quite unaffected by length differences when genes are >150 codons (Wright 1990). In addition, we use the frequency of optimal codons (Fop) index (Ikemura 1985) with the set of major codons defined by Akashi (1995) as a measure of departure from optimal codon usage in D. melanogaster.

Classification of amino acids: We classified amino acids into three groups, according to codon GC content (Jukes and Bshuan 1986; see also Li 1997). Group I consists of codons with a high GC: alanine (A), glycine (G), proline (P), and tryptophan (W); (e.g., alanine is encoded by GCU, GCC, GCA, or GCG). Group II consists of codons with an intermediate GC content: cysteine (C), aspartic acid (D), glutamic acid (E), histidine (H), glutamine (Q), serine (S), threonine (T), and valine (V) (e.g., aspartic acid is encoded by either GAU or GAC). Group III consists of codons with a low GC content: phenylalanine (F), isoleucine (I), lysine (K), methionine (M), asparagine (N), and tyrosine (Y) (e.g., phenylalanine is encoded by either UUU or UUC). Arginine (R) and leucine (L) are not included in these groups, because R is encoded by an intermediate (AGA, AGG) as well as a high-GC codon family (CGU, CGC, CGA, CGG), and L is encoded by a low (UUA, UUG) and an intermediate GC (CUU, CUC, CUA, CUG) codon family. If amino acid frequencies are impacted by nucleotide composition, f(I), the frequency of group I, will increase and f(II) will decrease as GC content increases, while f(II) will change little.

Directional mutation pressure and amino acid composition: As a measure of the intensity of the GC/AT mutation pressure on the gene regions investigated, we used the GC content at fourfold degenerate sites (GC4), because all nucleotide changes at these sites are synonymous. GC4 may be affected by codon use bias, but it is better for this purpose than the average GC content of a gene, because this is strongly impacted by the functional constraints of the proteins (Sueoka 1988; Li 1997). We use two additional measures of the GC/AT mutation pressure: the GC content of intron II (GC2) and the GC content of synonymous sites (GCsyn); see also Jermiin et al. 1994). All three measures are strongly correlated (r = 0.89, GC4 vs. GC2; f = 0.98, GC4 vs. GCsyn, and f = 0.91, GC4 vs. GCsyn; P < 0.001 in all three cases for Xdh). Using one or the other of them yields essentially the same results.

Species are part of a hierarchically structured phylogeny; therefore, treating them as statistically independent observations (Fel senst ein 1985) can lead to overestimation of the nominal significance level in hypothesis testing. To circumvent phylogenetic inertia we have studied the association between GC4, f(I), f(II), or f(III) by means of Fel senst ein's (1985) pairwise independent contrast test. Given a rooted phylogenetic tree with n species, a total of n − 1 independent contrasts can be obtained for each pair of characters X (e.g., the GC4) and Y (e.g., the amino acid frequency). Because little information is available for the saltans group of Drosophila, the contrast test was carried out with the tree inferred from the Xdh sequences (Figure 2). This can result in some circularity, because the same data are also used for investigating the relationship between GC4 and amino acid composition (Fel senst ein 1985). The substantial length of the sequences and the robustness of the maximum-likelihood method employed for inferring the tree, however, mitigate this potential problem.
closer. The GC content of D. melanogaster lineage and increased in the (gaster 1 saltans
obscura obscura is intermediate between the two other
to the values typical of genomes considered AT rich
saltans obscura obscura
package PHYLIP 3.5 (Felsenstein
Throckmorton and that proposed by
Moreover, we use two different topologies: the ML topology
Throckmorton and Magalhães (1962), which differ substantially in the arrangement of spe-
cial within the saltans group. However, using one or the other
topology yields essentially the same results. Contrast tests were
performed with the CONTRAST program in the computer
package PHYLIP 3.5 (Felsenstein 1993).

RESULTS

Xdh nucleotide composition and codon-use bias: Ta-
ble 1 shows the Xdh GC content for each codon position,
the fourfold degenerate sites, intron II, and intron B. The
largest compositional differences occur between the
obscura group (two species) and the saltans group
(six species). The obscura average GC content value for
the first (62.2%), second (42.8%), and third (78.3%)
position is typical of GC-rich genomes, while the saltans
averages, respectively 54.0, 41.6, and 41.4%, are closer
to the values typical of genomes considered AT rich
(Muto and Osawa 1987; Lloyd and Sharp 1993). D. melanogaster is intermediate between the two other
groups but closer to the obscura group (conspicuously
in the third position), to which it is also phylogenetically
closer. The GC content of D. melanogaster is also closest
to the outgroup S. lebanonensis, which is phylogenetically
equally distant from all the Sophophora species. If we
were to infer that the GC content of S. lebanonensis
remains similar to the ancestral composition, we could
conclude that after the saltans divergence from (melan-
gaster + obscura), the GC content decreased in the saltans
lineage and increased in the (melanogaster + obscura)
lineage. Be the ancestor content as it might, it is the
case that GC content of the saltans group has become
increasingly divergent from the obscura group and also,
but to a lesser extent, from melanogaster.

If a given locus experiences different mutation pres-
sures in different lineages, then positive correlations
should be observed between GC composition of the
codons and the introns (assuming that intron base com-
position reflects the mutational equilibrium of the ge-
nome). Intron B has arisen in the saltans lineage by a
duplication of intron II (Tarrio et al. 1998) and is
most divergent in GC content (see Table 1), hence it
is excluded from consideration. From Felsenstein’s
(1993) contrast test, intron II GC content correlates
significantly with the first (t = 0.76, P < 0.01) and
the third (t = 0.68, P = 0.04) codon positions. This
is apparent in Table 1, where we see that the GC content
of intron II is conspicuously lower in the saltans group
than in the obscura group (Mann-Whitney U-test, P <
0.05), as is the GC content in the third and (less so)
first codon positions. The G + C content of saltans
introns B and II is significantly lower (P < 0.001 and
P < 0.05, respectively) than the average G + C content
of the D. melanogaster introns (~40%; Shields et al.
1988; Moriyama and Hartl 1993), commonly assumed
to reflect the Drosophila mutational equilibrium (see
Akashi 1996).

Note that positive correlations between intron and
exon GC content are not necessarily indicative of vary-
ing mutation pressures that influence all nucleotide
positions alike. For example, Kliman and Eyre-Walke

<table>
<thead>
<tr>
<th>table1</th>
<th>GC content and codon-use bias in the Xdh gene of Drosophila (subgenus Sophophora) and Scaptodrosophila</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>First</td>
</tr>
<tr>
<td>Saltans group</td>
<td></td>
</tr>
<tr>
<td>D. saltans</td>
<td>57.5</td>
</tr>
<tr>
<td>D. prosaltans</td>
<td>57.4</td>
</tr>
<tr>
<td>D. emarginata</td>
<td>57.6</td>
</tr>
<tr>
<td>D. neocecorata</td>
<td>56.9</td>
</tr>
<tr>
<td>D. sturtwanti</td>
<td>57.5</td>
</tr>
<tr>
<td>D. subsaltans</td>
<td>56.3</td>
</tr>
<tr>
<td>Obscura group</td>
<td></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>59.3</td>
</tr>
<tr>
<td>D. pseudoobscura</td>
<td>62.5</td>
</tr>
<tr>
<td>D. subsaltans</td>
<td>61.9</td>
</tr>
<tr>
<td>Genus Scaptodrosophila</td>
<td></td>
</tr>
<tr>
<td>S. lebanonensis</td>
<td>58.4</td>
</tr>
</tbody>
</table>

The percentage GC content is given separately for each of the three codon positions (first, second, third); the third position in all fourfold synonymous codons; intron II; and intron B (which is only present in the saltans group). ENC is the effective number of codons, which may range from 20 to 61. Intron lengths are given in parentheses.
Figure 1.—Frequencies of amino acid groups I (black), II (white), and III (gray; high, medium, and low GC content, respectively) on the frequency of fourfold degenerate codons (GC4) for Xdh. Each dot in each group represents 1 of the 10 species studied.

(1998) found a consistent decline in GC content along the genes of D. melanogaster, which is reflected in the introns by a change in G, while it is due mainly to C in third codon positions. From the correlation values obtained by us, however, G and C appear to contribute equally to the interspecific variation in GC content in first (r = 0.960 and r = 0.940; correlation of GC with G and C, respectively) and third (r = 0.995 and r = 0.997) codon positions, and in introns (r = 0.712 and r = 0.890). Factors others than those shaping the GC content along genes (Kliman and Eyre-Walker 1998) must thus be responsible for the observed GC content variation across species.

Table 1 gives the ENC values. Consistent with previous results, there is little codon bias across all species in this study. Under the major codon preference model, this is expected for a region that is transcribed at very low levels (Riley 1989). Nevertheless, tests separately carried out for each amino acid in each species indicate that codon use in different species groups is not random for most amino acids (results not shown). Within a species group, sequence divergence is too low to detect any differences in codon use that may exist (χ² = 49 with 57 d.f., and χ² = 120 with 290 d.f., respectively, for Obscura and saltans groups; neither significant).

Xdh correlation between nucleotide and amino acid composition: Figure 1 shows that, as expected (see materials and methods), the high-GC amino acids (group I) are less used by species with low GC content (the saltans group), while the opposite is the case for group III (low GC content) amino acids, and less so for group II amino acids. Thus, the frequency of group I, f(I), is 21.7% (14.9% when only variable sites are considered) in D. subsaltans (GC4 = 39.4%), but it increases to 23.8% (22.2% of variable sites) in D. pseudobscura (GC4 = 77.2%). The correlation between f(I) and GC4 is significant by the contrast test (r = 0.68, P = 0.04). In contrast, f(III) is 25.5% (29% of variable sites) in D. subsaltans, but only 23.5% (23.2% of variable sites) in D. pseudobscura (r = −0.61, marginally significant P = 0.08). The association between f(II) and GC4 is not significant (r = −0.28, P = 0.47).

We have also conducted 2 × 2 chi-square tests for the null hypothesis that in the three Sophophora lineages there is no association between species group and the number of replacements that occurred toward GC-coded amino acids vs. those that occurred toward AT-coded amino acids. Unambiguous changes were estimated by maximum parsimony on the topology shown in Figure 2 (using the character trace function of McClade 3.0; Maddison and Maddison 1992), which are presented along the branches. All the saltans species but D. emarginata have undergone a significantly higher number of changes toward AT-coded amino acids (P < 0.05) than D. pseudobscura. Compared to D. subsaltans, differences are significant (P < 0.05) for D. saltans, D. prosaltans, and D. neocordata, and nearly so (P = 0.053) for D. sturtevanti. Pooling the total number of changes along the obscura and saltans lineages, the differences between both groups are highly significant (P < 0.001); moreover, the differences remain significant when the total number of changes in the obscura group is considered in conjunction with those that occurred in D. melanogaster (P < 0.05).

The topology in Figure 3 is largely consistent with previous studies (Kwitowski et al. 1994, 1997; Russo et al. 1995; Tatar enkov et al. 1999), except that we add several saltans species, which are closely related to the willistoni group. The topology of the species of the saltans group, based primarily on biogeographic data, places D. saltans and D. prosaltans as recently derived taxa within the group, and D. emarginata and D. neocordata as the oldest taxa (Throckmorton and Magalhaes 1962; see also O’Grady et al. 1998). When this topology is used, the correlations from the independent contrast tests and the chi-square tests remain significant.

We have not included two amino acids in the previous analyses: leucine, because it is encoded by a low-GC codon family (UUA, UUG) and an intermediate-GC codon family (CUU, CUC, CUA, CUG); and arginine because it is encoded by an intermediate-GC (AGA, AGG) and a high-GC codon family (CGU, CGC, CGA, CGG). In any case, the frequency of leucine in Xdh is not correlated with GC content (r = −0.22, P = 0.58), because the frequency changes in the two codon families largely cancel each other; that is, low-GC species use codons UUA and UUG more frequently (r = −0.77, P = 0.01) than codons CUU, CUC, CUA, and CUG (r = 0.76, P = 0.01). Arginine exhibits a similar pattern, except that the frequency of arginine increases with increasing GC4 (r = 0.59, P = 0.09), which occurs because high-GC codons (CGT, CGC, CGA, and CGG)
for arginine are more abundant than intermediate-GC codons (AGA and AGG).

**Xdh rates of substitution:** The relative-rate test is useful for comparing the substitution rates between a given pair of species (species 1 and 2 in Table 2) when the time since their split is not precisely known, but this time is the same for each pair-wise comparison within a set. We use the Xdh sequence of *S. lebanonensis* (species 3) as the outgroup. The values of $K (K_{1.3} - K_{2.3})$ in Table 2 represent the difference between the number of nonsynonymous substitutions per site (Wu and Li 1985) for lineages 1 and 2 after their divergence. If the value is negative, lineage 2 has evolved at a faster rate than lineage 1. We ignore synonymous substitutions because they are largely saturated and thus contain little information for the relative rate tests. The results in Table 2 indicate that Xdh has evolved at a faster rate in the saltans lineage than in the obscura or melanogaster

![Figure 2](image_url)

**Figure 2.**—Maximum-likelihood (ML) tree of the Xdh nucleotide sequences, obtained with the reversible model (Yang 1994; program PAML 1.3, Yang 1997), and allowing different nucleotide frequencies, transition/transversion rate ratios, assuming gamma-distributed rates among sites (eight rate categories), and different fixed rates at codon positions. Numbers above branches represent the unambiguous amino acid changes (AT-coded:GC-coded) estimated by maximum parsimony assuming the ML tree topology, with the character trace function of McClade 3.0 (Maddison and Maddison 1992). Numbers after the species names are the corresponding total changes along the branches.

![Figure 3](image_url)

**Figure 3.**—Frequency of optimal codons (Fop) values for the Xdh, Adh, Sod, and Per regions in the saltans group (solid arrows), *D. melanogaster* (white arrows), and the obscura group (gray arrows), plotted against the distribution of Fop values for 346 *D. melanogaster* genes (Sharp and Lloyd 1993).

**TABLE 2**

<table>
<thead>
<tr>
<th>Species 1</th>
<th>Species 2</th>
<th>$K$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. melanogaster</em></td>
<td><em>D. saltans</em></td>
<td>-0.0257**</td>
</tr>
<tr>
<td><em>D. prosaltans</em></td>
<td><em>D. saltans</em></td>
<td>-0.0226**</td>
</tr>
<tr>
<td><em>D. emarginata</em></td>
<td><em>D. saltans</em></td>
<td>-0.0175*</td>
</tr>
<tr>
<td><em>D. neocordata</em></td>
<td><em>D. saltans</em></td>
<td>-0.0175*</td>
</tr>
<tr>
<td><em>D. sturtevanti</em></td>
<td><em>D. saltans</em></td>
<td>-0.0172*</td>
</tr>
<tr>
<td><em>D. subsaltans</em></td>
<td><em>D. saltans</em></td>
<td>-0.0243**</td>
</tr>
<tr>
<td><em>D. pseudoobscura</em></td>
<td><em>D. saltans</em></td>
<td>-0.0381***</td>
</tr>
<tr>
<td><em>D. prosaltans</em></td>
<td><em>D. saltans</em></td>
<td>-0.0350***</td>
</tr>
<tr>
<td><em>D. emarginata</em></td>
<td><em>D. saltans</em></td>
<td>-0.0299***</td>
</tr>
<tr>
<td><em>D. neocordata</em></td>
<td><em>D. saltans</em></td>
<td>-0.0299***</td>
</tr>
<tr>
<td><em>D. sturtevanti</em></td>
<td><em>D. saltans</em></td>
<td>-0.0297***</td>
</tr>
<tr>
<td><em>D. subsaltans</em></td>
<td><em>D. saltans</em></td>
<td>-0.0367***</td>
</tr>
<tr>
<td><em>D. subobscura</em></td>
<td><em>D. saltans</em></td>
<td>-0.0211*</td>
</tr>
<tr>
<td><em>D. prosaltans</em></td>
<td><em>D. saltans</em></td>
<td>-0.0180*</td>
</tr>
<tr>
<td><em>D. emarginata</em></td>
<td><em>D. saltans</em></td>
<td>-0.0128*</td>
</tr>
<tr>
<td><em>D. neocordata</em></td>
<td><em>D. saltans</em></td>
<td>-0.0129*</td>
</tr>
<tr>
<td><em>D. sturtevanti</em></td>
<td><em>D. saltans</em></td>
<td>-0.0126*</td>
</tr>
<tr>
<td><em>D. subsaltans</em></td>
<td><em>D. saltans</em></td>
<td>-0.0197*</td>
</tr>
</tbody>
</table>

$K$ is the difference between the rates of species 1 and 2 when each is compared to Scaptodrosophila. A negative value indicates that species 2 has evolved faster than species 1. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
lineages. The tests pairing either D. pseudoobscura or D. melanogaster with the saltans group species are all significant, and the differences are consistently larger for the comparison with D. pseudoobscura. Comparisons between D. subobscura and the saltans species are significant in three cases.

A measure of the difference in rates between lineages is the ratio of the estimated substitution rates in each lineage (Gaut et al. 1992). Estimating this ratio for the second codon position data and averaging across saltans species, we see that the rate of nonsynonymous substitutions in Xdh is \( \sim 2.4 \) times faster in the saltans lineage than in the obscura lineage (\( \sim 3.15 \) and \( \sim 1.69 \) when the saltans lineage is paired with D. pseudoobscura and D. subobscura, respectively) and \( \sim 1.62 \) times faster than in D. melanogaster.

**Analysis of the Adh, Sod, Per, and 28S ribosomal RNA sequences:** We analyzed the Adh, Sod, and Per coding sequences in a similar fashion as those of Xdh, although for Sod and Per only the sequence of D. saltans is available for the saltans group. We also analyzed the base composition of the 28S ribosomal RNA untranslated region. Similar patterns emerge as with Xdh (Table 3). Across the Sophophora subgenus, the GC content in third and first codon positions of Adh, Sod, and Per is consistently lowest in the saltans species. For the more conserved 28S rRNA region and the second positions of Per, the pattern is the same, but the differences in base composition are less pronounced. Unlike the intron II of Xdh, the base composition of the Sod intron shows little variation: D. saltans has insignificantly less GC content than the two obscura species (chi-square test, \( P \sim 0.32 \)) and virtually the same as D. melanogaster. A closer inspection of this intron sequence with the program PRSS (W. R. Pearson, www.med.virginia.edu/~warp/csh197/prss.htm; default options used) reveals that it is substantially conserved. The PRSS program allows one to evaluate the significance of a pair-wise alignment by comparing its score against the empirical distribution of scores generated from 5000 random permutations of the sequences. While the intron II of Xdh renders nonsignificant alignments (except for some comparisons between the closely related species of the saltans group), the Sod intron of D. saltans can be aligned for most of its length with the introns of the distantly related D. pseudoobscura (\( P = 0.003 \)) and D. subobscura (\( P = 0.05 \)), and the latter two can be aligned with the intron of D. melanogaster (\( P = 10^{-7} \) and \( P = 0.02 \), respectively). Conservation of the Sod intron sequence over evolutionary time suggests that mutation bias is not the only factor influencing the base composition of this intron. It may be significant that this is the first intron in the Sod gene of Drosophila. Unlike downstream introns (e.g., intron II of Xdh), first introns are frequently larger, containing regulatory sequences, and their size covaries with the length of other elements of the host genes, including the leader, the coding region, and the 3' untranslated region (Maroni 1996), suggesting shared constraints among all of them.

Variation across the three Sophophora groups in the magnitude and pattern of codon bias in Adh, Sod, and Per (Table 3; Per contains few codons to calculate ENC) is similar to the pattern of the Xdh gene. Codon bias is least in D. saltans; for each species individually, Adh is more biased than Sod, and both genes are substantially more biased than Xdh in D. melanogaster. Averaged across the two species, Adh, Sod, and Xdh show fairly similar bias in obscura. In saltans, ENC values for the three genes parallel those of D. melanogaster.

ENC measures unequal usage regardless of the direction of the bias. It is interesting to know whether lower values of ENC for Adh and Sod than for Xdh in saltans are due to a greater use of optimal codons or, on the contrary, reflect an increased bias toward A- and T-ending codons. To ascertain this, we computed the Fop (Ikemura 1985) for Adh, Sod, and Xdh, assuming as major codons those of D. melanogaster as defined in Akashi (1995). Fop can be calculated for short sequences, which allows consideration of the Per region in the analysis. Only homologous codons that encode the same amino acid in all species are examined. Fop may range between 0 and 1, with closer values to 0 indicating greater similarity to the optimal codon use in D. melanogaster; i.e., less bias toward A- and T-ending codons. Figure 3 plots the Fop values for the four gene regions in D. saltans, D. melanogaster, and the two obscura species (averaged) against the distribution of Fop values of 346 D. melanogaster genes (compiled by Sharp and Lloyd 1993). All four genes reflect a dramatic reduction in the Fop values in D. saltans (\( P < 10^{-4} \) except for the short Per sequences; \( P = 0.29 \); \( 2 \times 2 \) chi-square tests). Thus, for example, in D. melanogaster Adh is among the 10% most biased genes. However, the Adh Fop value of D. saltans falls within the 10% lowest of D. melanogaster, and for Xdh this number is even more extreme (2.6%). Across loci, the amount of Fop decrease in saltans varies depending on which of the most biased species is compared (Figure 3). Consideration of the average Fop values over D. melanogaster and the two obscura species indicates, however, that all loci have experienced an equivalent reduction (by \( \sim 40\% \)) in major codon use in saltans.

There is no significant association between amino acid composition and GC content in either Adh, Sod, or Per. Interestingly enough, however, the Per region exhibits exactly the same pattern shown by Xdh: a higher proportion of AT-coded amino acids in D. saltans (21.7%) than in D. melanogaster (19.7%) and D. pseudoobscura (15.7%), and a lower number of GC-coded amino acids in the former species (25.5 vs. 27.4%). As to the intermediate amino acids, D. saltans has the same number as D. melanogaster (47%) and less than D. pseudoobscura (49%). With respect to Adh, the proportion of AT-coded amino acids is lower in D. saltans (29.9%) than
### TABLE 3

GC content for four genes and codon-use bias in Adh and Sod

<table>
<thead>
<tr>
<th>Species</th>
<th>Adh</th>
<th>Sod</th>
<th>Per</th>
<th>28SrRNA</th>
<th>Adh</th>
<th>Sod</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgenus Sophophora</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. saltans</td>
<td>37.3</td>
<td>29.6</td>
<td>37.3</td>
<td>52.1</td>
<td>59.6</td>
<td>43.8</td>
</tr>
<tr>
<td>D. prosaltans</td>
<td>36.2</td>
<td>29.8</td>
<td>39.0</td>
<td>52.1</td>
<td>59.6</td>
<td>43.8</td>
</tr>
<tr>
<td>D. emarginata</td>
<td>36.7</td>
<td>30.1</td>
<td>35.1</td>
<td>45.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D. neocordata</td>
<td>37.3</td>
<td>30.1</td>
<td>37.3</td>
<td>47.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D. sturtevanti</td>
<td>37.3</td>
<td>28.5</td>
<td>30.2</td>
<td>37.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D. subsaltans</td>
<td>37.3</td>
<td>30.1</td>
<td>37.1</td>
<td>49.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Subgenus melanogaster</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>39.6</td>
<td>29.1</td>
<td>60.4</td>
<td>84.0</td>
<td>61.6</td>
<td>44.6</td>
</tr>
<tr>
<td>D. pseudoobscura</td>
<td>39.0</td>
<td>30.1</td>
<td>57.1</td>
<td>82.7</td>
<td>61.4</td>
<td>44.8</td>
</tr>
<tr>
<td>D. subobscura</td>
<td>39.0</td>
<td>29.1</td>
<td>49.9</td>
<td>71.2</td>
<td>61.4</td>
<td>43.9</td>
</tr>
<tr>
<td><strong>Genus Scaptodrosophila</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. lebanonensis</td>
<td>36.8</td>
<td>26.9</td>
<td>46.1</td>
<td>57.6</td>
<td>56.9</td>
<td>41.8</td>
</tr>
</tbody>
</table>

Sequences span 135 codons for Adh, 146 codons for Sod, except D. pseudoobscura and D. subobscura with 114 codons, and 51 codons for Per. The 28SrRNA sequences are 544 nucleotides long. Abbreviations and other conventions are as in Table 1.
in D. pseudoobscura (31.9%) and D. melanogaster (31.1%), and the three species have almost exactly the same number of GC-coded amino acids (~22.2%). Sod has equal proportions of AT-coded amino acids in D. saltans and D. melanogaster (22.7%), and the number of GC-coded amino acids is significantly higher in the former species (28.9 vs. 27.5%). The two shorter Sod amino acid sequences from the two obscura species are effectively identical to D. melanogaster in this respect. When using S. lebanonensis as an outgroup, the null hypothesis of equal rates of nonsynonymous substitution for Adh and Sod in D. saltans and D. melanogaster or the two obscura species is not rejected.

DISCUSSION

GC content differences: mutation pressure or selection? The interspecific differences in GC content between the three species groups of the Sophophora subgenus are larger than had been previously observed in Drosophila, even between species of different subgenera. The observation of similar patterns present in the five gene regions investigated (Xdh, Adh, Sod, Per, and 28S rRNA) suggests that they reflect genome-wide GC content differences between lineages. The changes in GC content can be attributed to an increase of AT content in the lineage that gave rise to the saltans group (Figure 2).

The GC content differences between the species groups might be a consequence of natural selection favoring lower GC content in the saltans group. Thermostable amino acids are encoded by GC-rich codons, and high GC content in third codon positions and in introns and untranslated flanking regions increases the thermal stability of the primary mRNA transcripts. Adaptation to heat has been suggested, which accounts for high GC content in the thermophilic bacteria (Kaga wa et al. 1984) and in the isochores of warm-blooded vertebrates (Bernardi et al. 1985). In Drosophila, solar heating of necrotic fruit may expose larvae to temperatures >45° even in temperate latitudes (Feder 1996). However, this hypothesis does not fit the biogeography of the species groups we have investigated: the highest GC content occurs in the obscura group species, which evolved in the cold and temperate climates of the Palearctic and Nearctic regions (Powell 1997), and the lowest in the saltans species, which evolved in tropical and subtropical regions (Powell 1997).

An alternative explanation is that the higher AT content of the saltans group species is not due to a functional advantage of the DNA base composition but simply results from a shift in the direction of the GC/AT mutation pressure shifting the group toward a new composition equilibrium. This predicts that directional changes will be more conspicuous in the neutral parts of the genome than in functionally significant parts, where mutation pressure is counteracted by selective constraints (Sueoka 1962, 1988). Our observations are fairly consistent with this prediction. GC% in the 28S rRNA locus, which is presumably under direct sequence selection, is the most conserved across the species investigated, while the GC content of Xdh, putatively the most unconstrained gene examined (amino acid divergence: $K_s = 0.0903$ in Xdh vs. $K_s = 0.0779$ in Sod; $K_s = 0.0667$ in Adh; and $K_s = 0.0645$ in Per), as averaged across the comparisons of D. saltans with D. melanogaster, D. pseudoobscura, and D. subobscura—the latter species not available for Per; estimated by the method of Wu and Li 1985), is the most variable. Moreover, GC content variation in Xdh, Adh, and Sod is highest in the third codon positions, whereas in the second codon positions it remains virtually identical across the species groups.

The specific molecular mechanisms that might account for a shift in mutation bias in the saltans lineage are unknown. They could involve, for instance, altered replication fidelities or replication repair systems, or changes in the availability of triphosphate nucleosides (dNTPs) during DNA synthesis. The shift might ultimately be traced to mutations affecting enzymes involved in DNA metabolism (mutator mutations; reviewed in Filipski 1990) or in the case of altered dNTP pools, be related to a shift in the trophic resources associated with speciation events. The patterns of GC content variation that we have observed might be a stimulus to explore these and related hypotheses in Drosophila.

Switch in the codon-usage pattern: In Drosophila, it is most commonly held that mutation bias is basically unimportant for codon bias, which rather results from the constrictions imposed on codon usage by tRNA availability and other factors related to translational efficiency and/or accuracy (review in Akashi et al. 1998). This view, usually referred to as the “major codon preference model,” is supported by several observations: (i) codon usage bias increases with the use of G and C, while nucleotide regions thought to reflect the mutational equilibrium of the genome are A + T rich; (ii) anecdotal evidence suggests a positive association between codon bias and expression levels; (iii) preferred codons in highly biased genes appear to match fairly well the most abundant isoaccepting tRNAs (Shields et al. 1988; Powell and Moriyama 1997); (iv) silent divergence between Drosophila species is inversely related to codon usage bias (Sharp and Li 1989; Moriyama and Gojobori 1992; Carulli et al. 1993); (v) regional variation in mutation patterns cannot explain the GC content variation at synonymous sites among highly biased genes and could account for only a minor fraction (~16%) of this variation among low-bias genes (Kliman and Hey 1994); (vi) lower codon usage bias in regions of lowest recombination in the D. melanogaster genome is consistent with theoretical predictions of the reduced efficacy of selection in such regions (Kliman and Hey 1993); (vii) in D. melanogaster codon bias is correlated
with functional constraints at the protein level (Akashi 1994); and (viii) estimates of the ratio of polymorphism to divergence for preferred (toward major codons) and nonpreferred (toward suboptimal codons) changes (Akashi 1995, 1996) and their frequency spectra in populations (Akashi and Schaeffer 1997) indicate differences in the evolutionary trajectories of the two categories of synonymous DNA mutations.

In the saltans group species, preferred codons for Adh, Sod, Per, and Xdh do not correspond to the postulated more abundant isoaccepting tRNAs in Drosophila (Powell and Moriyama 1997). Moreover, in opposition to the situation in D. melanogaster, none of these four genes is strongly biased in the saltans species. This situation is precisely the opposite of what would be expected, because alcohol dehydrogenase and superoxide dismutase are among the most abundant proteins in Drosophila (among the set of the 10% most abundant proteins), whereas xanthine dehydrogenase is not (Riley 1989). Under the major codon preference model this putative genome-wide shift in codon use in saltans could be due to either a change in the population of cognate tRNAs, relaxed selection for metabolic efficiency, or a reduction in the effectiveness of natural selection at silent sites.

If adaptation based on tRNA pools were the major factor for the atypical codon usage in the saltans group, we would have to assume that the relative abundance of the isoaccepting tRNAs changed during the evolution of the Sophophora species group. Some 35–50 million years have elapsed since the last common ancestor of D. melanogaster, one of these four genes is strongly biased in the saltans species. This situation is precisely the opposite of what would be expected, because alcohol dehydrogenase and superoxide dismutase are among the most abundant proteins in Drosophila (among the set of the 10% most abundant proteins), whereas xanthine dehydrogenase is not (Riley 1989). Under the major codon preference model this putative genome-wide shift in codon use in saltans could be due to either a change in the population of cognate tRNAs, relaxed selection for metabolic efficiency, or a reduction in the effectiveness of natural selection at silent sites.

If adaptation based on tRNA pools were the major factor for the atypical codon usage in the saltans group, we would have to assume that the relative abundance of the isoaccepting tRNAs changed during the evolution of the Sophophora species group. Some 35–50 million years have elapsed since the last common ancestor of the subgenus (Kwiatowski et al. 1994, 1997; Russo et al. 1995). Even if this time span were sufficient for changing the complete translation machinery, it would be expected that highly expressed genes (those experiencing greater selective constraints on codon usage) would change more rapidly toward the new adaptive equilibrium than lowly expressed genes, which should remain largely unaffected. But, contrary to this expectation, Adh and Sod codon biases in the saltans group are more similar to the optimal codon use in D. melanogaster (representing the hypothetical ancestral codon use pattern) than in the case of Xdh. Note that a slight nonoptimal shift in tRNA abundance would surely result in a reduction in translation efficiency (Shields 1990). It thus seems unlikely that a change in the abundance of the cognate tRNAs would have been a major reason for the switch in codon usage patterns along the saltans lineage.

Relaxed constraints do not appear to explain the codon use pattern in the saltans group either. The level of Adh enzyme activity in this species group is about the same as for the Slow allele in D. melanogaster and D. simulans and is approximately the mean for species breeding in rooting fruits (Mercot et al. 1994). Also the expression of Xdh is known to be largely unaffected by position effects (Spradling and Rubin 1983), which could occur because of structural changes undergone by the saltans genome. Unless there are significant differences in specific activity of these enzymes, it would seem that the level of expression is not the cause of the relaxation of selection.

Alternatively, the unusual codon usage pattern in the saltans group might not be a change in codon bias itself but rather an epiphenomenon caused by a reduction in the effectiveness of natural selection. The effectiveness of natural selection in determining the fate of mutations depends on the product of the effective population size and the coefficient of selection, Ns. Assuming constant s, a reduction of Ns is achieved in reduced populations or by a reduction in the rate of recombination; i.e., when recombination drops, the effect of natural selection at a given site essentially accelerates genetic drift at linked sites. Kliman and Hey (1993) found lower codon usage in regions of reduced recombination in D. melanogaster. The effect, however, does not appear to be a linear function of recombination rate; rather, it seems limited to regions with the very lowest levels of recombination (i.e., near centromeres and telomeres and on the fourth chromosome; Kliman and Hey 1993). It seems quite unlikely that all four genes investigated in this study fall into regions of such a low recombination rate in all six saltans species, further taking into account that Xdh, Adh, Sod, and Per belong to different linkage groups (3R, 2L, 3L, and X, respectively, in D. melanogaster). A genome-wide drastic reduction of recombination does not appear to be the case either. The karyotype of the saltans group species consists of three pairs of mitotic chromosomes: the sex and second chromosomes are metacentric, and the third is acrocentric: the X chromosome corresponds to the X and left limb of the third chromosome, the second to the second, and the third to the right limb of the third chromosome of D. melanogaster (Spassky et al. 1950). A linkage map based on 26 morphological markers is available for D. prosaltans (Spassky et al. 1950). From this data, assuming the markers are randomly scattered along the chromosomes and correcting for multiple crossovers (Kosambi 1944), the map length of D. prosaltans turns out to be almost equal to D. melanogaster (285 cM vs. 280 cM, respectively; True et al. 1996). Considering the relative mitotic size of the chromosomes (the third chromosome is little more than half as long as the others), map length values for each chromosome individually suggest that recombination is reduced for the second chromosome (81.67 cM) relative to the sex and third chromosomes (121.75 and 82 cM, respectively). Lacking more precise estimates at a regional scale, this might account for some of the bias of Adh (putatively located in the second chromosome) but leaves unexplained the codon use pattern for Xdh, Sod, and Per. Note, however, that the above measures of recombination may not accurately reflect the long-term rates of recombination that affected codon usage in D. prosal-
tans. Other studies found that recombination rates vary widely among closely related species (True et al. 1996), suggesting that the phylogenetic inertia of this parameter is probably too weak to account for the fairly homogeneous codon use pattern in the saltans species group.

Reduced efficiency of natural selection can also result from a decline in the effective population number. The three- to sixfold smaller population size of D. melanogaster relative to D. simulans has been invoked to explain the barely ~2% difference in codon bias between the two species (Akashi 1995). A reduced population number, however, might be insufficient to explain the ~40% decline in major codon use observed in the saltans group (Figure 3). With major codon preferences, regions under the weakest selection pressure for base composition are expected to show the lowest sensitivity to changes in N_e (see Akashi 1996). Instead we see that codon bias in the low-expressed, allegedly less constrained Xdh gene shows a shift about as dramatic as the highly expressed, presumably more constrained Adh and Sod genes. We have calculated the average ratio of synonymous to nonsynonymous substitutions (K_s/K_a; Wu and Li 1985) for Adh and Xdh in the saltans group and in the group consisting of D. melanogaster and the two obscura species. K_s/K_a is lower for Adh (15.79) than for Xdh (20.82; or 17.34 for the comparison between the two obscura species less affected by saturation) in the latter species due to higher synonymous substitution rates in Xdh (0.9981 vs. 0.5758). In the saltans group the ratio decreases for both genes as expected if populations of these species were smaller; however, the ratio decreases notably less for Xdh (12.04) than for Adh (4.21), which can hardly be accounted for by a smaller effective size if silent sites of Xdh are under the weakest selection pressure. Thus, while a reduced effectiveness of selection associated with low population numbers might account for part, it cannot explain all of the shift in the codon use pattern of the saltans species group. In this respect, it may be significant that the proportion of preferred codons for Adh in the saltans species, many of which are widely distributed continental species, is by far more extreme than in the Hawaiian species (average F_pop = 0.56, four species; Thomas and Hunt 1991), which are known to have experienced repeated bottlenecks and to maintain reduced population sizes (Olht a 1993).

A likely explanation for the codon use pattern in the saltans group is that a shift in the mutation bias toward greater A + T content occurred early after the split in the common ancestor of the saltans group from other Sophophora and exerted enough pressure so as to switch codon preferences. The current codon use pattern in the saltans group may, then, represent a remnant of an ancestral codon bias that is being predominantly degraded by mutation pressure toward a new equilibrium composition bias. The historic pattern may persist longest in those family codons and genes that, as presumably is the case for Adh and Sod, are highly biased toward the ancestral pattern. This interpretation is consistent with the theoretical results of Shiel ds (1990), who has shown that, over a certain range, a shift in mutation bias can trigger a complete switch in codon preference.

Our results challenge currently held opinions about the importance of selection for codon bias in Drosophila (Powell 1997), although we do not exclude the possibility that selection may play a role once a new composition equilibrium has been reached. The significance of fluctuating mutation biases for switches in codon preferences has been discussed for several unicellular lineages (Shiel ds 1990; Li 1997). Existing information about Drosophila comes from limited evidence. For the most part, it consists of extrapolations from what has been observed in D. melanogaster and, to a much lesser extent, in a few other species, particularly D. virilis and D. psaoobscura (Powell 1997). GC content differences among these species are substantially smaller than the ones detected in our study. Consequently, their genomes are likely to have been subjected to similar mutation pressures, whose effects would be difficult to detect below a critical range (Shiel ds 1990). Our observations may contribute to explaining some "atypical" patterns such as high incidence of A- or T-ending codons in the Adh (Anderson et al. 1993), Sod (Kwiatowski et al. 1994), and Per (Gleason 1996) genes of D. willistoni (Powell 1997). This species belongs to the willistoni group, which is the sister clade of the saltans group within the Sophophora subgenus (Patterson and Stone 1952). Hence, the codon use pattern of these genes in D. willistoni may be simply a result of the same mutation bias that has impacted the saltans group species.

**Mutation bias and the rate of protein evolution:** Accelerated amino acid substitutions in the saltans group could reflect either the fixation of deleterious amino acid mutations (N_eS = −1) or a faster rate of adaptive evolution. Directional selection for replacement changes can accelerate genetic drift at linked silent sites (see Akashi 1996), resulting in reduced effectiveness of selection for codon bias. In the saltans lineage, the rate of nonsynonymous substitutions in Xdh has been ~2-fold greater than in the obscura group or D. melanogaster (ranging between 1.69 and 3.15; see results). To account for this difference as a consequence of natural selection, one has to assume that amino acids encoded by GC-rich codons are advantageous in the GC-rich species of the obscura and melanogaster groups, whereas amino acids encoded by GC-poor codons are advantageous in the GC-poor species of saltans (Li 1997). As discussed above in connection with the differences in base composition, it seems unlikely that such is the case given the environments where these species live.

The observed differences can be better interpreted as a consequence of directional mutation pressure...
Codon Bias and Rate of Evolution

(Sueoka 1962, 1988). Sueoka’s theory predicts that for a set of nucleotides at equilibrium between mutation pressure, base composition, and selective constraints, GC content will remain essentially unchanged until a shift in the direction of mutation pressure occurs. The change in mutation bias will then provoke a burst of mutations that will rapidly decrease with time until the new composition equilibrium is reached. This dynamic applies both to neutral and nonneutral sets of nucleotides, although the effect is expected to be much less pronounced for nonneutral nucleotides. Consequently, when the direction of mutation pressure changes in association with phylogenetic branching, the reconstructed branches that lead to the two extant populations should become different in length (measured in terms of nucleotide substitutions along each branch). The shorter branch will likely be more similar to the parental branch from which the offspring populations have emerged (Sueoka 1993). It follows that the rate of replacement of quasi-neutral amino acids should increase, which is what we have observed for the amino acid replacements occurring in Xdh during the saltans group evolution. The effect cannot be detected in Adh, Sod, and Per, indicating that either it did not occur or is weak. Given that these three genes appear to be a more functionally constrained protein than Xdh, this is precisely what would be anticipated, because mutation pressure will have less impact on the amino acid composition of proteins subjected to stronger functional constraints.

As discussed above, in connection with the patterns of codon use, our results cannot be accounted for merely by a reduction in population size. In addition, it is not likely that long-term population bottlenecks have occurred regularly in the evolution of the saltans group, independently across the different species of the group. Nor can the higher number of amino acid replacements of Xdh in the saltans group be explained by differences in generation time. Even if nonsynonymous substitutions are so nearly neutral to behave effectively as if they were synonymous, the generation time in the saltans species is not shorter than in the obscura group and longer than in D. melanogaster.

In view of this evidence, it seems likely that in the evolution of the subgenus Sophophora, since ~35–50 mya, mutation bias may have remained largely unchanged in the obscura and melanogaster group lineages. However, at some time point after the origin of the saltans lineage the strength of mutation bias changed substantially. The pressure exerted thereafter by the new mutation pattern has been strong enough to change the nucleotide composition (including that of regions subjected to direct sequence selection; i.e., ribosomal RNA), drastically modify the pattern of codon usage (even in highly expressed genes; e.g., Sod), and significantly accelerate the rate of relatively unconstrained proteins such as Xdh. Confirmation of all the trends for a larger number of genes would strongly support that a substantial fraction of molecular variants are weakly selected in Drosophila.

We are grateful to Carlos Machado for suggestions and critical discussion. We thank Hafid Layyouni, Richard Hudson, Lars Jermain, and Mauro Santos for valuable suggestions, and Xun Gu, David Hewett-Emmett, and Wen-Hsiung Li for sending us their manuscript before publication. Jody Hey and the two anonymous reviewers made crucial comments and pointed to additional sources of data relevant to the hypotheses presented in this article. Antonio Barbadilla and Mario CaÁceres helped with map distance calculations. F.R.-T. received support from Ministerio de EducacioÁn y Cultura (Spain; Contrato de Reincorporacion) and grant PB96-1136 to A. Fontdevila. This work was supported by National Institutes of Health grant GM42397 to F.J.A.

LITERATURE CITED


Gaut, B. S., V. M. Muse, W. D. Clark and M. T. Clegg, 1992 Relative...