The DNA sequences of the Adh genes of three species of the Drosophila melanogaster species subgroup have
been determined. This completes the Adh sequences of the eight species of this subgroup. Two species, D. yakuba
and D. teissieri, possess processed Adh pseudogenes. In all of the species of the subgroup, a gene of unknown
function, Adhr, is located about 300 bp 3' to Adh. Although this gene is experiencing a higher rate of synonymous
substitution than Adh, it is more constrained at the amino acid level. Phylogenetic relationships between all eight
members of the melanogaster subgroup have been analyzed using a variety of methods. All analyses suggested that
the D. yakuba and D. teissieri pseudogenes have a single common ancestor, rather than evolving independently
in each species, and that D. melanogaster is the sister species to D. simulans, D. sechellia, and D. mauritiana. The
evolutionary relationships of the latter three species remain equivocal.

Introduction

Drosophila offers many advantages for the study of molecular evolution. This is not only because of the
wealth of our knowledge of one species, D. melanogaster, but also because relationships within and between species
groups have been studied experimentally for many years, not least by the powerful method of polytene chromo-
some banding pattern analysis. Drosophila melanogaster itself belongs to the melanogaster species group of the
subgenus Sophophora. Most closely related to the obscura species group, the melanogaster species group in-
cludes 158 species and 12 species subgroups. The great majority of these subgroups are endemic to the Oriental
region. A few, in particular the large montium species subgroup, with Afrotropical and Eastern Palearctic
members, and the small melanogaster species subgroup, which is essentially Afrotropical in its distribution, are
exceptional. The relationship between the montium and melanogaster species subgroups is not close, despite their
coincident distribution in the Afrotropical region. Poly-
tene chromosome studies (F. Lemeunier and M. Ash-
burner, unpublished data) show that the melanogaster

species subgroup is more closely related to subgroups essentially limited to the Oriental region, e.g., the taka-
hashii species subgroup. The purpose of this paper is to
consider the relationships within the melanogaster spe-
cies subgroup by an analysis of the sequences of their
Adh genes. The relationships between the major branches
of the melanogaster species subgroup, based on polytene
chromosome analysis, have been studied by Lemeunier
and Ashburner (1976, 1984). These provide independent
criteria for evaluating phylogenetic hypotheses from the
sequence data.

The Adh gene itself is interesting for a study of mo-
lecular evolution for several reasons (for a review, see
Sullivan et al. 1989). Not only is it involved in the ad-
aptation of Drosophila to fermenting substrates, but also
the molecular basis for its transcriptional control is the
object of considerable study, in D. melanogaster and
several other species. For this reason we have evaluated
conservation and change in sequence, with respect to
knowledge of the sequences to which trans-acting tran-
scriptional factors bind.

The D. melanogaster species subgroup consists of
eight sibling species (Lemeunier et al. 1986). Two species
(D. melanogaster and D. simulans) are cosmopolitan,
two (D. mauritiana and D. sechellia) are island endemics,
and the remaining four (D. erecta, D. orena, D. teissieri,
and D. yakuba) are endemic to western and central Af-
rica. Many studies have been made to determine the phylogenetic relationships between these species, perhaps
the most comprehensive being cytological studies on polytene chromosome banding patterns (Lemeunier and
Ashburner 1976, 1984), allozyme electrophoretic distances (Cariou 1987), and the molecular analysis of the Adh gene (Bedmer and Ashburner 1984; Coyne and Kreitman 1986; Cohn and Moore 1988). Lachaise et al. (1988) reconciled these data with biogeographic evidence relevant to the evolutionary history of the subgroup. The origin of the subgroup was considered to be by immigration of an ancestral population into equatorial Africa from Eurasia. This, it was suggested, occurred 17–20 Mya, at the time when faunal interchange between Africa and Eurasia by a continental route was first possible. As a consequence of vicariance, differentiation of this population occurred, probably in west equatorial Africa, first giving rise to the erecta-arena lineage and then to the yakuba-teissieri and proto-melanogaster lineages. Differentiation within the melanogaster complex began in the late Pleistocene (about 2.5 Mya), as a result of tectonic activity in eastern Africa and the migration of a proto-simulans population to the less arid islands of the Indian Ocean.

Material and Methods

Strains

Strains of Drosophila erecta (154.1), D. teissieri (128.2), and D. yakuba (115) were obtained from the Laboratoire de Biologie Génétique Évolutive, Gif-sur-Yvette, France.

Molecular Methods

All of the molecular methods used in this study have been described by Jeffs and Ashburner (1991).

Data Sets for Analysis

Because evolutionary constraints are likely to vary across the Adh-Adhr region, phylogenetic trees were constructed separately for the following five parts of the Adh-Adhr region: Adh coding region, Adh introns, Adhr coding region, Adhr introns, and the 5' noncoding region. Furthermore, a phylogenetic analysis was also undertaken on a concatenation of these five regions.

Phylogenetic Analysis

Sequence alignments were carried out either using the GAT program in the sequence analysis package (version 6.0) of the Genetics Computer Group, University of Wisconsin (Devereux et al. 1984) or with CLUSTAL V (Higgins et al. 1992) and then were checked by eye. Because of the difficulties in aligning some of the more divergent sequence regions, especially in the 5' noncoding region and the intron sequences, phylogenetic analyses were also conducted on data sets after regions where the alignment was most uncertain had been removed.

A number of different methods of phylogenetic analysis were used in order to increase confidence in the groupings found. All programs were taken from the PHYLIP (version 3.5c) package of Felsenstein (1993) DNAML constructs a phylogeny under a stochastic model of molecular evolution with maximum-likelihood inference. The relevant settings for the transition/transversion ratio were estimated from the empirical data and were found to be 1:1 for the Adh coding region, Adh introns, Adhr introns, and the complete Adh-Adhr coding region; 2:1 for the Adhr coding region; and 0.75:1 for the 5' noncoding region. The relevant settings for the rates of nucleotide substitution at codon positions 1, 2 and 3 were found to be 3:1:8 and 2:1:20 for the Adh and Adhr coding regions, respectively. "Global" branch swapping was used in all cases, and multiple runs of the programs were made (with the Jumble option) in order to have a better chance of finding the tree of highest likelihood. The likelihoods of competing trees were compared directly by a significance test based on the mean and variance of the differences in log likelihood (Kishino and Hasegawa 1989).

NEIGHBOR clusters a matrix of nucleotide sequence distances according to the neighbor-joining algorithm of Saitou and Nei (1987). The distances themselves were estimated under Kimura's two-paramate method by using the DNADIST program. In order to assign some degree of certainty to individual groupings, 2,000 bootstrap replications were used in the neighbor joining analysis (according to the suggestion of Hedge 1992) by making use of the SEQBOOT and CONSENS programs in PHYLIP. Neighbor-joining trees were also constructed on a matrix of silent substitution distance as estimated by the method of Li et al. (1985) (Progran LWL89 was provided by K. Wolfe, Department of Genetics, Trinity College, Dublin). Finally, phylogenetic trees were reconstructed under the parsimony criterion (with 2,000 bootstrap replications; PHYLIP program SEQBOOT, DNAPARS, and CONSENSE).

In order to root the phylogenetic trees, an outgroup species has to be specified. In the case of the Adh and Adhr coding sequences, D. pseudoobscura was used as a the outgroup. However, D. pseudoobscura sequence were not available for the 5' noncoding region and were too divergent in the cases of the intron sequences. In these cases the sequences from D. erecta were used as a the outgroup, and those from D. arena were removed from the analysis, as only a single outgroup can be specified. In the case of the complete Adh-Adhr region, the sequence from D. teissieri was used as the outgroup.

Pairwise comparisons of sequences were made with the TWO program of R. Staden. Each gap was scored as a single event and weighted equally with a base change
Unless stated otherwise, the estimates of pairwise differences between sequences are uncorrected for multiple substitutions. When correction has been used, the frequencies of substitution per site are estimated by \( d = -0.75\ln(1-4p/3) \) (Jukes and Cantor 1969).

Codon bias was estimated by calculating the expected use of synonymous codons (using M. J. Bishop's EQUALCODONS program) and then calculating \( \chi^2 \) for the difference between the observed and expected codon tables (with M. J. Bishop's CODONFIT program). For comparison between different genes, the total \( \chi^2 \) is normalized by dividing by the number of codons (\( L \)), and the measure of codon bias is expressed as \( \chi^2/L \) (Shields et al. 1988). Codon bias was also estimated by calculating \( N_c \), the effective number of codons (Wright 1990). \( N_c \) was estimated using the CODONS program of Lloyd et al. (1988). Codon bias was also estimated by calculating the expected use of synonymous codons (using M. J. Bishop's EQUALCODONS program) and then calculating \( \chi^2 \) for the difference between the observed and expected codon tables (with M. J. Bishop's CODONFIT program). For comparison between different genes, the total \( \chi^2 \) is normalized by dividing by the number of codons (\( L \)), and the measure of codon bias is expressed as \( \chi^2/L \) (Shields et al. 1988). Codon bias was also estimated by calculating \( N_c \), the effective number of codons (Wright 1990). \( N_c \) was estimated using the CODONS program of Lloyd and Sharp (1992). Were codon use random, then the value of \( N_c \) would be 61, and the most extreme bias would result in \( N_c = 20 \).

Data Deposition

The new sequence data are not published in full but have been deposited in the EMBL/GenBank/DDBJ nucleic acid sequence data banks, under the following accession numbers: X54116 (D. erecta), X54117 (D. teissieri pseudogene), X54118 (D. teissieri), X54119 (D. yakuba pseudogene), and X54120 (D. yakuba). The other melanogaster species subgroup sequences used were D. melanogaster Adh+ (Z00030), D. simulans (X00607), D. mauritiana (M19264 and Z00033), D. sechellia (X04672), and D. orena (Z00032). The coding sequence of the Adh gene of D. tsacasi, a member of the montium species subgroup of the melanogaster species group, is from Maruyama and Hartl (1991). Sequence data from species of the obscura group of Drosophila that have been used were D. pseudoobscura (Y00602), D. persimilis (M60997), D. miranda (M60998), and D. ambigua (X54813).

Sequence Numbering

The sequences are all numbered with respect to the distal messenger RNA (mRNA) start of D. melanogaster Adh. This base is +1, and bases are negative 5' to it and positive 3' to it. Base +1 is numbered 4771 in the data bank accession Z00030.

Results

Cloning and Sequencing the Adh Genes

Genomic phage libraries from Drosophila yakuba, D. teissieri, and D. erecta were screened with a D. melanogaster Adh probe, and a number of positive phage were recovered from each. Restriction mapping showed that all the phage from the D. erecta library came from a single genomic region but that the phage from the D. yakuba and D. teissieri libraries were of two classes, each corresponding to a different set of fragments seen to hybridize to total genomic DNA. One of these classes has been shown to correspond to clones of an Adh pseudogene in each of these species (Jeffs and Ashburner 1991; also see Long and Langley 1993). The other class of clone, as well as that from D. erecta, correspond to functional Adh genes. As expected, they map, by in situ hybridization, to 35B on the polytene chromosomes of D. melanogaster and to the corresponding sites in their own species (data not shown). The sequences of these clones show that, as in D. melanogaster and D. pseudoobscura (Coyne and Kreitman 1986; Schaeffer and Aquadro 1987), the Adh gene is immediately flanked by a gene homologous in structure and similar in sequence: Adhr (= Ahh-dup of other authors).

A summary of the structures of the Adh and Adhr genes in these species, as well as a comparison with those of the other five members of the melanogaster species subgroup and with the genes of D. pseudoobscura is shown in figure 1. For the interpretation of these structures we assume that all of these Adh genes are transcribed from alternative distal or proximal promoters (see Discussion). There are several differences between our D. yakuba sequence and those of McDonald and Kreitman (1991). These are noted in the feature table that accompanies the Nucleic Acid Sequence Data Library Accession X54120 (also see table 3). All 12 alleles sequenced by McDonald and Kreitman were from collections made in Brazzaville, Zaire. The allele sequenced in our study was from a long-established strain collected (in 1967) at Kounden, West Cameroon.

Rates of Substitution

The numbers of synonymous and nonsynonymous nucleotide substitutions between the Adh and Adhr genes of the subgroup were estimated using the method of Li et al. (1985). The values of \( K_A \) (the number of nonsynonymous substitutions per nonsynonymous site) and \( K_S \) (the number of synonymous substitutions per synonymous site) compared with those in D. pseudoobscura are given in table 1.

For Adh the rates of synonymous substitution are about 12-fold higher than the rates of nonsynonymous substitution. For the Adhr gene this difference is about 39-fold. Although Adhr shows a slightly (though not significantly) lower \( K_S \) than does Adh, implying a greater selective constraint, the disparity between the genes is greater at silent sites, where Adhr has a \( K_S \) approximately twice that of Adh. This difference correlates with a codon-use difference between the two genes (table 2). The Adh genes have a much more biased codon use than do the Adhr genes. As observed elsewhere (Ashburner et al.
FIG. 1.—Summary of the organization of the Adh and Adhr genes of the melanogaster species subgroup (with those of Drosophila pseudoobscura, for comparison). Both Adh and Adhr have three coding exons (thicker lines). Note that the position of the introns between these are very similar in the two genes. Adh is transcribed from either a distal (P_D) or proximal (P_p) promoter. The length (in bp) of each region in each of the nine species is shown below the diagrams (intron lengths are in italics). The third column of numbers refers to the sequence unique to the 5' end of the proximal Adh mRNA, and the fourth column refers to the sequence common to the 5' end of the proximal and distal Adh mRNAs.
Table 1

KS and KA, for the melanogaster Adh and Adhr Genes, Compared with Those of Drosophila pseudoobscura

<table>
<thead>
<tr>
<th>Gene and Species</th>
<th>KS</th>
<th>KA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>melanogaster</td>
<td>0.6408 ± 0.0830</td>
<td>0.0475 ± 0.0093</td>
</tr>
<tr>
<td>simulans</td>
<td>0.6402 ± 0.0832</td>
<td>0.0494 ± 0.0095</td>
</tr>
<tr>
<td>sechellia</td>
<td>0.6311 ± 0.0823</td>
<td>0.0494 ± 0.0095</td>
</tr>
<tr>
<td>mauritiana</td>
<td>0.6641 ± 0.0851</td>
<td>0.0531 ± 0.0099</td>
</tr>
<tr>
<td>yakuba</td>
<td>0.6246 ± 0.0810</td>
<td>0.0512 ± 0.0097</td>
</tr>
<tr>
<td>teissieri</td>
<td>0.6317 ± 0.0812</td>
<td>0.0438 ± 0.0089</td>
</tr>
<tr>
<td>oreina</td>
<td>0.5583 ± 0.0741</td>
<td>0.0530 ± 0.0099</td>
</tr>
<tr>
<td>erecta</td>
<td>0.5652 ± 0.0745</td>
<td>0.0531 ± 0.0099</td>
</tr>
</tbody>
</table>

| Adhr:            |        |        |
| melanogaster     | 1.3510 ± 0.1889 | 0.0367 ± 0.0077 |
| simulans         | 1.2688 ± 0.1845 | 0.0351 ± 0.0076 |
| sechellia        | 1.2951 ± 0.1817 | 0.0334 ± 0.0074 |
| mauritiana       | 1.2958 ± 0.1831 | 0.0318 ± 0.0072 |
| teissieri        | 1.4266 ± 0.2404 | 0.0344 ± 0.0075 |
| pseudoobscura    | 1.186   | 59.81  |
| persimilis       | 0.241   | 57.72  |
| miranda          | 0.240   | 57.72  |
| ambigua          | 0.470   | 48.24  |

1984), Adh codons very rarely have an A in the third position; for Adhr this is not so marked, at least in the melanogaster subgroup. Overall, Adhr is, in its coding region, less G+C rich than is Adh. The G+C content of Adh is 57.46%–59.92% in the melanogaster subgroup; that of Adhr is 47.5%–51.28%. The same trend is also clear in the obscura species group (see table 2).

Distal Enhancer Sequences

We have analyzed about 650-bp of sequence upstream of Adh in seven of the species. Within this region is located the “adult” Adh enhancer (Corbin and Maniatis 1989a, 1989b; Ayer and Benyajati 1990, 1992; England et al. 1990; Falb and Maniatis 1992a). This A+T-rich sequence (0.61 A+T in D. melanogaster) shows a markedly mosaic pattern of highly conserved and highly divergent regions (fig. 2). The 200 bp immediately distal to the distal mRNA cap site shows an identity of 0.735 between all seven species. From −201 to −430 this identity is only 0.196; from −431 to −603 there is a second conserved region, with an identity of 0.734. The regions of high identity include all of the sites that are protected against DNAse 1 digestion when bound to purified Adfl transcription factor in both D. melanogaster and D. oreina (Moses et al. 1990). The sites protected in only one of these species (i.e., d5(o) and d5(m)) are in the very variable region between −431 and −201.

Table 2

Analysis of Codon Use in Adh and Adhr

<table>
<thead>
<tr>
<th>Gene and Species</th>
<th>χ²/L</th>
<th>N,</th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>Overall G+C Content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>melanogaster</td>
<td>0.756</td>
<td>31.35</td>
<td>14.40</td>
<td>50.58</td>
<td>4.67</td>
<td>30.35</td>
<td>57.72</td>
</tr>
<tr>
<td>simulans</td>
<td>0.781</td>
<td>30.34</td>
<td>14.06</td>
<td>51.56</td>
<td>3.52</td>
<td>30.86</td>
<td>58.46</td>
</tr>
<tr>
<td>sechellia</td>
<td>0.801</td>
<td>30.28</td>
<td>15.18</td>
<td>50.97</td>
<td>3.89</td>
<td>29.86</td>
<td>57.85</td>
</tr>
<tr>
<td>mauritiana</td>
<td>0.808</td>
<td>30.28</td>
<td>14.40</td>
<td>51.36</td>
<td>4.28</td>
<td>29.86</td>
<td>58.11</td>
</tr>
<tr>
<td>yakuba</td>
<td>0.833</td>
<td>30.33</td>
<td>12.06</td>
<td>51.36</td>
<td>4.28</td>
<td>32.30</td>
<td>58.76</td>
</tr>
<tr>
<td>teissieri</td>
<td>0.846</td>
<td>29.67</td>
<td>9.34</td>
<td>54.47</td>
<td>3.89</td>
<td>37.30</td>
<td>59.92</td>
</tr>
<tr>
<td>oreina</td>
<td>0.780</td>
<td>32.41</td>
<td>15.18</td>
<td>49.81</td>
<td>3.89</td>
<td>31.13</td>
<td>57.46</td>
</tr>
<tr>
<td>erecta</td>
<td>0.820</td>
<td>32.59</td>
<td>13.62</td>
<td>51.75</td>
<td>4.28</td>
<td>30.35</td>
<td>58.11</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>0.631</td>
<td>36.74</td>
<td>20.47</td>
<td>45.67</td>
<td>3.15</td>
<td>30.71</td>
<td>56.17</td>
</tr>
<tr>
<td>persimilis</td>
<td>0.611</td>
<td>37.59</td>
<td>19.22</td>
<td>46.67</td>
<td>3.53</td>
<td>30.59</td>
<td>56.34</td>
</tr>
<tr>
<td>miranda</td>
<td>0.605</td>
<td>38.58</td>
<td>20.39</td>
<td>45.58</td>
<td>3.53</td>
<td>30.20</td>
<td>55.82</td>
</tr>
<tr>
<td>ambigua</td>
<td>0.840</td>
<td>36.35</td>
<td>18.43</td>
<td>45.10</td>
<td>2.35</td>
<td>34.12</td>
<td>56.86</td>
</tr>
<tr>
<td>Adhr:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>melanogaster</td>
<td>0.186</td>
<td>59.81</td>
<td>26.01</td>
<td>25.27</td>
<td>18.68</td>
<td>30.04</td>
<td>47.50</td>
</tr>
<tr>
<td>simulans</td>
<td>0.241</td>
<td>57.72</td>
<td>23.44</td>
<td>27.47</td>
<td>18.68</td>
<td>30.40</td>
<td>48.23</td>
</tr>
<tr>
<td>sechellia</td>
<td>0.246</td>
<td>56.72</td>
<td>22.34</td>
<td>27.47</td>
<td>19.05</td>
<td>31.14</td>
<td>48.59</td>
</tr>
<tr>
<td>mauritiana</td>
<td>0.239</td>
<td>56.59</td>
<td>23.81</td>
<td>26.74</td>
<td>18.68</td>
<td>30.77</td>
<td>48.35</td>
</tr>
<tr>
<td>teissieri</td>
<td>0.283</td>
<td>53.75</td>
<td>19.41</td>
<td>32.60</td>
<td>15.02</td>
<td>32.97</td>
<td>51.28</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>0.440</td>
<td>45.48</td>
<td>22.22</td>
<td>35.45</td>
<td>9.68</td>
<td>32.62</td>
<td>53.64</td>
</tr>
<tr>
<td>miranda</td>
<td>0.320</td>
<td>46.36</td>
<td>22.94</td>
<td>34.41</td>
<td>10.04</td>
<td>32.62</td>
<td>52.92</td>
</tr>
<tr>
<td>ambigua</td>
<td>0.470</td>
<td>48.24</td>
<td>21.90</td>
<td>32.98</td>
<td>10.99</td>
<td>34.84</td>
<td>52.72</td>
</tr>
</tbody>
</table>
Fig. 2.—Alignment of a 650-bp region upstream to the distal Adh mRNA cap site. The regions binding the Adfl transcription factor in both Drosophila melanogaster and D. orena or in only one of the species are indicated by the blackened and unblackened boxes, respectively (data are from Moses et al. 1990). Also shown, as shaded boxes, are the DEP1-2, DEP3, and DEP4 binding sites.
Ayer and Benyajati (1990) have identified three other putative transcription factor binding sites: for AP-1, between −493 and −486; for DTF-1, between −532 and −525; and for octamer-binding factors, between −525 and −512. All three sites are absolutely conserved between the seven species, except for a single base change in the D. erecta-Adult enhancer (Falb and Maniatis 1992a). The overlapping DEP4 site is one that binds the mammalian CCAAT/enhancer-binding protein (C/EBP) (Falb and Maniatis 1992a). The C/EBP binding sequence is absolutely conserved in all species, except for a C-to-A change in D. sechellia (fig. 2).

Figure 3a and b show alignments of the 5'-untranslated regions of the distal and proximal Adh mRNAs. Within the melanogaster species subgroup, two species—D. erecta and D. sechellia—differ from the others by a 12-bp deletion in the sequence common to both leaders (i.e., the sequence encoded by the untranslated part of exon 2; fig. 3b). Apart from this deletion, the overall similarity between the leader sequences is very high—0.874 for the distal mRNA leader, 0.862 for the proximal. The distal leader of D. pseudoobscura is shorter than that in the melanogaster subgroup, 92 bases versus 111-123 bases. However, in that region of the leader encoded by exon 1 (which is 87 bp in all species except D. pseudoobscura, where it is 79 bp), the D. pseudoobscura sequence is reasonably conserved (a similarity of 0.655 with D. melanogaster). The proximal mRNA leader is also shorter in D. pseudoobscura than in D. melanogaster, 46 bases versus 70 bases, and is diverged...
In sequence to the extent that a convincing alignment has not been achieved.

Intron Sequences

The first intron of Adh is specific to the distal transcript. It separates an entirely noncoding exon 1 from the noncoding 5' portion of exon 2 (fig. 3b). The length of this intron is similar in all melanogaster subgroup species, ranging from 613 bp (D. erecta) to 666 bp (D. teissieri). In D. pseudoobscura, intron 1 is 793 bp. Between-species comparisons of the sequences show a pattern similar to that seen for the sequences 5' to the mRNA (see above), i.e., a mosaic of conservation and divergence. This is probably for very similar reasons, since intron 1 also includes sequences that bind specific transcription factors (Heberlein et al. 1985). From a divergence. This is probably for very similar reasons, since intron 1 also includes sequences that bind specific transcription factors. These sequences are also conserved in all other members of the melanogaster subgroup (fig. 4), but not obviously so in D. pseudoobscura.

Within the Adh protein coding region there are two small introns (63–67 bp and 60–75 bp). These show a much lower degree of sequence conservation than does intron 1, both within the melanogaster subgroup and in comparison with D. pseudoobscura (which is essentially random in intron 2, compared with D. melanogaster). The two introns within Adhr (312–435 bp and 51–58 bp) show a very similar pattern.

Adh-Adhr Spacer Sequence

The “spacer” region that separates Adh and Adhr (taken as being the sequence from the Adhr terminator codon to the Adhr ATG codon) may include both the sequences required for the correct termination of transcription of Adh and the sequences required for the correct initiation of transcription of Adhr (fig. 5). The sequence varies in length, between 296 bp and 326 bp. It is, overall, very high in A+T (0.634 in D. melanogaster) and, within the melanogaster subgroup, reasonably conserved in sequence (>0.85, for all pairwise comparisons). Between these species and D. pseudoobscura the conservation is less, of the order of 0.55. However, achieving a satisfactory alignment of the D. pseudoobscura and melanogaster subgroup sequences is not trivial.

It has been claimed that Adhr has both CAAT and TATA boxes upstream of its coding region (Schaeffer and Aquadro 1987; Marfany and González-Duarte 1991). The sequence thought to represent a CAAT box in D. pseudoobscura is very well conserved between D. pseudoobscura and the melanogaster subgroup. It is in the same position, relative to the ATG codon of Adhr (but not relative to the terminator codon of Adh), in D. pseudoobscura and D. melanogaster and is identical in sequence in all species. The sequence thought to represent the Adhr TATA box in D. pseudoobscura is not conserved in position between D. pseudoobscura and the melanogaster subgroup; the similar sequence in these species is probably that at +1880 to +1888. However, this does not look to be a very convincing TATA box, since the region in general has a very high A+T content and since the putative box is not in a G+C-rich local environment.

The sequence immediately upstream of the ATG codon of Adhr is very highly conserved within the melanogaster subgroup and with D. pseudoobscura. It departs from the Drosophila translational start consensus in having a G in position −2 (with respect to the ATG) (Cavener 1987).

Adh Protein Sequences

The translated Adh DNA sequences of all eight species of the melanogaster species subgroup and of D. pseudoobscura are given in figure 6. The sequences of the melanogaster subgroup species differ from that of D. pseudoobscura (and, in fact, from all species in other species groups or subgroups of Drosophila) by two extra residues (positions 3 and 4). Apart from this difference there are 32 variable sites, of which 13 are unique to D. pseudoobscura (table 3).

The sequences of the available Adhr proteins are shown aligned in figure 7. The conservation between these proteins is even higher than that for Adh (table 4). All amino acid substitutions are unique to a single species. The Adhr proteins of the obscura group species are extended, with respect to those of the melanogaster subgroup, at their C-termini, by six amino acids in D. pseudoobscura and D. miranda and by nine amino acids in D. ambigua.

Phylogenetic Analysis

Our analysis has concentrated on the relationships of the four species of the melanogaster species complex: mauritiana, melanogaster, sechellia, and simulans. There are three possible topologies (I–III) relating simulans, mauritiana, and sechellia if melanogaster is assumed to be the sister group to these three. This assumption is consistent with all the data presented here (see below), as well as with those presented elsewhere (Lachaise et al. 1988). The results of all phylogenetic analyses of the simulans complex are summarized in table 5, with the columns representing the competing topologies and with the rows representing the different data sets.
Adh Coding Sequence

The analysis of 771 bp of the Adh coding sequence is especially informative, as it is the only region for which sequences from all eight members of the melanogaster species subgroup, as well as from the teissieri and yakuba pseudogenes, were available. The maximum-likelihood tree for these sequences is presented in figure 8a. Identical trees were produced from the neighbor-joining analysis of both complete nucleotide distances and silent substitutions only (not shown), while the parsimony analysis differed only in the grouping of the teissieri and yakuba functional Adh genes (fig. 8b).

The close relationship of the four members of the melanogaster species complex, as well as the orena-ecta and teissieri-yakuba species pairings, are clear from these trees. Drosophila teissieri and D. yakuba share a more recent common ancestry with the melanogaster species complex than do D. orena and D. erecta, although this relationship is not significantly better than one that puts D. orena and D. erecta closer to the melanogaster species complex. Drosophila tsacasi, as expected for a member of the montium species subgroup, occupies an “intermediate” position between pseudoobscura and the melanogaster species subgroup.

Of less certainty are the relationships between the teissieri and yakuba pseudogenes and their functional counterparts; however, the fact that the two pseudogenes are depicted as having a common ancestry, rather than as grouping with their functional paralogues, supports the notion that they were formed as a single event in the common ancestor of both species. It might then be
expected that the functional Adh genes of teissieri and yakuba would group together. Although this is not the case in either the maximum-likelihood or the neighbor-joining analyses, the length of the branch between teissieri and yakuba in the former analysis was so short as to be insignificant under the maximum-likelihood model. In the neighbor-joining analysis only 58% of bootstraps grouped the functional teissieri gene with the two pseudogenes. These two sequences did group together in the parsimony analysis, however, although with support from only 44% of the bootstrap replications (fig. 8b).

All methods of analysis suggest that mauritiana and sechellia are sister species in the melanogaster species complex, thereby supporting topology II (table 5). The closeness of branching events, however, is evident from the fact that, on the basis of the maximum-likelihood significance test, this topology is not significantly better than either of the other possibilities and that, in the bootstrap neighbor-joining and parsimony analyses, only 71% and 58% of replications, respectively, supported the mauritiana-sechellia grouping.

Adh Coding Sequence

The 837 bp of the Adh coding sequence supports topology I under all three methods of analysis, although not significantly so in the case of the maximum-likelihood analysis (table 5). However, 95% of both bootstrap-parsimony and neighbor-joining replications support the sechellia-simulans grouping, the highest percentage in any of the analyses of the melanogaster species complex.

Adh Introns

Sequences taken from the Adh introns were especially divergent and difficult to align. In an analysis of 874 bp of aligned sequence, the maximum-likelihood

...
Table 3
Amino Acid Differences between the ADH Proteins of the Species of the melanogaster Species Subgroup and Drosophila pseudoobscura

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>melanogaster</th>
<th>simulans</th>
<th>sechellia</th>
<th>mauritiana</th>
<th>yakuba</th>
<th>teissieri</th>
<th>orena</th>
<th>erecta</th>
<th>pseudoobscura</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>(1)</td>
</tr>
</tbody>
</table>

* No. of amino acids unique to each species (the no. in parentheses considers only the eight melanogaster subgroup species). (There are three coding differences between our D. yakuba Adh sequence and those of McDonald and Kreitman (1991); these are H211 to O, T213 to N, and G233 to R.)

5' Noncoding Region

Maximum-likelihood analysis of 723 bp of the 5' noncoding region found all trees to be of very similar likelihood, with highest support being given to the tree linking sechellia and simulans (topology I). The closeness of the likelihoods—they are not significantly different—and the fact that both the maximum-parsimony and neighbor-joining analyses back the mauritiana-sechellia grouping (topology II) suggests that there is not enough information in the 5' noncoding region to resolve the phylogeny, although the parsimony method found this topology in 96% of bootstrap replications. None of the topologies was significantly supported in trees reconstructed on 376 bp of sequence with the regions of difficult alignment removed.

Complete Adh-Adhr Region

The maximum-likelihood analysis of a concatenation of the five separate Adh-Adhr regions (n = 3,666)
was particularly informative, as it was the only one in which one topology — i.e., topology III grouping *D. sechellia* and *D. mauritiana* — was significantly disfavored. Furthermore, all three methods of phylogenetic reconstruction supported the grouping of *mauritiana* and *sechellia* (topology II), with 93% of bootstrap replications in the case of the neighbor-joining analysis. However, in a second analysis of the 2,944 bp of sequence that could be aligned without difficulty, the *sechellia-simulans* grouping was supported by 90% of replications in the parsimony analysis (the maximum-likelihood and neighbor-joining analyses still favored topology II). Thus the safest inference appears to be that there is a great deal of internal variation in the *Adh-Adhr* regions, with respect to the phylogenetic relationships between these species.

**Discussion**

**General Structure of the Adh Gene**

All eight species of the *D. melanogaster* species subgroup possess a single functional *Adh* gene on chromosome arm 2L. Just 300 bp or so downstream of this gene is *Adhr*, a gene of unknown function yet one that is clearly related to *Adh* in structure and sequence (see below). As summarized in figure 1, all of these genes are highly conserved in their organization, both within the *melanogaster* subgroup and between these species and members of the *obscura* species group.

For *D. melanogaster, D. simulans,* and *D. arenae* there is experimental evidence that *Adh* is transcribed from different promoters at different stages of development (Benyajati et al. 1983; Dickinson et al. 1984; Thomson et al. 1991). The distal promoter is used predominantly in adult stages, and the proximal promoter is used predominantly in larval stages (Savakis et al. 1986). The structure and sequences of the *Adh* genes of all other members of the *melanogaster* subgroup, and of the genes of the *obscura* species group, suggest that this transcriptional organization is conserved. Indeed, it appears to be conserved in all drosophilids, with the particular exception of species in the *repleta* group of the subgenus *Drosophila*. For example, the two-pro-moter arrangement has been found in *Scaptodrosophila lebanonensis* (Alabat and González-Duarte 1990; Jou et al. 1990), and in several members of the Hawaiian genus *Idiomyia* (Bishop and Hunt 1988; Rowan and Dickinson 1988; Rowan and Hunt 1991; R. Thoma and J. A. Hunt, personal communication). (We use the revised classification of the drosophilids that was proposed by Grimaldi [1990].) In *D. affinisdisjuncta* there is experimental evidence that the transcription of *Adh* is similar to that in *D. melanogaster* (Rowan and Dickinson 1988).

In the *repleta* species group of the subgenus *Drosophila* the structure of the *Adh* genes is different from that just described. These species have two functional *Adh* genes arranged in tandem (and usually a pseudogene as well). Each of these two genes is transcribed from a single promoter, with different developmental specificities (Fischer and Maniatis 1985; Atkinson et al. 1988; Menotti-Raymond et al. 1991). It is interesting that in *I. affinisdisjuncta* there is evidence of a past duplication of *Adh*. Immediately 3' to the single functional gene there is a 525-bp sequence that is a direct duplication of sequences just 5' to *Adh* itself (Rowan and Dickinson 1988). A similar duplication is seen in *I. heteroneura* (Rowan and Hunt 1991). Whether this is a relic of a functional duplication of *Adh* in Hawaiian drosophilids or simply an abortive duplication remains to be seen.

**The Adhr Gene**

The *Adhr* gene is known from species in the subgenus *Sophophora* — i.e., from species in the *melanogaster* species subgroup and the *obscura* species group (Coyne and Kreitman 1986; Schaeffer and Aquadro 1987; Marfany and González-Duarte 1991) — and from

---

**Table 4**

Amino Acid Differences between the ADHR Proteins of the Species of the melanogaster Species Subgroup and *Drosophila pseudoobscura*

<table>
<thead>
<tr>
<th></th>
<th>melanogaster</th>
<th>simulans</th>
<th>sechellia</th>
<th>mauritiana</th>
<th>teissieri</th>
<th>erecta</th>
<th>pseudoobscura</th>
<th>U^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>melanogaster</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>(5)</td>
<td>18 + 6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>simulans</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>(3)</td>
<td>16 + 6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sechellia</td>
<td>0</td>
<td>3</td>
<td>(5)</td>
<td>17 + 6</td>
<td>(17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mauritiana</td>
<td>1</td>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>teissieri</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>erecta</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pseudoobscura</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15 + 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.**—The data for *erecta* are incomplete (see text). The six extra amino acids at the carboxy-terminal of the *pseudoobscura* protein are counted separately.

^a No. of amino acids unique to each species.
Table 5
Results of Phylogenetic Analyses of Different Sequence Data Sets from the *melanogaster* Species Complex

<table>
<thead>
<tr>
<th>Topology</th>
<th>Topology I ((((sechellia.simulans) mauritiana)melanogaster)</th>
<th>Topology II ((((sechellia.mauritiana) simulans)melanogaster)</th>
<th>Topology III ((((simulans.mauritiana) sechellia)melanogaster)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh (n=771)</td>
<td>-6.733 ML: -2,591.114 MP: 58%</td>
<td>-9.848 ML: -2,009.739 MP: 95%</td>
<td>-7.935 ML: -2,437.258 MP: 95%</td>
</tr>
<tr>
<td>Adhr (n=837)</td>
<td>-9.848 ML: -2,009.739 MP: 95%</td>
<td>-9.848 ML: -2,009.739 MP: 95%</td>
<td>-0.170 ML: -1,517.044 MP: 96%</td>
</tr>
<tr>
<td>Adh introns (n=874)</td>
<td>-4.472 MP: 87%</td>
<td>-6.733 ML: -2,591.114 MP: 58%</td>
<td>-8.170 MP: 95%</td>
</tr>
<tr>
<td>Adhr introns (n=511)</td>
<td>-6.517 MP: 74%</td>
<td>-9.848 ML: -2,009.739 MP: 95%</td>
<td>-0.839 MP: 96%</td>
</tr>
<tr>
<td>5' Noncoding (n=723)</td>
<td>-6.517 MP: 74%</td>
<td>-9.848 ML: -2,009.739 MP: 95%</td>
<td>-0.170 MP: 96%</td>
</tr>
<tr>
<td>Complete Adh-Adhr region (n=3,666)*</td>
<td>-11.638 ML: -8,204.887 MP: 76%</td>
<td>-23.174-* ML: -8.204.887 MP: 76%</td>
<td></td>
</tr>
</tbody>
</table>

NOTE.—The most favored topology, as inferred by three different methods of phylogenetic analysis, is denoted as follows: ML = maximum likelihood; MP = maximum parsimony; and NJ = neighbor joining. For the ML analysis the log likelihood of the ML topology is presented along with the difference in likelihood between this and the competing two topologies. For the ML and NJ analyses, the percentage of bootstrap replications (of a total of 2,000 replications) supporting the most favored topology is also shown. n = length of sequence analyzed.

* Total is less than that of the five regions combined, because some gaps introduced when *Drosophila erecta* was used as the outgroup are removed when *D. teissieri* is used as the outgroup.

b Topology that is found to be significantly worse than the ML topology.

S. lebanonensis (E. Juan, personal communication). In these species it is conserved in structure, sequence, and position, relative to *Adh* (see fig. 1). The predicted protein product of *Adhr* is clearly related to that of *Adh*, in both primary sequence and secondary structure (see fig. 5 of Schaeffer and Aquadro 1987). The overall level of similarity between the two proteins is about 36%. The positions of the two introns within *Adhr* are very similar to those of introns 2 and 3 in *Adh*. These features all suggest an origin of *Adhr* from *Adh* by duplication (or vice versa) (Schaeffer and Aquadro 1987).

The function of *Adhr* and its protein is enigmatic, although its sequence is well conserved. It is not an alcohol dehydrogenase (ADH), or at least not one with significant activity with short-chain primary or secondary alcohols as substrates. This is because no ADH activity can be detected in many different *D. melanogaster* Adh-null mutant strains. In *D. melanogaster*, at least, *Adhr* is not a vital gene, since many deletions that include both it and *Adh* are viable when homozygous (for genetic evidence that these deletions remove *Adhr*, see, e.g., Ashburner et al. 1982; for molecular evidence that these deletions remove *Adhr*, see, e.g., Chia et al. 1985). *Adhr*, like *Adh*, is almost certainly included within an intron of the *outspread* gene, at least in *D. melanogaster* (Chia et al. 1985; S. McNabb and M. Ashburner, unpublished data).

ADH is an NAD-dependent dehydrogenase and, in common with other enzymes of this class, has the characteristic glycine residues at the NAD-binding site (see Persson et al. 1991). The amino acid sequence FVAGLGGIG (residues 11–19) is conserved in 35 of 36 *Drosophila* ADH sequences. Gly14 is known to be essential for both ADH activity and NAD binding, since the EMS-induced *Adhn* mutation of *D. melanogaster* is substituted by an aspartic acid residue at this site. ADH from this mutant strain is inactive and has a lower affinity for 5'-AMP, a competitive inhibitor of coenzyme binding (Thatcher 1980; Thatcher and Retzios 1980).

The predicted ADHR protein lacks Gly14; the amino acid sequence in all eight known *Adhr* genes in this region is YVADCGGIA. Of the three glycyl residues (Gly14, Gly17, and Gly19) conserved in NAD-dependent dehydrogenases, only Gly17 is conserved in ADHR. (We should say that the ADH of *S. lebanonensis* also lacks Gly14, it being replaced by an alanyl residue [see Villarroya and Juan 1991]. However, *D. melanogaster* ADH in which Gly14 is replaced [in vitro] by Ala14 is an active enzyme, albeit with a somewhat reduced specific activity and decreased thermal stability [Chen et al. 1990]).

These considerations indicate that, whatever its function, ADHR is not an NAD-dependent dehydrogenase or, if it is, then it is a most unusual one. That
being said, AHDR does share other conserved amino acid residues with the NAD-dependent short-chain dehydrogenases. From a comparison of 20 such protein sequences (from bacteria to humans and including D. melanogaster ADH), Persson et al. (1991) have identified nine amino acid residues (in addition to those of the NAD-binding domain) that are very highly conserved. ADHR has all of these except one: the conserved Gly94 is replaced by threonine in all known ADHR sequences. Asp38, conserved in many but not in all NAD-dependent dehydrogenases (Persson et al. 1991), is Glu38 in Adhr. Substitution of Asp38 by Asn38 allows Drosophila ADH to use both NAD+ and NADP+ as cofactors, but it increases the thermolability of the enzyme in the absence of cofactors (Chen et al. 1991).

ADHR is extended, with respect to ADH, at its carboxy-terminus. Indeed the C-terminus of ADHR is remarkably variable between species (see above). ADHR also differs from ADH in having a rather high (4.4%) methionine content. This is interesting because Henikoff and Wallace (1988) identified, in the fly Sarcophaga (Matsumoto et al. 1985), a fat-body protein that is about as related to ADH as ADH is to ADHR. More recently, Rat et al. (1991) have cloned and sequenced the homologous gene from D. melanogaster. It encodes the fat-body protein FP6. The function of this protein is not known, yet it has an extraordinarily high (20%) methionine content. The similarities between FP6, ADH, and ADHR extend beyond the 28%–30% identity of amino acids, for the introns of Adh and Adhr are also conserved in position in the Sarcophaga homologue of Fp6 (Rat et al. 1991).

The Adhr gene has not yet been sequenced from any species of the subgenus Drosophila (although Visa et al. [1991] imply that it is found in D. immigrans). If Adh and Adhr are related by descent, i.e., one arose by duplication of the other, then this event must predate the radiation of the drosophilids. This must be so, in light of the very high divergence (only about 36% similarity) between the ADH and ADHR proteins. The most obvious hypothesis is that in the subgenus Drosophila (or at least in those species that have been studied) Adhr is no longer directly adjacent to Adh.

Phylogenetic Relationships of the Adh Pseudogenes

All the trees constructed from the Adh coding region suggest that the teissieri and yakuba pseudogenes have a single common origin, grouping together on the phylogenetic tree rather than with their functional counterparts. This coincides with the molecular evidence for
single origin of the pseudogenes (Jeffs and Ashburner 1991). Both the maximum-likelihood and neighbor-joining phylogenetic trees depict an unrealistic topology whereby the functional Adh gene from teissieri gave rise to the pseudogenes from both teissieri and yakuba. This is likely to be a reflection of the fact that the pseudogenes originated very shortly before the species themselves diverged (see Jeffs and Ashburner 1991). Such a speciation event, coupled with the change in rates of nucleotide substitution concomitant with a change in gene function (i.e., the production of a processed pseudogene), is likely to muddy any picture of phylogenetic relationships. The acceleration of rate along the lineage leading to the pseudogenes is clearly depicted in the long branch that connects them to their functional counterparts.

### Synonymous and Nonsynonymous Substitutions

We have estimated the frequencies of substitution at synonymous and nonsynonymous codon positions within both Adh and Adhr, using the weighted-pathway method of Li et al. (1985). The frequencies of synonymous and nonsynonymous substitution are very different between Adh and Adhr. To take only the comparisons between the melanogaster subgroup sequences and the sequences of \( D. \) pseudoobscura, \( K_s \) is in the range 0.56–0.65 for Adh but is in the range 1.27–1.42 for Adhr; \( K_A \) is in the range 0.044–0.053 for Adh and in the range 0.031–0.037 for Adhr (table 1). Thus, at the protein level, Adhr is evolving slightly, though not significantly, more slowly than Adh; but, with respect to changes that do not affect the protein, the reverse is true.

Shields et al. (1988; also see Sharp and Li 1989) have pointed out that, in *Drosophila*, there is a negative correlation between \( K_s \) and codon bias; that is, there is a selective constraint against change at silent sites. Genes that are highly constrained with respect to codon choice have a lower rate of synonymous substitution than do genes that are less constrained. We have estimated the extent of codon bias for Adh and Adhr genes (table 2) (as did Shields et al. 1988). It is clear that Adh genes are, in general, more biased in their codon choice than are Adhr genes, confirming the hypothesis of Shields et al. (1988). It is interesting that the codon bias of Adh is higher, but that of Adhr is lower, in the melanogaster subgroup than in \( D. \) pseudoobscura. However, the codon bias of the Adh gene of \( D. \) ambigua, a European representative of the obscura species group, is similar to that of *D. melanogaster*.

A major contributor to codon bias in these genes is that A is very rarely used in the third position of codons. In Adh genes the third base is A in only 3.15%–4.67% of cases; in Adhr this frequency range is 9.68%–19.05%. Starmer and Sullivan (1989) point out that the very biased codon use of the sophophoran Adh genes is not characteristic of these genes in the subgenus *Drosophila*.

Although the correlation between high codon bias and low rates of substitution at synonymous codon sites is very clear, and must be so if codon bias is to be preserved over time, the biological significance of the bias is less so. Shields et al. (1988) suggest that genes with low codon bias are expressed at a lower level than are those with high codon bias. This may be so, although it is very difficult to obtain good data on relative rates of gene expression in multicellular organisms.

### Relationships within the melanogaster Species Complex

In this study we have assessed, using six DNA sequence data sets from the Adh-Adhr region and three different methods of phylogenetic reconstruction, the relative merits of three hypotheses of relationship of the four species of the melanogaster complex—i.e., topologies I–III (table 5). In all cases it was assumed that melanogaster is the sister group to the three other species. Such an assumption not only conforms to evidence from other studies (see Lachaise et al. 1988) but also to all the trees inferred in the present study (with a high degree of significance).

Despite the fact that a total of 3,666 sequence positions were analyzed, no single topology was conclusively favored, although topology III, placing mauritiana and simulans as the closest species pair, was consistently the least favored, significantly so in the maximum-likelihood analysis of the complete Adh-Adhr region. This suggests that within the Adh-Adhr region there is simply not enough evolutionary information to resolve the phylogeny of the melanogaster species complex (for a similar conclusion, see Coyne and Kreitman 1986); and what information there is is not consistent in its support, since different parts of the Adh-Adhr region analyzed by different methods give their support to different hypotheses of relationship. For example, in an analysis of the complete Adh-Adhr region, with the areas of difficult alignment removed \( (n = 2,944) \), both the maximum-likelihood and neighbor-joining analyses support the sechellia-mauritiana grouping, yet, for the same data set, the parsimony analysis supports, with 90% of bootstrap replications, the simulans-sechellia grouping. Such a situation also illustrates the differences between methods of phylogenetic reconstruction and argues against the overinterpretation of results obtained from a single analysis.

Therefore, perhaps the most reasonable interpretation of the phylogenetic relationships of the melanogaster species complex as inferred in this study is that *D. melanogaster* itself is the most divergent of the four species, with the remaining three—mauritiana, sechellia,
and simulans—separating from each other within a very short time period, hence their appearance as a trichotomy on the present data (see Coyne and Kreitman 1986). The fact that two of these species, i.e., D. sechellia and D. mauritiana, are endemic to small, relatively isolated, islands, while the third, D. simulans, is broadly distributed in East Africa and Malagasy, would suggest that topology II, with D. simulans being the ancestral species, is correct.

Phylogenetic Relationships of the D. melanogaster Species Subgroup

The center of diversity of the melanogaster species subgroup is in western Africa, where four species—erecta, orena, yakuba, and teissieri—as well as melanogaster itself are now found. Trees derived from three different methods of phylogenetic reconstruction—i.e., maximum likelihood, maximum parsimony, and neighbor joining—all show that the first cladistic event within the subgroup was that which lead to the separation of the erecta and orena lineages. It is just these two species that, in the present day, are restricted, in their geographical distribution, to western Africa. The next cladistic event is the separation of yakuba-teissieri from the lineage that gave rise to melanogaster and its three closest relatives, i.e., simulans, sechellia, and mauritiana. Today, both yakuba and teissieri are found in a broad swath across western and eastern central Africa. The topologies of the trees derived from phylogenetic reconstructions of the Adh sequences are identical to that suggested by Lachaise et al. (1988) on the basis of a variety of other data, principally biogeography and polytene chromosome banding patterns.

Acknowledgments

This work was supported by an MRC Programme Grant to M.A. P.S.J. was supported by a Research Studentship from the Science and Engineering Research Council. We thank Roger Staden, Des Higgins, Martin Bishop, and Ken Wolfe for programs. We also thank K. Maruyama, S. Schaeffer, and M. Kreitman for sequence data prior to their publication, and we thank J. Felsenstein for discussions.

LITERATURE CITED


Felsenstein, J. 1993. PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.


Rowan, R. G., and J. A. Hunt. 1991. Rates of DNA change and phylogeny from the DNA sequences of the alcohol de-


