J Mol Evol (1994) 38:443–454

Journal of Molecular Evolution

Phylogeny of *Drosophila* and Related Genera Inferred from the Nucleotide Sequence of the Cu,Zn *Sod* Gene

Jan Kwiatowski,[1,2] Douglas Skarecky,[1] Kevin Bailey,[1] Francisco J. Ayala[1]

[1] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92717, USA
[2] Institute of Botany, Warsaw University, 00-478 Warsaw, Poland

Received: 20 March 1993 / Revised: 31 August 1993 / Accepted: 30 September 1993

Abstract. The phylogeny and taxonomy of the drosophilids have been the subject of extensive investigations. Recently, Grimaldi (1990) has challenged some common conceptions, and several sets of molecular data have provided information not always compatible with other taxonomic knowledge or consistent with each other. We present the coding nucleotide sequence of the Cu,Zn superoxide dismutase gene (*Sod*) for 15 species, which include the medfly *Ceratitis capitata* (family Tephritidae), the genera *Chymomyza* and *Zaprionus,* and representatives of the subgenera *Dorsilopha, Drosophila, Hirtodrosophila, Scaptodrosophila,* and *Sophophora.* Phylogenetic analysis of the *Sod* sequences indicates that *Scaptodrosophila* and *Chymomyza* branched off the main lineage before the major *Drosophila* radiations. The presence of a second intron in *Chymomyza* and *Scaptodrosophila* (as well as in the medfly) confirms the early divergence of these two taxa. This second intron became deleted from the main lineage before the major *Drosophila* radiations. According to the *Sod* sequences, *Sophophora* (including the *melanogaster, obscura, saltans,* and *willistoni* species groups) is older than the subgenus *Drosophila;* a deep branch splits the *willistoni* and *saltans* groups from the *melanogaster* and *obscura* groups. The genus *Zaprionus* and the subgenera *Dorsilopha* and *Hirtodrosophila* appear as branches of a prolific "bush" that also embraces the numerous species of the subgenus *Drosophila.* The *Sod* results corroborate in many, but not all, respects Throckmorton's (King, R.C. (ed) *Handbook of Genetics.* Plenum Press, New York, pp. 421–469, 1975) phylogeny; are inconsistent in some important ways with Grimaldi's (*Bull. Am. Museum Nat. Hist.* 197:1–139, 1990) cladistic analysis; and also are inconsistent with some inferences based on mitochondrial DNA data. The *Sod* results manifest how, in addition to the information derived from nucleotide sequences, structural features (i.e., the deletion of an intron) can help resolve phylogenetic issues.

Key words: Superoxide dismutase gene — *Drosophila* phylogeny — Nucleotide sequence — Medfly *Ceratitis capitata* — Intron evolution — G + C content

Introduction

The taxonomy and systematics of *Drosophila* have been the subject of many investigations. A few landmarks are the monographs by Sturtevant (1921), Patterson and Stone (1952), Throckmorton (1975), and Wheeler (1981, 1986). Important recent contributions include a cladistic and revisionist monograph by Grimaldi (1990), and several molecular studies, the most notable and inclusive of which is DeSalle and Grimaldi (1991).

Throckmorton's (1975) assessment of previous taxonomic, phylogenetic, and biogeographic studies moved him to conclude that the genus *Drosophila* originated in the Old World tropics, probably in Asia. Throckmorton's other important conclusions include that (1) the first major radiation of the genus is represented by the subgenus *Scaptodrosophila,* primarily distributed throughout the Old World tropics from Africa to Australia and the Pacific, although some species groups

Correspondence requests to: F. J. Ayala

(including *subtilis* and *victoria*) occur in the New World; (2) the radiation of the subgenus *Sophophora* (comprising the *melanogaster, obscura, saltans,* and *willistoni* groups) preceded the radiation of the subgenus *Drosophila;* (3) the genus *Chymomyza* is part of the *Sophophora* radiation; and (4) the genus *Zaprionus* emerged as part of the *Drosophila* subgenus radiation (which also includes the subgenera *Hirtodrosophila* and *Dorsilopha*).

Grimaldi (1990) has carried out a cladistic analysis of morphological characters and produced a phylogeny that challenges Throckmorton's conclusions in important respects; in particular, Grimaldi places *Chymomyza, Zaprionus,* and *Hirtodrosophila* outside the lineage of the genus *Drosophila.* He also places *Scaptodrosophila* outside the *Drosophila*-genus lineage (thus agreeing with Throckmorton) and raises it (as well as *Hirtodrosophila*) to the genus category. DeSalle and Grimaldi (1991) as well as DeSalle (1992) have shown that molecular data (derived particularly from mitochondrial DNA) disagree with some of Grimaldi's (1990) conclusions.

We present here the DNA coding sequence of the gene *Sod* (which codes for the Cu,Zn superoxide dismutase) in 15 species representing the drosophilid genera and subgenera just mentioned. Our results are largely consistent with the phylogenetic relationships proposed by Throckmorton (1975)—more so, in fact, than with those proposed by Grimaldi (1990) or DeSalle and Grimaldi (1991). The propitious discovery of a second intron, present in the medfly *Ceratitis capitata* (family Tephritidae) as well as in *Scaptodrosophila* and *Chymomyza,* places the latter two taxa outside the genus *Drosophila.* The absence of this second intron from *Hirtodrosophila* also locates the branching of this taxon after the split of *Chymomyza* from the genus *Drosophila,* thus contradicting the mtDNA-based conclusion of DeSalle and Grimaldi (1991) and DeSalle (1992).

Materials and Methods

Species. The 15 species studied are listed in Table 1. Strains of *Chymomyza amoena, C. procnemis, Zaprionus tuberculatus, Drosophila busckii, D. lebanonensis, D. pictiventris, D. quinaria,* and *D. virilis* were obtained from the National Drosophila Species Stock Center at Bowling Green, Ohio; *D. hydei* and *D. saltans* were obtained from the Stock Center at Indiana University, Bloomington; other *Drosophila* species derived from cultures available in our laboratory; for the source of the *Ceratitis capitata* DNA see Kwiatowski et al. (1992a).

DNA Preparation, Amplification, Cloning, and Sequencing. We prepared genomic DNA from about 10–20 flies following the method of Kawasaki (1990). The *Sod* gene was amplified by the high-fidelity PCR technique and cloned into plasmids (puc19 or puc21) (Kwiatowski et al. 1991b). Double-stranded DNA templates were sequenced as described earlier (Kwiatowski et al. 1992a). The primers for PCR amplification were designed by comparing available dipteran *Sod* sequences, using conserved regions of *Sod* as well as a downstream unidentified gene (Fig. 1). Three oligonucleotides were used for amplification and sequencing (5′ → 3′): *N,* CCTCTAGAAATGGTGGTTAAAGCTGTNTGCGT; *C,* CTTGCTGAGCTCGTGTCCACCCTTGCCCAGATCATC; and *O,* ACGGAAGTCTAGAAGGGCTTTTTGGGCTTTGCCACCTG. Three additional oligonucleotides were used only for sequencing: *I,* GACATGCAGCCATTGGTGTTGTC; *IR,* GACAACACCAAYGGCTGCATGTC; and *CR,* CAAGGGTGGACACGAGCTGAGCAAG. The *IR* primer failed in three species (*C. procnemis, Z. tuberculatus,* and *D. lebanonensis*), for which it was replaced by IR140 (TGTACCTTCGGCACGTCTGG). In addition we used standard M13 sequencing primers. In some cases (nine species) primers that were different for different species (but are all represented by *A, B,* and *D* in Fig. 1) were designed using noncoding gene regions so as to sequence the coding fragments from both DNA strands. All compressions and ambiguities were resolved by multiple sequencing of both strands.

Computer-Assisted Sequence Analysis. DNA and protein sequences were assembled and analyzed using the Darwin package written by Mr. Robert Tyler from our laboratory. Phylogenetic analyses were made with the PHYLIP 3.4 and 3.5c package (Felsenstein 1989). The codon usage table was computed with the CODONS program (Lloyd and Sharp 1992).

Results

Structure of the Sod *Gene*

The structure of the *Sod* gene is outlined in Fig. 1. In all species the coding sequence is interrupted after the 22nd codon by an intron 300–700 bp in length. Several *Drosophila* species that we had earlier sequenced exhibit no other intron; but a second short intron (<100 bp), between codons 95 and 96, occurs in *Chymomyza* and in the medfly, *Ceratitis capitata* (Kwiatowski et al. 1992a,b), which belongs to a different dipteran family.

D. lebanonensis, a species of the subgenus *Scaptodrosophila,* also exhibits the second intron, which is, however, absent from *Zaprionus tuberculatus* as well as from all other *Drosophila* species now sequenced.

Nucleotide Sequences

The 15 nucleotide sequences of the *Sod* coding region are given in Fig. 2. The following sequences have been published before: *Drosophila melanogaster* (Kwiatowski et al. 1989b), *D. simulans* (Kwiatowski et al. 1989a), *D. virilis* (Kwiatowski and Ayala 1989), *Chymomyza amoena* (Kwiatowski et al. 1992b), and *Ceratitis capitata* (Kwiatowski et al. 1992a).

The noncoding regions are not shown in Fig. 2. They were not used for phylogenetic analysis because they are so highly diverse that their alignment becomes uncertain in many cases. The primer sequences are not shown either. The complete coding region is amplified by means of the N and O primers (see Fig. 1), which yield single or multiple PCR bands 1,150–1,850 bp in size

Table 1. Taxonomy of the 15 species according to Wheeler (1981)

Family	Genus	Subgenus	Group	Species
Tephritidae				
	Ceratitis			*capitata*
Drosophilidae				
	Chymomyza			*amoena*
				procnemis
	Zaprionus			*tuberculatus*
	Drosophila			
		Dorsilopha		*busckii*
		Hirtodrosophila[a]		*pictiventris*
		Scaptodrosophila[a]		*lebanonensis*
		Drosophila		
			quinaria	*quinaria*
			repleta	*hydei*
			virilis	*virilis*
		Sophophora		
			melanogaster	*melanogaster*
				simulans
			obscura	*subobscura*
			saltans	*saltans*
			willistoni	*willistoni*

[a] Raised to genus category in the revision by Grimaldi (1990)

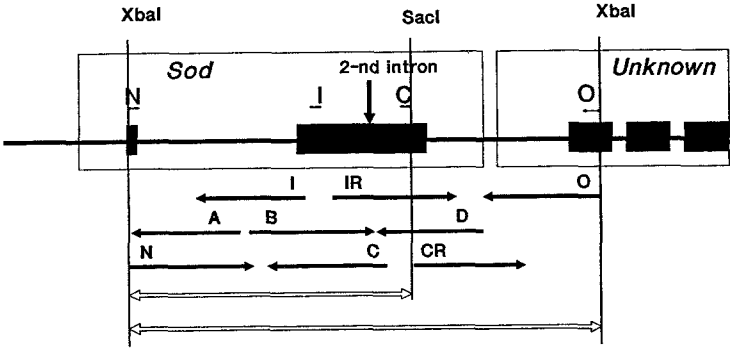

Fig. 1. Structure of the Cu,Zn *Sod* gene region and strategy for amplification, cloning, and sequencing. The *black boxes* represent exons: the three on the right represent an ORF (open reading frame) with unidentified function. The *vertical arrow* indicates the position of a second intron present in some species. The two *hollow arrows* represent the segments amplified, cloned, and sequenced. The *thinner arrows* indicate the extent and direction of sequencing.

that were cloned into *Xba*I sites. The O primer, however, does not work for *D. pictiventris* and *D. subobscura.* These two species have been amplified by means of the N and C primers (and cloned into *Xba*I and *Sac*I sites), and thus lack 96 bp in the carboxyl end of the coding region; 439 bp are listed for all other species.

The sequences have been obtained by double-strand sequencing of single clones. This procedure fixes PCR-derived nucleotide errors. However, we have determined in a preliminary experiment (Kwiatowski et al., 1991b) that the cumulative transversion-plus-transition error generated by our procedures is 3×10^{-4}. This rate would be expected to yield fewer than two erroneous nucleotide determinations in the whole data set given in Fig. 2, which is trivial compared to the average of about 100 bp differences between species pairs.

Table 2 gives the number of pairwise nucleotide differences out of the 439 bp sequenced in the 13 species for which the sequences are complete. The table also gives the number of inferred amino acid differences out of the 153 encoded by the gene. (The first seven amino acids covered by the N primer are identical in the six species for which sequences have been published, including *C. capitata,* except for *C. amoena,* which differs by one amino acid from all others.)

Table 2 shows that *D. lebanonensis* is more different from *Zaprionus* and all other *Drosophila* species than any of these is from the rest. For example, the average number of nucleotide differences between *D. lebanonensis* and the other *Drosophila* species plus *Zaprionus* is 107, whereas it is less than 100 between any of these other species. This result is consistent with the presence, noted above, of a second intron in the *D. lebanonensis Sod* gene, which places *Scaptodrosophila* (as well as the genus *Chymomyza,* which also has the second intron) outside the subgenera *Sophophora* and *Drosophila.* Since the genus *Zaprionus* shares with all other *Drosophila* species the lack of the second intron, it can be concluded that *Zaprionus* is closer to the *Sophophora* and *Drosophila* subgenus than either *Chymomyza* or *Scaptodrosophila.* The distances shown in Table 2 are consistent with this conclusion.

```
D.melanogaster  AATTAACGGC GATGCCAAGG GCACGGTTTT CTTCGAACAG GAGAGCAGCG GTACGCCCGT  60
D.simulans      .......... .......... .......... .......... .......... ..........
D.subobscura    .......... .......... ....C..... ......G... ....CT.... AGG.T.....
D.saltans       T.....T... .......... .T..C..... ...T...... ...G.T.A.. .AG.T..T..
D.willistoni    T.....T... .......... .T..C..... ...T..G... ...GA..AT. .AG.T.....
D.virilis       T..C..T... .......... ....A..... .......... ...G..GAG. ..TGC.....
D.hydei         T..C..T... ..C....... .T..A..... .......... ...TCGGA.. ..TGC..G..
D.busckii       T.....T..T .......... ....A..... ......G..A ...TCGGAGA AGTGC.....
D.quinaria      T......... ........A. ....A..... ...T...... ...GCTGAA. ...GC.....
D.pictiventris  T.....T... .......... .T..C..... T.....G... ...GCTGAAA AC.GT.....
Zaprionus       TC.A.....A ........A. ....T..C.. T..T...... C..GCTGCA. AGTGT.....
C.amoena        T.....T..T ........A. .T..C..C.A T..T.....A ...GATGCTT .CG.T..A..
C.procnemis     G.....T... .......... ....T..C.A ...T...... ...T..GCA. ..G.T..A..
D.lebanonensis  T.....T..A .....T.... .G..C....A T..T..G... ....AAGAA. .CG.T..T..
Ceratitis       .........A ..C.TT.... .A..C...CA ...T...... C..GATGCAA AGT.T.....

D.melanogaster  GAAGGTCTCC GGTGAGGTGT GCGGCCTGGC CAAGGGTCTG CACGGATTCC ACGTGCACGA 120
D.simulans      .......... .......... .......... .........C .......... ..........
D.subobscura    ...A...A.. ..G......C TG.....T.. ......C... .......... .T..T.....
D.saltans       C..A..TA.. ..C......A CT..TT.... T.........C ..T..C.... .T..C..T..
D.willistoni    C..A..TA.. ..C......A C.........G ...A..A... ..T..T.... .T..C..T..
D.virilis       ......TA.. ..C.....AA C....T.... A........A. ..T..C.... .T.....T..
D.hydei         ...A..TA.. ..A..A..AA C....T.A.. G........A. ..T..C.... .T.....T..
D.busckii       T.....TA.. ..C.....AA CA........ .......... ..T..C.... .T........
D.quinaria      ...A...A.. .....A..AA AT..G..T.. GCCA...... ..T..C.... .T........
D.pictiventris  A..A..AA.. ..A..A...G CT..T..C.. T..A..C... .....C.... .T........
Zaprionus       T.....GA.A .........A CT..A...T. A.......... ..T..T.... .T..T..T..
C.amoena        A..A..A.GT ..A...A.TA CT..AT..AA ........A. .....T.... .T..A..T..
C.procnemis     T..A..G.G. ......A.AA CT...T..AG ......C.A. ..T..... .T.....T..
D.lebanonensis  C..A..GA.T ..C..A..AA A...G....A T.......... .....C.... .T.....T..
Ceratitis       CCTA...A.T .........TA A...A..T.. T.......... ..T..C.... .T..A.....

D.melanogaster  GTTCGGTGAC AACACCAATG GCTGCATGTC GTCCGGACCG CACTTCAATC CGTATGGCAA 180
D.simulans      ...T..A... .......... .......... .........C .......... ..........
D.subobscura    ......G... .....T..C. .......... ......G..C .....T.... .TCGCAA...
D.saltans       A..T..C... ..T....... .......... A..T..T..C .......... .CC..A....
D.willistoni    A..T..C... ........C. .A........ C..T..C..C ..T....... .AC..A....
D.virilis       ......C... .......... .......... ...G...... .......... .C..CCAG..
D.hydei         ...T..C... ........C. .......... C..T...... ..T..T.... .A..CCAG..
D.busckii       A..T..C... .....A.... .......... ...T..C..C .......... .TC.A.....
D.quinaria      ...T..C... .....A.... .......... ...T......A ..T....... .CC..AAA..
D.pictiventris  A..T..G... .......... .T........ A..T..G..A .....T.... .TC.CAAG..
Zaprionus       ...T..A..T ..T..A.... .A........ C..T...... .....T..C. .T...CAG..
C.amoena        A..T..A... .....A.... .G........ .........C ..T....... .C.TGAAT..
C.procnemis     A..T..C..T ..T..A.... .T........ A..G.....A ..T....... .C.TGAAT..
D.lebanonensis  ...T..A..T ..T.....C. .......... CG.T......T ..T....... .CC.CAA...
Ceratitis       A.....G... ........C. .T....CT.. .G.T..C..T ..T.....C. .T.....A..

D.melanogaster  GGAGCATGGC GCTCCCGTCG ACGAGAATCG TCACCTGGGC GATCTGGGCA ACATTGAGGC 240
D.simulans      ......C... .......... .......C.. .......... ...T...... ..........
D.subobscura    ......C..T ..C..AAC.. .......... C...T..... ..CT...... ....CC....
D.saltans       .........T .....T.A.. .T..C..C.. ...TT..... ...T...... .T........
D.williston     .......... ..C..T.GA. .T........ ...TT..... ...T...... ..........
D.virilis       .......... ..A..GAC.. .......... C..T...... ...T...... ....CATC..
D.hydei         ......C... .....AAAT. .T........ ...TT....T ..C..... ....AATC..
D.busckii       .........T ..A...ACT. .T.....C.. C..TT....T ...T...... ....AACT..
D.quinaria      ...A.....T ..A...AC.. ...GTG.... C..TT....T .........T. ....AACC..
D.pictiventris  .........T ..A...ACT. ....CG.... ...TG....T ...T...... .....ACTT.
Zaprionus       ...A...... .......GA. .T........ C..TT....A ..CT....T. ....AACT..
C.amoena        A..A.....T ......ACT. .T........ ...TT....T ..C..A.... .T.....A..
C.procnemis     .........T ..A..A.CT. .T..A..... ...TT....T ...T...... .......A..
D.lebanonensis  ......C... AGC..TTG.. .......C.. C.........T ..CT...... .......A..
Ceratitis       TTC......T ..A..ATC.. .TTT...C.. ...TT..... ...T.A.... .T.....A..

D.melanogaster  CACCGGCGAC TGCCCCACCA AGGTCAACAT CACCGACTCC AAGATTACGC TCTTCGGCGC 300
D.simulans      .......... .........A .......... .......... .....C.... ..........
D.subobscura    TG....A... A........GG CC..T.GT.. .......... .....C.... .G.....T..
D.saltans       A..T..A..T G.....A..T. .A..G..... T..T..T... ...T.A.... .AG.T..T..
D.willistoni    AT.T...AGT G.T..A..TG CT..T..T.. ......T... ..A..C..C. .AG.....T..
D.virilis       ..AT..A... G.T.....TC CC..G..... .TG...T.G. .....C..AT .GC.......
D.hydei         ..GT.....T G.T.....TC C...TG.... ATGT..T.G. ..A..A..AT .A..T..A..
D.busckii       ...A.....T G.......TG CA..GG.... TTG...T.G. .........A .G..T..T..
D.quinaria      .T....T..T G....G..TC C...T..T.. A.G....AA. C........A. ....T....A
D.pictiventris  ..TA.....T G....T...G CA..A..T.. ..G...TAG. C.....C..A. ....T..T.A
Zaprionus       .T....T..T G.A..T...G C....G.... TTG...T.GT .....C.... .T..T..T.A
C.amoena        TC.A.....T G.T.....A. ....GTG... T.A......T ..A.....CT .G.....T.A
C.procnemis     AC.......T G.T.....A. ....GTG... T.A...T.G. ..A..A.... .G..T..T.A
D.lebanonensis  .G.T...A.. .CTG.T.... .A........ T.....T.GT C....A..AT .G.....A..
Ceratitis       ..GT.....T G.AG.A..G. .....G.A.. ..G....AAG CT......CT .G..T..T.A
```

Fig. 2. Nucleotide sequence of the coding region of the Cu,Zn *Sod* gene from 15 species. The sequences begin 23 bp downstream from the start of the coding region and end at the stop codon. The sequences of *D. subobscura* and *D. pictiventris* are missing the last 96 bp (indicated by *dashes*) because they failed to be amplified by the O primer. (See Fig. 1.) *Dots* indicate nucleotides identical to those of *D. melanogaster*. The species are as in Table 1.

G + C Content of the Sod *Gene*

Table 3 gives the G + C content for the 343 bp sequenced in all 15 species (column labeled *b*), as well as for the complete set of nucleotides sequenced in 13 of the species (column *a*). The percent G + C of any given species is virtually identical in both sets. When all codon sites are taken into account, the G + C content is about 50% in all species, except *melanogaster, simulans, subobscura,* and *virilis,* for which it is about 60%.

```
D.melanogaster  CGACAGCATC ATCGGACGCA CCGTTGTCGT GCACGCCGAT GCCGATGATC TTGGCCAGGG 360
D.simulans      .......... .......... .......... .......... .......... .C........
D.subobscura    .......... ..T....... ....G..... T......... ..T------- ----------
D.saltans       T..T...... ..T....... .A.....T.. C..T...... C........CT .G...A....
D.willistoni    .A.T..T... ..T....... .......... C..T...... C.......... .G...A....
D.virilis       .A.T.....T ..T....... .......G.. .......... C.A.....C. .G...A....
D.hydei         .A.T.....T .......... .T.....G.. .......T... C........C. .C...A.A..
D.busckii       .A........ ..T....... .T..G..G.. A..T...... C.........T .G...A....
D.quinaria      GA........ ..T....... .A.....T.. .......... C....C.... .C...A....
D.pictiventris  AA.......T ..T....... .A.....G.. ...T...... C..------- ----------
Zaprionus       AA.T...... ..T....... .T........ C......T... C.......... .G...A....
C.amoena        GA........ G.T....... .G..G..A.. T..T...... C.T.......T .G...A....
C.procnemis     GC.T.....T G.T....... .G..G..A.. ...T...... C.A.......T .G...A....
D.lebanonensis  .A........ ..T....... .TA.A..T.. T..T..T... C.........T .G...A.A..
Ceratitis       AA.T...... G.T.....T. .TA....T.. A..T..T... C.T.......T .G..TA.A..

D.melanogaster  TGGACACGAG CTGAGCAAGT CAACGGGCAA CGCTGGTGCC CGCATCGGGT GCGGCGTTAT 420
D.simulans      .......... .......... .......... .......... .........C .........
D.subobscura    ---------- ---------- ---------- ---------- ---------- ----------
D.saltans       C.....T... .........A. .G..T..T.. T......... .....T..T. ....T..C..
D.willistoni    .......... .........A .G..T..T.. T......A..T .....T..T. ....T..C..
D.virilis       C......... .........A ....T..... ...G..A..T .....T..C. ....T..C..
D.hydei         ...C..T... ..T......A ....T..A.. .........T .....T..A. ....T..A..
D.busckii       C......... ..A......A ....T..A.. T......C..T .....T..A. ....T..C..
D.quinaria      ......T... T.........A ....T..... T......C..T .....T..A. ....T..C..
D.pictiventris  ---------- ---------- ---------- ---------- ---------- ----------
Zaprionus       ...C...... T......... ....T..A.. T.........T ........T. ....T..C..
C.amoena        ......T... T......... ....A..A.. T......... ..T..T..T. ....T..A..
C.procnemis     .......... .......... .C..A..... T..C..C..A ..T..T..C. ....T..C..
D.lebanonensis  ...C..T... .......... .G..T..A.. T......A..A ..T..T..A. ....A..C..
Ceratitis       ...T..T..A T.A......A ....T..... ......G..T ...C.T..T. .T.....C..

D.melanogaster  TGGCATTGCC AAGGTCTAA 439
D.simulans      .......... .........
D.subobscura    ---------- ---------
D.saltans       ...T...... ..AA.....
D.willistoni    .......... ...A.....
D.virilis       C......... ..AA.....
D.hydei         ...T...... ..AA.A...
D.busckii       .......... ..AA.....
D.quinaria      .........T ..AA....G
D.pictiventris  ---------- ---------
Zaprionus       .........T ..AA.....
C.amoena        ...A...TG. ..AA.....
C.procnemis     .......TG. ..AA.....
D.lebanonensis  .......TG. ...A.T...
Ceratitis       ...A...TG. ...A.T...
```

Fig. 2. Continued.

Table 2. Number of nucleotide (above the diagonal) and amino acid differences (below the diagonal) between the 13 species: numbers compared are 439 nucleotides and 153 amino acids

Species	1	2	3	4	5	6	7	8	9	10	11	12	13
1 *D. melanogaster*	–	11	94	92	84	106	98	98	107	117	108	111	129
2 *D. simulans*	0	–	90	91	86	107	97	101	108	118	107	108	131
3 *D. saltans*	15	15	–	60	90	97	82	101	103	100	90	99	120
4 *D. willistoni*	19	19	10	–	86	95	82	100	97	105	96	95	120
5 *D. virilis*	20	20	16	15	–	63	68	85	96	120	96	110	132
6 *D. hydei*	19	19	18	15	6	–	79	89	96	116	108	107	130
7 *D. busckii*	18	18	17	15	10	9	–	79	84	103	94	104	115
8 *D. quinaria*	24	24	20	18	16	17	16	–	89	114	107	115	121
9 *Zaprionus*	23	23	21	17	15	12	12	19	–	107	104	113	123
10 *C. amoena*	25	25	22	21	22	22	22	26	22	–	65	108	118
11 *C. procnemis*	24	24	21	22	21	20	22	26	20	7	–	109	130
12 *D. lebanonensis*	24	24	21	20	23	24	23	23	26	24	22	–	113
13 *Ceratitis*	33	33	32	30	31	29	28	27	28	29	30	28	–

The difference in G + C content between the two sets of species becomes larger when only the third coding position is considered: the four exceptional species noted have 69–80% G + C, whereas the average is about 50% for the other species. (The G + C content is particularly low in *Ceratitis, Chymomyza,* and *Zaprionus.*)

Variation in *Sod* G + C content among dipterans has been noted earlier (Kwiatowski et al. 1992a,b). Similar heterogeneity occurs in the *Adh* gene (Starmer and Sullivan 1989), and it reflects a bias in codon preferences that is generally encountered in well-expressed genes of other taxa, ranging from bacteria to humans (Sharp et al. 1988).

Phylogenetic Analysis

Differences in the transition/transversion ratio and G + C composition are a potential problem when inferring phylogenies from sequence data. It is difficult to assess the transition/transversion bias for the *Sod* coding region

Table 3. Percent G + C content in the coding sequence of *Sod*

Species	All sites[a]		Third positions[a]	
	a	b	a	b
D. melanogaster	61	61	76	79
D. simulans	61	61	77	80
D. subobscura		60		75
D. saltans	50	50	46	45
D. willistoni	53	52	52	52
D. virilis	58	57	69	69
D. hydei	51	51	49	53
D. busckii	53	54	55	57
D. quinaria	52	52	50	50
D. pictiventris		50		48
Z. tuberculatus	50	49	44	43
C. amoena	46	46	39	39
C. procnemis	50	49	47	44
D. lebanonensis	51	51	52	54
C. capitata	48	48	42	43

[a] a, includes the complete set of 439 nucleotides sequenced in 13 species; b, subset of 345 nucleotides sequenced in all 15 species. The two initial nucleotides derived from the N primer, which are conserved in all species, have been added to the sequences for estimating G + C content

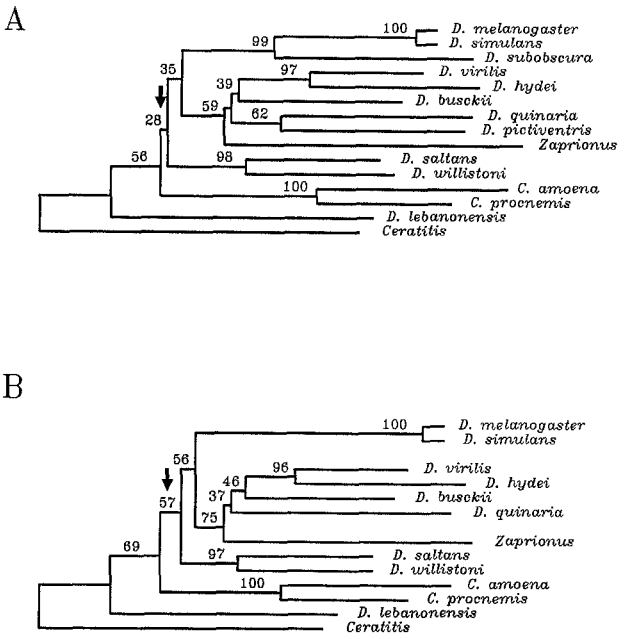

Fig. 3. Phenetic trees of *Sod* coding sequences, based on Kimura's (1980) distances (FITCH program of the PHYLIP package; Felsenstein 1989). The branch lengths correspond to the distances between sequences; the *numbers* represent the bootstrap results out of 100 replications. **A** Based on 345 bp for all 15 species. **B** Based on 439 bp for 13 species. The *vertical arrow* indicates the deletion of the second intron.

in our data since the taxa used are fairly distant in most cases. For the two most closely related species, *D. melanogaster* and *D. simulans,* the ratio is 1.2. For the more distantly related pairs, *virilis/hydei* and *willistoni/saltans,* the ratio is 1.6 and 1.9, respectively. The within-species ratio for 11 *Sod* alleles sequenced in *D. melanogaster* is 2.5 (Hudson et al. 1994). For a set of closely related species and subspecies of the *D. willistoni* group the ratio is about two (Antezana et al. unpublished data from our laboratory). Likelihood analysis (see below) suggests a transition/transversion ratio of two, and this ratio is used in calculating differences between sequences. In any case, the results of our analysis are not sensitive to the value of this ratio: a broad range (from one to 20) of transition/transversion ratios yields very similar trees.

Figure 3 shows the *Sod* phylogenetic relationships obtained by the distance method (Fitch and Margoliash 1967; FITCH algorithm in the PHYLIP package, Felsenstein 1989). The phylogeny on top (A in the figure) is based on the 343 bp sequenced in all 15 species; the one below (B) is based on the 439 bp sequenced in 13 species. Distances are calculated according to the two-parameter model of nucleotide substitutions (Kimura 1980). We have explored the degree of confidence that can be attached to the various phylogenetic relationships by means of the bootstrap method of resampling (Felsenstein 1985b). We have run SEQBOOT and DNADIST (100 replications) and the CONSENSE algorithm of PHYLIP in order to find a consensus tree.

The phylogenies show *D. lebanonensis* and the two *Chymomyza* species as early branches; the remaining clade includes all other *Drosophila* species as well as *Zaprionus.* This early separation of *D. lebanonensis* and *Chymomyza* is consistent with the presence of the second intron noted earlier, which became deleted in the short evolutionary interval that preceded the divergence of any other species in the remaining clade.

The subgenus *Sophophora* appears in Fig. 3 as polyphyletic, with the *willistoni* and *saltans* groups as the sister clade of a grouping that includes the *melanogaster* and *obscura* groups as one clade and the rest of the species as the other clade. This latter clade appears as a bush that includes the genus *Zaprionus* as well as *D. busckii, (Dorsilopha), D. pictiventris (Hirtodrosophila),* and species of the subgenus *Drosophila.* This particular clade appears as monophyletic in Fig. 3 in a relatively high number of replications (59% in A and 75% in B). The alternative grouping (which places *Hirtodrosophila* and/or *Zaprionus* outside the subgenera *Sophophora* and *Drosophila*) occurs with much lower frequencies (1% in A and 4% in B). Therefore, the distance phenetic analysis of the *Sod* sequences supports Throckmorton's rather than Grimaldi's hypothesis concerning the phylogeny of these species. However, our bootstrap values for this grouping (59 and 75% for A and B, respectively) are far from robust, since bootstrap results below 70% are suspect (Hillis and Bull 1993).

A deep split between the two *Sophophora* clades (the *willistoni* plus *saltans* and the *melanogaster* plus *obscura* groups) had been recognized by Throckmorton (1975). Rousset et al. (1991) obtained a phylogeny, based on the D1 and D2 variable domains of the large

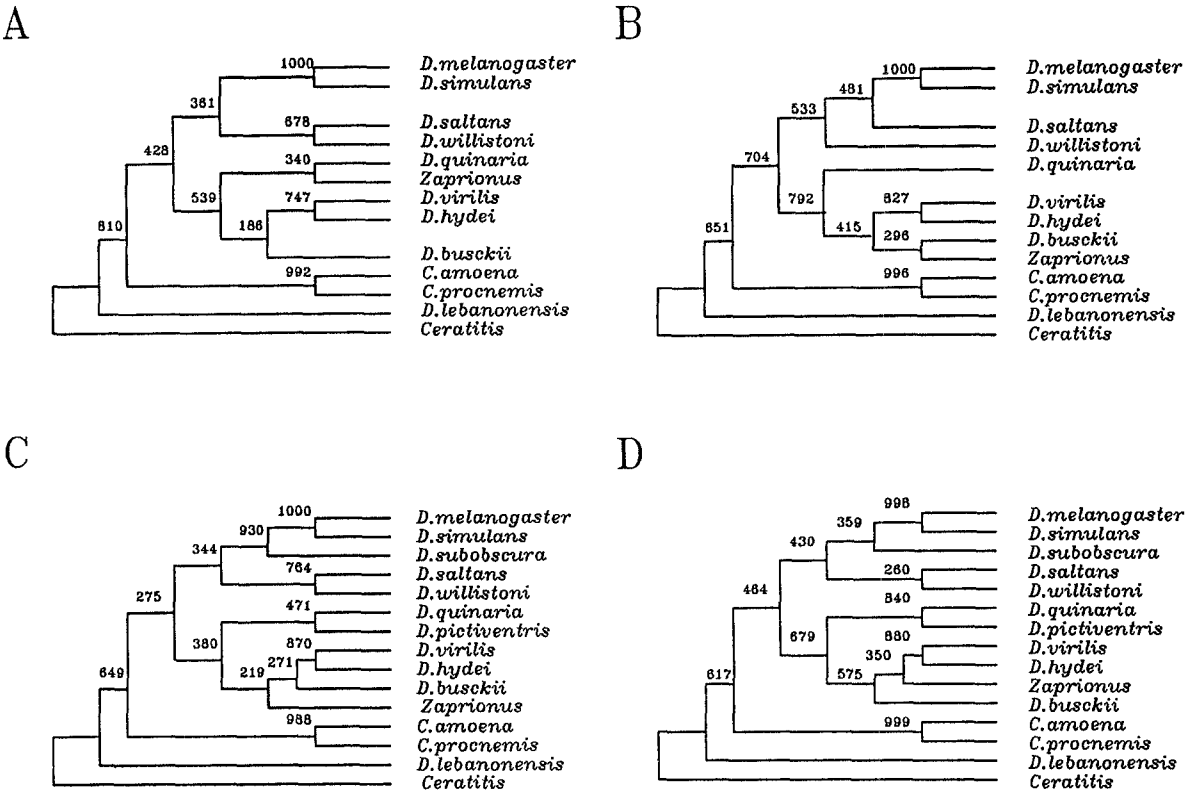

Fig. 4. Phylogenetic trees of *Sod* coding sequences, using the DNA parsimony algorithm with bootstrap resampling (DNABOOT program of PHYLIP package; Felsenstein 1989). The trees are unrooted; *Ceratitis* is used as the outgroup. The *numbers* indicate how many times the sequences to the right occurred as a clade among the 1,000 replicates. **A** and **B,** based on 439 bp (13 species); **C** and **D,** based on 345 bp (15 species); A and C, all sites; B and D, third codon positions are excluded from the analysis.

ribosomal RNA subunit, which shows, as in Fig. 3, the *willistoni* and *saltans* groups as a clade that has as a sister clade the other *Sophophora* species as well as the *Drosophila* subgenus. (See also Pelandakis et al. 1991.) However, the neighbor-joining method (Saitou and Nei 1987; result not shown), as well as other phylogenetic analyses of our data (see below), shows the *Sophophora* species as a monophyletic subgenus, although deeply divided.

A parsimony analysis (DNAPARS of PHYLIP) of the 343 nucleotides sequenced in all species, with *Ceratitis* as the outgroup, yields trees that are somewhat different, one from the other, as a function of the order in which the species are added. The six shortest trees independently obtained are three with 654 steps, one with 655, and two with 656. All trees show the same three deepest branches: successively *Scaptodrosophila* (*D. lebanonensis*), *Chymomyza,* and the subgenus *Sophophora.* The various trees differ, however, with respect to the topological position of *Zaprionus,* and the *Hirtodrosophila, Dorsilopha,* and *Drosophila* subgenera. Somewhat lesser sensitivity to the order in which the sequences are added occurs when we analyze the 439 nucleotides sequenced in 13 species. The four shortest trees all have 666 steps, but differences still occur in the topological positions of *Zaprionus, Hirtodrosophila,* and subgenus *Drosophila.*

Figure 4 gives the bootstrap results (1,000 runs) obtained with the DNABOOT algorithm of PHYLIP. The two trees on the top use the 439 bp sequenced in all 13 species; the two trees at bottom are based on the 343 bp sequenced in all 15 species. The trees on the left use all sites; the two on the right exclude the third codon bases so as to allow evaluation of the possible bias introduced by differences in G + C composition and by superimposed substitutions at particular sites. The trees in Fig. 4 are among the most parsimonious ones obtained with the DNAPARS algorithm; they also are quite similar to one another and to the distance trees shown in Fig. 3.

The bootstrap results show fairly high confidence (617 to 810 times out of 1,000 replicates) for the deepest branch, the split of *Scaptodrosophila* (*D. lebanonensis*) from the rest. The second branch, leading to the *Chymomyza* species, appears more often (704 and 464 times) in the bootstrap replications that exclude the third codon sites than when all sites are considered (428 and 275 times). This suggests that, at this depth, repeated substitutions in the third positions are obscuring the phylogenetic relationships. The clustering of the *Sophophora* subgenus shows a lower level of confidence (344–533 occurrences) than the cluster including the subgenera *Drosophila* and *Hirtodrosophila* and the genus *Zaprionus* (380–792 occurrences). This supports

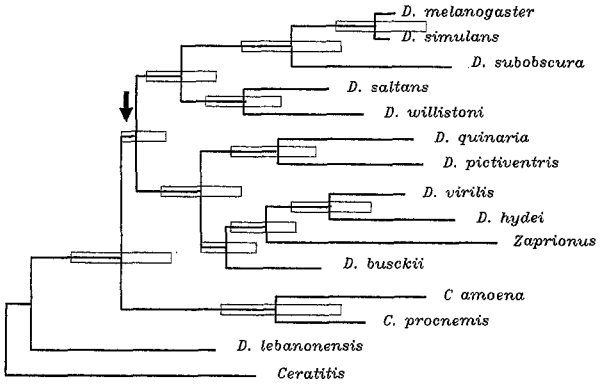

Fig. 5. Maximum likelihood tree of *Sod* coding sequences based on 345 bp. The DNAML program of the PHYLIP package is used, assuming identical rates for all sites. The maximum log likelihood was obtained with a transition/transversion ratio = 2. Multiple runs of the program, performed with randomized order of input sequences and different transition/transversion ratios seeking to maximize log likelihood, all give the same tree. Base frequencies are calculated from the data. The branch lengths are proportional to the number of nucleotide substitutions along the branches, with *hollow bars* indicating confidence intervals calculated according to Felsenstein (1989). The deletion of the second intron is shown by the *vertical arrow.*

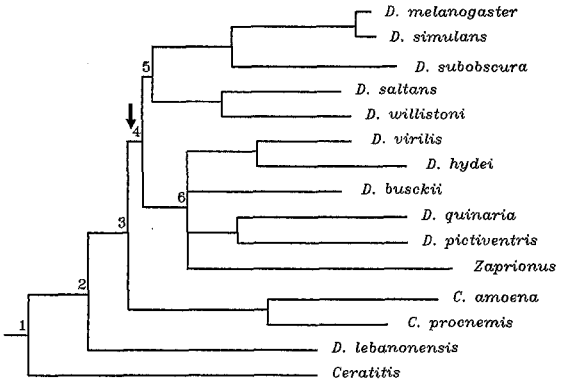

Fig. 6. Consensus phylogenetic tree of *Sod* coding sequences: an approximation to the relative temporal succession of phylogenetic events (FITCH program of the PHYLIP package with user-defined trees obtained by the DNABOOT and DNAML programs). The deletion of the second intron is represented by the *vertical arrow.*

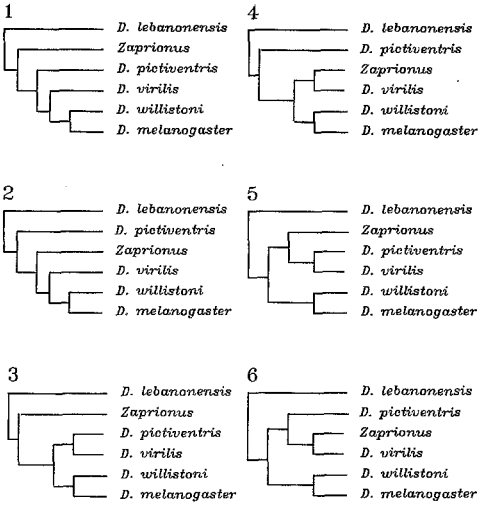

Fig. 7. Six different cladograms showing different phylogenetic hypotheses for the Drosophilidae. Tree 2 reflects the morphological (Grimaldi 1990) and mtDNA hypotheses (DeSalle and Grimaldi 1991; DeSalle 1992) whereas trees 5 and 6 are consistent with the hypotheses proposed by Throckmorton (1975), Beverley and Wilson (1984, based on larval hemolymph), and Thomas and Hunt (1993, based on the *Adh* gene), and with the *Sod* results herein presented.

the radiation of all the species in the latter cluster as a sister group to *Sophophora.* It manifests as well the deep split between the two *Sophophoran* clades: *willistoni + saltans* vs *melanogaster + obscura.* The four trees in Fig. 4 differ with respect to the branching order within the cluster that includes *Zaprionus, Hirtodrosophila,* and subgenus *Drosophila.*

Figure 5 shows a maximum likelihood phylogeny obtained with the program DNAML (Felsenstein 1981). Approximate confidence limits of branch lengths (assuming a transition/transversion ratio = 2) show that these are significantly positive ($P < 0.01$) in all cases except for the one between *Chymomyza* and the large clade that includes most *Drosophila* species, but this split is confirmed by the deletion of the second intron in all species of the larger clade. The two deepest branches (*D. lebanonensis* and *Chymomyza*) are the same, and in the same order, as in all previous trees. Also, the *Sophophora* species appear as a cluster, distinct from the cluster that includes the subgenus *Drosophila, Hirtodrosophila,* and *Zaprionus.* The relationships just mentioned also persist in the phylogeny obtained with the FITCH program for distance data, provided with the trees obtained with DNABOOT and DNAML (Fig. 6).

We have tested alternative hypotheses of the Drosophilidae phylogeny by statistical evaluation of pertinent trees. For simplicity we have pruned our trees so that they only include six species of representative taxa: *Scaptodrosophila* (*D. lebanonensis*), *Zaprionus* (*Z. tuberculatus*), *Hirtodrosophila* (*D. pictiventris*), subgenus *Drosophila* (*D. virilis*), and *Sophophora* (*D. willistoni* and *D. melanogaster*). The trees are shown in Fig. 7. Tree 2 is advanced by Grimaldi (1990) based on morphology and DeSalle (1992) based on mtDNA, while trees 5 and 6 represent Throckmorton's (1975) view and the *Sod* results. We have used the methods of Kishino and Hasegawa (1989) and Templeton (1983; Felsenstein 1985a) to calculate the mean and variance of log likelihood, and step differences between trees, respectively. According to these tests, none of the six trees is significantly worse than the best one when maximum likelihood and maximum parsimony analyses are performed on all 343 sites. However, when only the first and second codon positions are considered for maximum parsimony analysis, trees 1, 2, and 3 are significantly worse than tree 6, which is the best (Table 4). Similar results are obtained with the full set of 13 sequences. Here again, the two trees (similar to 1 and 2) with *Hirtodrosophila* and *Zaprionus* branches outside

Table 4. Statistical comparison of six possible different tree topologies for six Drosophilidea species: the tests are based on the first and second codon positions for 343 bp of the *Sod* coding region and follow the method of Templeton (1983)

		Comparison with tree No. 6		
Tree[a]	Steps	Difference	SE	Significantly worse?
1	92.0	8.0	2.8	Yes
2	90.0	6.0	2.8	Yes
3	91.0	7.0	3.3	Yes
4	88.0	4.0	2.4	No
5	85.0	1.0	2.2	No
6	84.0			

[a] The six topologies are shown in Fig. 7

Table 5. Mean number of PAM differences and estimated time since the phylogenetic events indicated

Node[a]	Phylogenetic event	PAM[b]	Million years
1	Tephritidae/Drosophilidae	23.0 ± 0.6	77
2	*Scaptodrosophila* divergence	16.9 ± 0.4	56
3	*Chymomyza/Drosophila* genus	16.6 ± 0.3	55
4	*Drosophila* genus radiation	13.3 ± 0.6	44
5	*Sophophora* radiation	12.0 ± 1.6	40
6	*Drosophila* subgenus radiation[c]	10.0 ± 0.8	33

[a] The nodes are as numbered in Fig. 6

[b] PAM is the estimated percent of amino acid differences corrected for superimposed and back replacements

[c] *Zaprionus, Dorsilopha,* and *Hirtodrosophila* are included in the *Drosophila* subgenus radiation

the *Drosophila* clade (subgenera *Drosophila* and *Sophophora*) are rejected relative to the best tree (similar to 6), which has the *Hirtodrosophila* and *Zaprionus* branches inside, as part of a monophyletic cluster that includes the subgenus *Drosophila.* Similar results with respect to the position of *Zaprionus* have been recently obtained with *Adh* sequences (Thomas and Hunt 1993). A molecular study based on immunological distances of larval hemolymph protein also places *Hirtodrosophila* within a clade that includes the subgenus *Drosophila,* but not *Sophophora* (Beverley and Wilson 1984).

Divergence Time

Table 5 shows time estimates for various phylogenetic events. The relevant nodes are labeled in Fig. 6. The estimates are based on the number of amino acid replacements given in Table 2. We use amino acid rather than nucleotide differences, because the former are likely to be more reliable in this case for two reasons. One is that the substantial differences observed in G + C composition among species, particularly in the third codon sites, introduce a bias that is difficult to evaluate as to its magnitude and significance (Gillespie 1986; Woese 1991). The other reason is that the rate of amino acid substitutions in the Cu,Zn SOD of diverse organisms has been shown to be fairly constant during the last 60 MY (million years); the rate is approximately 15 aa/100 aa/100 MY for PAM-corrected data (Kwiatowski et al. 1991a, 1992a; for the PAM correction, see Dayhoff 1978).

The evolutionary time estimates given in Table 5 are only rough estimates, because they not only depend on the assumption of a molecular clock, but also on the particular rate previously established for SOD, and on the limited amount of information provided by a dozen or so *Sod* sequences. Nevertheless, the estimates shown in Table 5 may be as reliable as any currently available in the literature, since sequence data are more precise than data sets based on immunological distances, two-dimensional protein electrophoresis, and restriction analysis.

Our time estimates are somewhat lower than those of Collier and MacIntyre (1977) based on microcomplement fixation studies of alpha-glycerophosphate dehydrogenase, and those of Spicer (1988), based on two-dimensional protein electrophoresis. Collier and MacIntyre (1977) estimate the Tephritidae radiation at 90 MY (our estimate, 77 MY). Spicer (1988) estimates the *Drosophila* genus radiation at about 60 MY (ours, 44). The radiation of the *Drosophila* subgenus is estimated by Collier and MacIntyre (1977) as well as by Spicer (1988) at 50 MY (ours, 33 MY). The estimates of Beverley and Wilson (1984), based on immunological distances for a larval hemolymph protein, also are somewhat higher than those shown in Table 5. However, lower numbers than ours are estimated by Thomas and Hunt (1993) based on the nucleotide sequence of the *Adh* gene: *Scaptodrosophila* divergence, approximately 45 MY (ours, 56 MY); *Drosophila* genus radiation, 40 MY (ours, 44); *Drosophila* subgenus radiation, 27 MY (ours, 33 MY).

Discussion

The Cu/Zn superoxide dismutase is an abundant enzyme in eukaryotic organisms, with highly specific superoxide dismutation activity that protects aerobic cells against the harmfulness of free oxygen radicals (Fridovich 1986). Cu,Zn SOD is distinctly interesting for investigating phylogenetic issues, because (1) it is apparently present in all eukaryotes, and (2) it evolves at a fairly high rate, so as to be informative for recent evolutionary events, i.e., within the last 100 MY (Lee et al. 1985; Ayala 1986; see Table 2). Yet (3), it is well conserved over long time spans, so 60% of the amino acid residues remain identical between organisms from different kingdoms, such as humans and yeasts (Ayala 1986; Kwiatowski et al. 1991a). In higher Diptera the *Sod* gene consists of a 462-bp coding region interrupt-

ed by one or two introns (Kwiatowski et al. 1992a; see Fig. 1).

The Drosophilidae are well-studied organisms with respect to genetics and systematics. Yet many issues remain controversial. The commonly accepted taxonomy (Wheeler 1981, 1986) and the evolutionary account of Throckmorton (1975) have been recently challenged by Grimaldi (1990) in important respects. A profusion of molecular investigations have failed to settle the issues and have often yielded incongruous outcomes. The results conveyed in the present paper provide helpful evidence toward resolving some issues.

On the basis of biogeographical, morphological, and other considerations, Throckmorton (1975) has argued that the divergence of *Scaptodrosophila* (represented in our study by *D. lebanonensis*) precedes the first major radiation of the genus *Drosophila.* This is supported by all our data. The presence of the second *Sod* intron in *Scaptodrosophila* (as well as in *Ceratitis* and *Chymomyza*), and its absence in the species of the *Sophophora* and *Drosophila* subgenera, definitely places the phylogenetic divergence of *Scaptodrosophila* before the *Drosophila* radiations (a position also endorsed by Grimaldi 1990; and DeSalle and Grimaldi 1991; but see Villarroya and Juan 1991). Grimaldi (1990) has accordingly raised *Scaptodrosophila* to the taxonomic status of "genus."

The presence of the second *Sod* intron in the two *Chymomyza* species investigated, and its absence from the species of the subgenus *Sophophora, Drosophila, Hirtodrosophila,* and *Dorsilopha* (as well as from the genus *Zaprionus*), places the *Chymomyza* lineage outside the *Drosophila* radiations. This is also supported by the analysis of the *Sod* sequence data. (See Figs. 3–7.) Our results therefore contradict Throckmorton's inclusion of *Chymomyza* as a member of the *Sophophora* radiation and support the phylogenetic position of *Chymomyza* proposed by Grimaldi (1990, his Fig. 542, p. 100) and DeSalle and Grimaldi (1991).

The *Sod* sequence data indicate that *Scaptodrosophila* and *Chymomyza* diverged from the *Drosophila* lineage within a short time interval (between 56 and 55 MY ago, according to our date estimates). The data are thus insufficient to decide which one of the two lineages is the sister clade of the *Drosophila* clade. The radiation of the genus *Drosophila* happened shortly afterward (estimated at 44 MY in Table 4); but it was during this brief time preceding the *Drosophila* radiation that the *Drosophila* lineage lost the second *Sod* intron.

The absence of the second *Sod* intron from *D. pictiventris* (subgenus *Hirtodrosophila*) excludes the position of *Hirtodrosophila* outside the *Chymomyza* + *Drosophila* clade, as proposed by DeSalle (1992) on the basis of mtDNA data. (See also DeSalle and Grimaldi 1991.) The *Sod* sequence data support Throckmorton's (1975) position of *Hirtodrosophila* within the radiation of the subgenus *Drosophila* (*sensu lato,* i.e., inclusive also of the genus *Zaprionus* and the subgenus Dorsilopha) and are inconsistent with Grimaldi's (1990) opposite conclusion (as well as DeSalle's).

The *Sod* sequence data also support Throckmorton (1975) on the phylogenetic position of the genus *Zaprionus,* which he sees as part of the subgenus *Drosophila* (*s.l.*) radiation. The *Sod* phylogenies contradict Grimaldi (1990), DeSalle and Grimaldi (1991), and DeSalle (1992), who place *Zaprionus* outside the clade comprising the subgenera *Drosophila* and *Sophophora.* Throckmorton's (1975) proposal that the *Sophophora* radiation preceded the radiation of the subgenus *Drosophila* is also supported by the *Sod* sequence data.

Several systematists, Throckmorton (1975) among them, have noted that the evolution of the drosophilids is modulated by rapid radiations, or bursts of cladistic expansion. The short time spans between cladistic events that follow one another in rapid succession are unlikely to leave conspicuous traces in the organisms' morphology or genetic makeup. The sequence of phylogenetic events may then be difficult to determine. This hardship is further intensified in the evolution of the drosophilids by the relative scarcity of fossil specimens, substantial conservation of morphology, and occasional homoplasy. It is thus not surprising that the systematics of the Drosophilidae has remained controversial in the face of extensive and authoritative investigations.

Will molecular information eventually provide the definitive answers concerning phylogenetic matters? Nucleotide sequence data have indeed the potential to do so. The DNA of an organism has a record of its evolutionary history. There are many genes (and other DNA sequences) in each organism, so more and more data can in principle be accumulated until a particular phylogenetic issue of interest is settled. But possibility in principle and securing the data are very different matters. Obtaining a DNA sequence is a laborious process (compare it with a morphological observation such as eye color or wing length), so at best only a few relevant DNA sequences (or other highly informative molecular data) are known for most groups of organisms. Whenever the molecular data are very limited, as currently they are in most cases of interest, variance in evolutionary rates, homoplasy, and other difficulties can yield erroneous conclusions when taken at face value. (Homoplasy is a particularly nagging problem, since the nucleotide bases provide only four possible alternatives at each particular site in a DNA sequence.)

The present investigation illustrates some of the virtues and potential pitfalls of molecular data. The pitfalls are apparent in the variation of outcomes concerning details of the phylogeny obtained by different analytical methodologies: compare Figs. 3–6. An attempt to attenuate the difficulties of homoplasy, by ignoring the third codon positions, increases only slightly the stability of branches in the phylogeny. (Compare

B and D with A and C in Fig. 4.) The contributing reasons are three: (1) the twofold degeneracy of the genetic code implies that silent sites occur in first codon positions, not only in the third positions; (2) the evolution of codon preferences, reflected in G + C content in the third positions, is itself phylogenetically informative; (3) the data set becomes reduced when one-third of the nucleotides are excluded from consideration. The potential pitfalls of molecular data are illustrated de facto by the observation that mtDNA sequence data (DeSalle and Grimaldi 1991; DeSalle 1992) yield conclusions that are inconsistent with the *Sod* sequence data.

One virtue of DNA sequences that is illustrated by the *Sod* data is that phylogenetic information derives not only from the direct comparison of nucleotides at particular sites in a sequence, but also from the organization of the DNA sequences. Deletions of well-defined DNA segments may be particularly informative, as exemplified by the deletion of the second *Sod* intron, owing to the low probability of independent occurrence or restoration of such an event within a defined phylogeny; that is, homoplasy is particularly unlikely.

There can be little doubt that the accumulation of DNA sequence data may eventually settle any given phylogenetic issue. It would seem equally certain that in the interim, or for that matter at any time, the only reasonable approach to settling phylogenetic relationships is to use all available information—molecular, morphological, biogeographical, etc.—and to weigh it according to its value in a particular case.

Acknowledgments. Research supported by NIH grant GM42397. We thank R.R. Hudson for giving us access to his computer hardware; Marcos Antezana and Leila Fakhrabadi for help with DNA sequencing; J. Felsenstein for making available the PHYLIP package; and W. Heed (University of Arizona) for morphological corroboration of the species *D. pictiventris*.

References

Ayala FJ (1986) On the virtues and pitfalls of the molecular evolutionary clock. J Hered 77:226–235

Beverley SM, Wilson AC (1984) Molecular evolution in *Drosophila* and higher Diptera. II. Time scale for fly evolution. J Mol Evol 21:1–13

Collier GE, MacIntyre RJ (1977) Microcomplement fixation studies on the evolution of α-glycerophosphate dehydrogenase within the genus *Drosophila*. Proc Natl Acad Sci USA 74:684–688

Dayhoff MD (1978) Atlas of protein sequence and structure. Natl Biomed Res Found, Washington, DC

DeSalle R (1992) The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol Phylogenet Evol 1:31–40

DeSalle R, Grimaldi DA (1991) Morphological and molecular systematics of the Drosophilidae. Annu Rev Ecol Syst 22:447–475

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–374

Felsenstein J (1985a) Confidence limits on phylogenies with a molecular clock. Syst Zool 34:152–161

Felsenstein J (1985b) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Felsenstein J (1989) PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–287

Fridovich I (1986) Superoxide dismutases. Adv Enzymol 58:61–97

Gillespie JH (1986) Rates of molecular evolution. Annu Rev Ecol Syst 17:637–655

Grimaldi D (1990) A phylogenetic revised classification of genera in the Drosophilidae. Bull Am Museum Nat Hist 197:1–139

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182–192

Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ (1994) Evidence for positive selection at the superoxide dismutase (*Sod*) locus of *Drosophila melanogaster*. Genetics (in press)

Kawasaki ES (1990) Sample preparation from blood, cells, and other fluids. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 146–152

Kimura M (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179

Kwiatowski J, Ayala FJ (1989) *Drosophila virilis* Cu,Zn superoxide dismutase gene sequence. Nucleic Acids Res 17:2133

Kwiatowski J, Gonzalez F, Ayala FJ (1989a) *Drosophila simulans* Cu,Zn superoxide dismutase gene sequence. Nucleic Acids Res 17:6735

Kwiatowski J, Patel M, Ayala FJ (1989b) *Drosophila melanogaster* Cu,Zn superoxide dismutase gene sequence. Nucleic Acids Res 17:1264

Kwiatowski J, Hudson RR, Ayala FJ (1991a) The rate of Cu,Zn superoxide dismutase evolution. Free Radic Res Commun 12–13: 363–370

Kwiatowski J, Skarecky D, Hernandez S, Pham D, Quijas F, Ayala FJ (1991b) High fidelity of the polymerase chain reaction. Mol Biol Evol 8:884–887

Kwiatowski J, Skarecky D, Ayala FJ (1992a) Structure and sequence of the *Cu,Zn Sod* gene in the Mediterranean fruit-fly, *Ceratitis capitata:* intron insertion/deletion and evolution of the gene. Mol Phyl Evol 1:72–82

Kwiatowski J, Skarecky D, Burgos M, Ayala FJ (1992b) Structure and sequence of the Cu,Zn superoxide dismutase gene of *Chymomyza amoena:* phylogeny of the genus and codon-use evolution. Insect Mol Biol 1:3–13

Lee YM, Friedman DJ, Ayala FJ (1985) Superoxide dismutase: an evolutionary puzzle. Proc Natl Acad Sci USA 82:824–828

Lloyd AT, Sharp PM (1992) CODONS: a microcomputer program for codon usage analysis. J Hered 83:239–240

Patterson JT, Stone WS (1952) Evolution in the genus *Drosophila*. MacMillan, New York

Pelandakis M, Higgins DH, Solignac M (1991) Molecular phylogeny of the subgenus Sophophora of *Drosophila* derived from large subunit of ribosomal RNA sequences. Genetica 84:87–94

Rousset FM, Pelandakis M, Solignac M (1991) Evolution of compensatory substitutions through G-U intermediate state in *Drosophila* rRNA. Proc Natl Acad Sci USA 88:10032–10036

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F (1988) Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schistosaccharomyces pombe, Dro-*

sophila melanogaster and *Homo sapiens:* a review of the considerable within-species diversity. Nucleic Acids Res 16:8207–8211

Spicer GS (1988) Molecular evolution among some *Drosophila* species groups as indicated by two-dimensional electrophoresis. J Mol Evol 27:250–260

Starmer WT, Sullivan DT (1989) A shift in the third-codon-position nucleotide frequency in alcohol dehydrogenase genes in the genus *Drosophila.* Mol Biol Evol 6:546–552

Sturtevant AH (1921) The North American species of Drosophila. Carnegie Institution, Washington, DC

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. Evolution 37:221–244

Thomas RH, Hunt JA (1993) Phylogenetic relationships in Drosophila: a conflict between molecular and morphological data. Mol Biol Evol 10:362–374

Throckmorton LH (1975) The phylogeny, ecology and geography of Drosophila. In: King RC (ed) Handbook of genetics. Plenum Press, New York, pp 421–469

Villarroya A, Juan E (1991) ADH and phylogenetic relationships of *Drosophila lebanonensis* (Scaptodrosophila). J Mol Evol 32:421–428

Wheeler MR (1981) The Drosophilidae: a taxonomic overview. In: Ashburner M, Carson HL, Thompson JN Jr (eds) The genetics and biology of drosophila. Academic Press, New York, pp 1–97

Wheeler MR (1986) Additions to the Catalog of the World's Drosophilidae. In: Ashburner M, Carson HL, Thompson JN Jr (eds) The genetics and biology of drosophila, vol 3e. Academic Press, New York, pp 395–409

Woese CR (1991) The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria. In: Selander RK, Clark AG, Whittam TS (eds) Evolution at the molecular level. Sinauer Associates, Sunderland, MA, pp 1–24