

## Functional Constraints of the Cu,Zn Superoxide Dismutase in Species of the *Drosophila melanogaster* Subgroup and Phylogenetic Analysis

Kyriaki Arhontaki,<sup>1</sup> Elias Eliopoulos,<sup>1</sup> George Goulielmos,<sup>1</sup> Petros Kastanis,<sup>1</sup> Spyros Tsakas,<sup>1</sup> Michael Loukas,<sup>1</sup> Francisco Ayala<sup>2</sup>

<sup>1</sup> Department of Genetics, Agricultural University of Athens, Iera Odos 75, Athens 11855, Greece

<sup>2</sup> Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA

Received: 13 November 2001 / Accepted: 3 July 2002

**Abstract.** The phylogenetic relationships among the *Drosophila melanogaster* subgroup species were analyzed using approximately 1550-nucleotide-long sequences of the Cu,Zn SOD gene. Phylogenetic analysis was performed using separately the whole region and the intron sequences of the gene. The resulting phylogenetic trees reveal virtually the same topology, separating the species into distinct clusters. The inferred topology generally agrees with previously proposed classifications based on morphological and molecular data. The amino acid sequences of the Cu,Zn SOD of the *D. melanogaster* subgroup species reveal a high-conservation pattern. Only 3.9% of the total amino acid sites are variable, and none affects the major structural elements. Comparison of the *Drosophila* Cu,Zn SOD amino acid sequences with the Cu,Zn SOD of *Bos taurus* and *Xenopus laevis* (whose three-dimensional structure has been elucidated) reveals conservation of all the protein's functionally important amino acids and no substitutions that dramatically change the charge or the polarity of the amino acids.

**Key words:** Cu,Zn SOD — *Drosophila melanogaster* subgroup — Phylogeny — Amino acid sequence — 3D structure

### Introduction

Superoxide dismutase (SOD) is a metalloenzyme which protects organisms from toxic reactive oxygen metabolites by catalyzing the conversion of the superoxide anion to hydrogen peroxide and molecular oxygen (McCord and Fridovich 1969). Oxygen free radicals (superoxide O<sup>2-</sup>, hydrogen peroxide H<sub>2</sub>O<sub>2</sub>, and hydroxyl anion OH) are generated as intermediates during normal metabolism (Halliwell and Gutteridge 1984), and even more so when organisms are exposed to such environmental conditions as active compounds (ozone, SO<sub>2</sub>, herbicides, acid rain, cigarette smoke, etc.), radiation, and other stress agents. Increased levels of oxygen radicals (oxidative stress) can damage cells by oxidizing many biomolecules (DNA, proteins, membrane lipids), which result in several diseases, such as mutation, carcinogenesis, aging, increased male crossing-over frequencies in *Drosophila melanogaster* (Vontas et al. 1999), and cell death (Fleming et al. 1982; Harman 1982).

SOD is classified into Mn, Fe, and Cu,Zn SOD, depending on the metal in the active site. Cu,Zn SODs are found in the cytosol of all eukaryotic cells (Puget and Michelson 1974) but not generally in prokaryotic cells, with a few exceptions (e.g., *Photobacter leiognathi*, *Pseudomonas diminuta*, *P. maltophilia*, *Caulobacter crescentous*, and several species of *Brusella* and *Haemophilous*). However, according to Benov and Fridovich (1994), who found that *Esc-*

*herichia coli* contains a kind of monomeric variant copper and zinc SOD in the periplasm, it is likely that this enzyme is a common attribute of Gram-negative bacteria. Furthermore, the Mn SODs occur in bacteria, chloroplasts, eukaryotic algae, and mitochondria; the Fe enzymes occur in bacteria, chloroplasts, and prokaryotic algae (Fridovich 1979; Asada et al. 1980).

Eukaryotic Cu,Zn SODs consist of two identical subunits, each of which contains one atom of copper and one atom of zinc in the active site of the enzyme, with a total molecular weight of 31.5 kDa. In *Drosophila melanogaster* the Cu,Zn SOD locus is on the left arm of chromosome 3 and encodes a polypeptide with 151 amino acids in the functional enzyme.

In addition to the essential physiological role of SOD and the high enzymatic activity of the protein, *Sod* is useful in phylogenetic studies. Cu,Zn *Sod* is a single-copy gene and the protein is abundant in almost all eukaryotes and fairly conserved in amino acid sequence (Ayala 1986). In *Drosophila* it evolves at a fairly rapid rate that makes it informative for research of recent evolutionary events (Lee et al. 1985).

The family *Drosophilidae* is one of the most diverse and widely distributed dipteran families, consisting of more than 3000 species, which are divided into 61 genera (Wheeler 1986). Its taxonomy and phylogeny, however, remain controversial, due to the lack of knowledge of the phylogenetic relationships of the species. To this end, several studies have been done based either on internal morphology and biogeography (Throckmorton 1975) and morphological characters (Grimaldi 1990) or on molecular approaches including the microcomplement fixation technique (Beverly and Wilson 1984), DNA-DNA hybridization, mitochondrial DNA (mtDNA), ribosomal RNA (rRNA), and sequences of various nuclear genes.

The *melanogaster* subgroup is one of the most intensively studied groups in the genus *Drosophila*. However, questions remain about the phylogenetic relationship among *D. simulans*, *D. melanogaster*, *D. mauritiana*, and *D. sechellia*. According to Solignac and co-workers' (Solignac et al. 1986; Solignac and Monnerot 1986) data (RLFP analysis of the mtDNA), the subgroup is polyphyletic, with *D. mauritiana*, *D. simulans*, and *D. sechellia* associating with *D. melanogaster*. *D. yakuba* and *D. teissieri* are placed outside of this cluster, while the species pair *D. erecta*/*D. orena*, which is also placed outside the *D. melanogaster* cluster, has appeared more recently. Caccone et al. (1988), based on scnDNA-DNA hybridization, found that *D. mauritiana* and *D. sechellia* are the most closely related species, followed by *D. simulans* and *D. melanogaster*. *D. yakuba*, *D. teissieri*, and *D. erecta* represent another cluster,

while *D. orena* is outside all the former groups. In contrast, many other studies proposed the species pairs *simulans*/*sechellia* (Cariou 1987) and *simulans*/*mauritiana* (Lachaise et al. 1986; Kliman et al. 2000; Ting et al. 2000) as the most closely related ones. The phylogenetic tree topography was a little different when the same method was applied in mtDNA. The latter was verified by Lachaise et al. (1988), who prepared a consensus tree using molecular, chromosomal, morphological, and behavioral data, in which *D. mauritiana* and *D. simulans* are most closely related, followed by *D. sechellia* and *D. melanogaster*. *D. teissieri*-*D. yakuba* and *D. erecta*-*D. orena* are sister taxa and they cluster with the rest of the *melanogaster* complex separately.

Considering the circumstances mentioned above, a reexamination of the phylogenetic relationships of the species of the *D. melanogaster* subgroup is expected to provide valuable information. To assess these relationships we have determined the nucleotide sequences of the SOD coding region in six species belonging to the *melanogaster* subgroup. Together with homologous sequences published previously, we used the information obtained from the analysis of these nucleotide sequences in an attempt to infer the phylogeny of the *melanogaster* subgroup. We used the Cu,Zn SOD nucleotide sequence also seeking to investigate the amino acid homology of the protein among the *Drosophila* species and any inferred substitution impacting the three-dimensional (3D) structure. The functional constraints of the enzyme, reflected by the very low amino acid variation between the *Drosophila* species, are extensively discussed.

## Materials and Methods

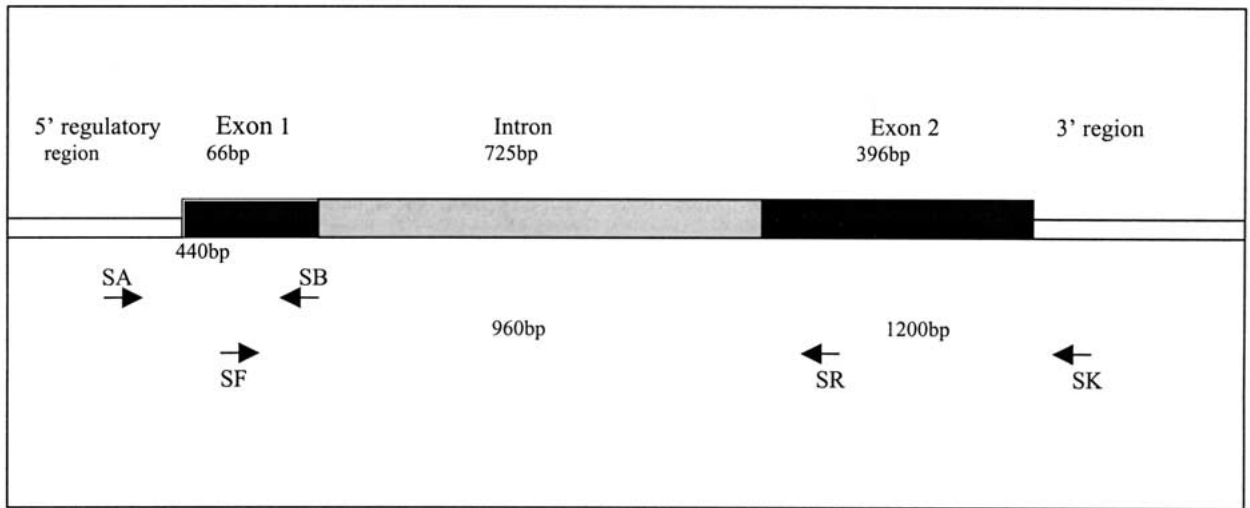
*Drosophila Species.* Nine species of the *Drosophila melanogaster* and *montium* subgroup were used: *D. sechellia* (stock number 14021-0248.1), *D. mauritiana* (stock number 14021-0241.5), *D. yakuba* (stock number 14021-0261.0), *D. orena* (stock number 14021-0245.0), *D. teissieri* (stock number 14021-0251.0), *D. erecta* (stock number 14021-0224.0), *D. isakasi* (stock number 14028-0681.4), *D. serrata* (stock number 14028-0681.4), and *D. kikkawai* (stock number 14028-0561.5), obtained from the National *Drosophila* Species Resource Center at Bowling Green. The nucleotide sequences of *D. melanogaster* (Canton S), *D. simulans*, and *D. willistoni* were recovered from the NCBI Genbank database with accession numbers X17332, X15685, and X13831, respectively.

*DNA Preparation and Amplification.* Insects were frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Genomic DNA was extracted from about 20–30 flies on ice following the method of Henry et al. (1990). Two pairs of primers were designed for the PCR reaction method (McPherson et al. 1991) based on the conserved regions of the *D. melanogaster* Cu,Zn SOD nucleotide sequence.

First pair:

SA, 5'-d[AGTTCCCGCCACTGTCATTG]-3';

SB, 5'-d[GGATTCTCACCTCCTGTTTCG]-3'.



**Fig. 1.** Structure of the Cu,Zn Sod gene of *Drosophila melanogaster* and strategy for amplification, cloning, and sequencing. Arrows refer to the location and direction of the amplification and sequencing primers.

Second pair:

SF, 5'-d[AATTAACGGCGATGCCAAGG]-3';

SR, 5'-d[CCTCAATGTTGCCAGATCG]-3'; or

SK, 5'-d[GAGTAGATCAGCACAGTGGC]-3'.

The amplification was carried out using high-fidelity conditions (Kwiatowski et al. 1991b). Pwo polymerase (a proofreading enzyme) (Boehringer–Mannheim) was used to get amplification products of a high fidelity. The total amplification product consists of approximately 1670 bp including the two exons of the Cu,Zn SOD gene, the intron, and a segment of about 350 bp of the 5' untranslated region (Fig. 1).

**DNA Cloning and Sequencing.** The amplified samples were run on 1% agarose gel and the bands excised out using the Gel Extraction Kit (Boehringer–Mannheim, Catalog No. 1696505). Samples were then filled in with Klenow polymerase and ligated in the vector pBluescript II SK(-) in the *EcoRV* restriction site (Stratagene) and transformed into competent XL1 blue cells in *Escherichia coli* (Sambrook et al. 1989). Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Catalog No. 27104). Sequencing of the double-stranded plasmid was carried out using an ABI PRISM 310 Genetic Analyzer (Perkin–Elmer). The method is based on the Sanger et al. (1977) dideoxy termination method. For the amplification reaction we used the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with Ampli Taq DNA Polymerase FS and four dideoxy terminators labeled with different fluorescent dyes (Perkin–Elmer). The PCR reaction products were electrophoresed in the ABI PRISM 310 Genetic Analyzer instrument. The PCR products were separated according to size under a constant voltage at the edge of a capillary filled with POP6 polymer. Detection of the dye-probed products was achieved with a laser detector and a lens system. For sequencing, we used the same PCR primers, SA, SB, SF, SR, SK, and the external -21 M13 primer. For each genomic region both strands were completely sequenced and the consensus nucleotide sequence was obtained for two different clones from at least two sequencing reactions.

**DNA Sequence Analysis and Phylogenetic Tree Construction.** The sequence alignment was performed using the program Clustal W (Tompson et al. 1994). The phylogenetic trees were constructed with the neighbor-joining (NJ) (Saitou and Nei 1987), UPGMA (Sneath and Sokal 1973), maximum parsimony (MP) (Fitch 1971),

and minimum evolution (ME) (Rzhetsky and Nei 1992) methods. All phylogenetic and molecular evolutionary analyses were conducted using MEGA Version 2.0 (Kumar et al. 2001) and PHYLIP software package Version 3.5c (Felsenstein 1995). The confidence level of the clusters was evaluated with the bootstrap test with 1000 replications (Felsenstein 1985) and with the *t* test (Rzhetsky and Nei 1992, 1993). The nucleotide sequences reported in this study have been deposited in the NCBI/Genbank nucleotide sequence database under accession numbers AF127155 to AF127160.

## Results

### Structure of the Sod Gene and Nucleotide Sequences

The structure of the *Sod* gene is outlined in Fig. 1. We determined the nucleotide sequence of the Cu,Zn SOD coding region for six species of the *Drosophila melanogaster* subgroup and a segment of the 5' SOD flanking region for nine species (Fig. 2). In addition, for the sequence alignments we used two published sequences: *Drosophila melanogaster* (Kwiatowski et al. 1989a) and *Drosophila simulans* (Kwiatowski et al. 1989b). Cu,Zn SOD consists of two exons and one intron. The first exon is 66 bp long and codes for the first 22 amino acids including the N-terminal methionine. The second is 396 bp and codes for the other 131 amino acids plus the TAA termination signal. The intron ranges in length from 708 to 783 bp. The intron–exon insertion points exhibit the GT–AG consensus sequence (Breathnach et al. 1978). The sequence before the first methionine is CGAA (Fig. 2; boldface characters) in all species, including *D. melanogaster* and *D. simulans*, instead of C/A AA A/C as given by Cavener (1987) for *Drosophila* genes.

The 5' flanking region is lowly conserved. The variation of the segment that is transcribed but not translated is 15.8% (Fig. 2; light-shaded sequences); it

<i>D. kikkawai</i>	TGGAAAGCTTCCAGGTGAACCACTGGTTCGTTTAAACAGCTCAAAAAAGTGAACGTCATTT	60
<i>D. serrata</i>	TGGAAAGCTTCCAGGTGAACCACTGGTTCGTTTAAACAGCTCAAAAAAGTGAACGTCATTT	60
<i>D. mauritiana</i>	TGGAAAGCTTCCACGTGAACCACTGGTTCGTTTAAACAGCTCAAAAAAGTGAACGTCATTT	60
<i>D. erecta</i>	TGGAAAGCTTCCACGTGAACCACTGGTTCGTTTAAACAGCTCAAAAAAGTGAACGTCATTT	60
<i>D. tsacasi</i>	TGGAAAGCTTCCAAAGTGAACCACTGGTTCGTTTAAACAGCTCAAAAAAGTGAACGTCATTT	60
<i>D. simulans</i>	TGGAAAGCTTCCAAAGTGAACCACTGGTTCGTTTAAACAGCTCAAAAAAGTGACGTAATTT	60
<i>D. orena</i>	TGGAAAGCTTCCACGTGAACCACTAGTTCGTTGACAAGTTAAAAAGTGAACGCCATTT	60
<i>D. sechellia</i>	TGGAAAGCTTCCACGTGAACCACTAGTTCGTTGACAAGTTAAAAAGTGAACGCCATTT	60
<i>D. melanogaster</i>	TGGAAAGCTTCCAAAGTGAACCACTGGTTCGTTGACAAGTTAAAAAGTGAACGCCATTT	60
<i>D. teissieri</i>	TGGAAAGCTTCCAAAGTGAACCACTGGTTCGTTGACAAGTTAAAAAGTGAACGCCATTT	60
<i>D. yakuba</i>	TGGAAAGCTTCCGGATGAACCACTGGTTCATTGAACAGCTCAAAAAATTGAACCCGCTT	60
	*****	
<i>D. kikkawai</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTCAAATTTTG	96
<i>D. serrata</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTCAAATTTTG	96
<i>D. mauritiana</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTCAAATTTTG	96
<i>D. erecta</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTCAAATTTTG	96
<i>D. tsacasi</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTCAAATTTTG	96
<i>D. simulans</i>	GTGCAGTGAC-----TACAAATGTTCCGTTTTAAAATTTTG	96
<i>D. orena</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTAAAATTTTG	96
<i>D. sechellia</i>	GTGCAGTAAC-----TACAAATGTTCCGTTTTAAAATTTTG	96
<i>D. melanogaster</i>	GTGCAGTAA-----AATTGTTCCGTTTTCAAATTTTG	92
<i>D. teissieri</i>	GGGCTATAA-----ACACAATTTACGAAATTTGTAATTAAGAGCTTTA	104
<i>D. yakuba</i>	GGGCTATAACCACAGTGGTCTGTACACAATTTTATAATATAGTTTTAGGAAGCACTA	120
	* * * * *	
<i>D. kikkawai</i>	-----CAATTTGTAGCTTTTAGCCT-AAAA-----AAAGCAAACACATG	135
<i>D. serrata</i>	-----CAATTTGTAGCTTTTAGCCT-AAAA-----AAAGCAAACACATG	135
<i>D. mauritiana</i>	-----CAATTTGTAGCTTTTAGCCT-AAAA-----AAAGCAAACACATG	135
<i>D. erecta</i>	-----CAATTTGTAGCTTTTAGCCT-AAAA-----AAAGCAAACACATG	135
<i>D. tsacasi</i>	-----CAATTTGTAGCTTTTAGCCT-AAAA-----AAAGCAAACACATG	135
<i>D. simulans</i>	-----CAATTTGTAGCTTTTAGCCT-AAAA-----AAAGCAAACACATG	135
<i>D. orena</i>	-----CCATTTGTAGCTTTTAGCCTTAAAA-----AAAGCAAACACATG	136
<i>D. sechellia</i>	-----CCATTTGTAGCTTTTAGCCTTAAAA-----AAAGCAAACACATG	136
<i>D. melanogaster</i>	-----TAATTTGTAGCTTTTATCCTTAAAAATGTAAATAAATG <u>CCCAATA</u> AAACATG	145
<i>D. teissieri</i>	GGGAAGCTAATAATTTAG-ATTGCTTATAAAGTAAGTAAACAAATTTCCCGTAACACACG	163
<i>D. yakuba</i>	GGGAAGCTAATTTTTGGAGTGTGTGTAAGTAAGAAAACAAATG <u>CCCGTA</u> ACCGCATG	180
	* * * * *	
<i>D. kikkawai</i>	AGCTTGAAATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	190
<i>D. serrata</i>	AGCTTGAAATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	190
<i>D. mauritiana</i>	AGCTTGAAATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	190
<i>D. erecta</i>	AGCTTGAAATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	190
<i>D. tsacasi</i>	AGCTTGAAATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	190
<i>D. simulans</i>	AGCTTGAAATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	190
<i>D. orena</i>	AGCTTTATATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	191
<i>D. sechellia</i>	AGCTTTATATTTTCCAAAGCATT-----TACTTCCAGTAACGGGATAATAGTGTGAACA	191
<i>D. melanogaster</i>	AGCTTGAAATATTACAAAGAAAAA---TGCTTCCAGTAACGGCATAATAGTGTGAACG	201
<i>D. teissieri</i>	TCTAAAGATTAAATATTTTAAAGATAATACGTCTTGTA-----CCCGTGTGAACG	217
<i>D. yakuba</i>	TGCTTAAATATTTTAAACATTAATAATTTTCTAAATATAATCTT-TCCAGTGTGAACG	239
	* * * * *	
<i>D. kikkawai</i>	GAACTTTTACACAACCGAAGCT---ACAGTAAAAGTTTCTCTGGTGT <u>TTCT</u> TAAGCTGCTCT	247
<i>D. serrata</i>	GAACTTTTACACAACCGAAGCT---ACAGTAAAAGTTTCTCTGGTGT <u>TTCT</u> TAAGCTGCTCT	247
<i>D. mauritiana</i>	GAACTTTTACACAACCGAAGCT---ACAGTAAAAGTTTCTCTGGTGT <u>TTCT</u> TAAGCTGCTCT	247
<i>D. erecta</i>	GAACTTTTACACAACCGAAGCT---ACAGTAAAAGTTTCTCTGGTGT <u>TTCT</u> TAAGCTGCTCT	247
<i>D. tsacasi</i>	GAACTTTTACACAACCGAAGCT---ACAGTAAAAGTTTCTCTGGTGT <u>TTCT</u> TAAGCTGCTCT	247
<i>D. simulans</i>	GAACTTTTGCACAACCTGAAGCT---ACAGTAAAAGTTTCCCTGGTGT <u>TTCT</u> TAAGCTGCTCT	247
<i>D. orena</i>	GAACTTTTGCACAACCTGAAGCT---ACTGTAAAAGTTTCCCTGGTGT <u>TTCT</u> TAAGCTGCTCT	248
<i>D. sechellia</i>	GAACTTTTGCACAACCTGAAGCT---ACTGTAAAAGTTTCCCTGGTGT <u>TTCT</u> TAAGCTGCTCT	248
<i>D. melanogaster</i>	AAACTTTTGCACAACCTGAACACTAACAGTAAAAGTTCCGCATG <u>TATTTCT</u> TAAGCTGCTCT	261
<i>D. teissieri</i>	GAACTTTTGCATGCCTGTT---ATATTAAGTTCCCGTGGTGT <u>TATTT</u> TAAGCTCATCT	273
<i>D. yakuba</i>	GAGCTTTTTGCATACCTGTA---ATAGTAAAAGTTCCGCATG <u>TATTTCT</u> CGAGCTACTCT	294
	* * * * *	

**Fig. 2.** Nucleotide sequence of the 5' flanking region of the Cu,Zn SOD gene from nine species of the *melanogaster* group. The light- and dark-shaded sequences represent, respectively, the beginning of the transcribed and translated region. The underlined characters

represent the TATA box for *D. melanogaster* and *D. simulans* (Seto et al. 1989; Kwiatowski et al. 1989) and the CAAT box for *D. melanogaster*, given by Kwiatowski et al. (1989).

<i>D.kikkawai</i>	GCTACGGT <b>CACACCTTAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGCTTTGATAAAAT</b>	307
<i>D.serrata</i>	GCTACGGT <b>CACACCTTAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGCTTTGATAAAAT</b>	307
<i>D.mauritiana</i>	GCTACGGT <b>CACACCTTAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGCTTTGATAAAAT</b>	307
<i>D.erecta</i>	GCTACGGT <b>CACACCTTAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGCTTTGATAAAAT</b>	307
<i>D.tsacasi</i>	GCTACGGT <b>CACACCTTAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGCTTTGATAAAAT</b>	307
<i>D.simulans</i>	GCTACGGT <b>CACACCTTAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGCTTTGATAAAAT</b>	307
<i>D.orena</i>	GCTACGGT <b>CACACCATAGAAGATACCTGGAAAGTTCTCAACATTTTTTCGTTTTGATAAAAT</b>	308
<i>D.sechellia</i>	GCTACGGT <b>CACACCATAGAAGATACCTGGAAAGTTCTCAACATTTTTTCGTTTTGATAAAAT</b>	308
<i>D.melanogaster</i>	GCTACGGT <b>CACACCATAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGTTTTGATAAAAT</b>	321
<i>D.teissieri</i>	GCGACGGT <b>CACACTAGAAGATACCTGGAAAGTTCTCAAGTTTTTTTCGTTTTGATAAAAT</b>	333
<i>D.yakuba</i>	GCGACGGT <b>CACACCGTAAAAGATACCTGGAAAGTTCTCAAGTTTTTTTCGTTTTGATAAAAT</b>	354
	** *	
<i>D.kikkawai</i>	-----TGATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	359
<i>D.serrata</i>	AATCAAATTAATTAATTCCT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGCAATTAATGGCGA</b>	367
<i>D.mauritiana</i>	-----TGATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	359
<i>D.erecta</i>	-----TGATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	359
<i>D.tsacasi</i>	-----TGATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	359
<i>D.simulans</i>	-----TGATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	359
<i>D.orena</i>	-----TAATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	360
<i>D.sechellia</i>	-----TAATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	360
<i>D.melanogaster</i>	-----TGATTAATTCAT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAACGGCGA</b>	373
<i>D.teissieri</i>	AATCAATCTAATTAATTCCT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAATGGCGA</b>	393
<i>D.yakuba</i>	AATCTACTGAATAATTCCT <b>TCGAAATGGTGGTTAAAGCTGTCTGCGTAATTAATGGCGA</b>	414
	* *	

Fig. 2. Continued.

increases in the 5' region to 38.1%. Deletion/insertion segments are observed in this region, so that it varies from 324 bp in *D. mauritiana* to 379 bp in *D. yakuba*. The AT content is high, 64.6% for the 5' region up to the start of translation. According to Seto et al. (1989) the 5' flanking region of *D. melanogaster* Cu,Zn SOD shows a putative imperfect TATA sequence, TATTTCT, and a consensus sequence, CCCAAT (Fig. 2; boldface characters), best matched to the CAAT box. These putative regulatory sequences are found in homologous positions in *D. simulans* (Kwiatowski et al. 1989). The putative TATA box is similar to that of *D. simulans* in all the species, except *D. melanogaster*, *D. teissieri*, and *D. yakuba*. Similarly in the putative CAAT box the latter three species have different nucleotide sequences than the other eight species.

#### Phylogenetic Analysis

We determined the phylogenetic relationships of the *melanogaster* subgroup considering only the intron sequence, including the *D. willistoni* Cu,Zn SOD intron as an outgroup. The coding region is highly conserved and thus not phylogenetically informative. In addition, we used the complete nucleotide sequence of 1673 bp, including the 5' flanking region, the two exons, and the intron.

The alignment of the eight SOD introns of the *melanogaster* subgroup consists of 812 nucleotide positions, of which 448 (55.17%) are variable. The percentage G+C contents of any given species are virtually identical, having a value of 35.0%. The

transition/transversion ratio is 0.7. With the addition of the intron of *D. willistoni*, the nucleotide positions increase to 824, with 86.4% variation (712 variable positions). The transition/transversion ratio has an average value 0.6. The G+C content becomes 35.6% because the G+C content of the outgroup species is high (42.2%). The observed variation in G+C content has been reported earlier in other dipterans (Kwiatowski et al. 1992, 1994) and probably reflects a bias in codon preferences that is revealed in well-expressed genes of other taxa (Sharp et al. 1988).

Given the A+T composition bias, we estimated nucleotide divergence using the Jukes-Cantor (JC), Kimura two-parameter, and Tamura-Nei models (unpublished results) as well as Tamura's three-parameter model (Table 1). All four methods give similar distance patterns. We used Tamura's distances for constructing phylogenetic trees. The NJ, UP-GMA, ME, and MP methods all yield largely similar topologies. Apparently, there is a deep split between two main clusters; one includes *D. sechellia*, *D. simulans*, *D. mauritiana*, and *D. melanogaster*. The phylogenetic trees consistently show that these four species form a monophyletic clade which is supported by a high bootstrap value. The other cluster is separated into two subclusters, one with the pair *D. yakuba*/*D. teissieri* and the other with the pair *D. oreana*/*D. erecta*. The only difference among the phylogenies concerns the species *D. simulans*, *D. mauritiana*, and *D. sechellia* in the case where the MP method was applied. Due to the higher accuracy of the other three methods used, we can disregard the different tree, which is characterized by a pairing of *D. simulans* with *D. sechellia* instead of *D. mauritiana*,

**Table 1.** Pairwise nucleotide divergence between Cu,Zn SOD intron sequences for all substitutions (lower triangle) and standard errors (upper triangle), according to Tamura's distance

Species	1	2	3	4	5	6	7	8	9
<i>D. sechellia</i>		0.013	0.014	0.017	0.028	0.027	0.028	0.028	0.113
<i>D. simulans</i> ,	0.060		0.008	0.014	0.026	0.024	0.025	0.026	0.108
<i>D. mauritiana</i>	0.070	0.021		0.014	0.027	0.025	0.025	0.026	0.109
<i>D. melanogaster</i>	0.090	0.056	0.060		0.026	0.023	0.024	0.025	0.104
<i>D. teissieri</i>	0.219	0.188	0.192	0.184		0.017	0.020	0.018	0.112
<i>D. yakuba</i>	0.200	0.165	0.168	0.153	0.107		0.020	0.019	0.112
<i>D. erecta</i>	0.219	0.180	0.184	0.176	0.132	0.135		0.011	0.107
<i>D. oreana</i>	0.203	0.165	0.169	0.161	0.104	0.114	0.037		0.116
<i>D. willistoni</i>	1.030	1.008	1.006	0.973	1.036	1.019	1.007	1.054	

as found in the trees constructed according to the other methods as well as from the thoroughly studied triplet consisting of *D. simulans*–*D. mauritiana*–*D. sechellia* (see Discussion). The UPGMA tree, the NJ tree, and the ME method (Fig. 3) show that *D. simulans* and *D. mauritiana* are the most closely related species, followed by *D. sechellia* and, finally, *D. melanogaster*. That is, *D. melanogaster* is more distantly related to the *D. simulans*–*D. mauritiana* cluster than *D. sechellia*.

We aligned the complete region, consisting of 1673 sites that include the two exons, the intron, and approximately 350 bp of the 5' flanking region. Nucleotide polymorphism occurs at 458 (27.4%) positions and the G+C content is 42.6%. The nucleotide distances according to the Kimura two-parameter, JC, and Tamura models were used in constructing phylogenetic trees using the NJ and UPGMA methods. Trees were also constructed with the ML method, assuming the molecular clock. The main clusters of the trees are the same, arising from the intron (Fig. 3).

#### Sequence–Structure Selection of the Cu,Zn SOD Protein

The alignment of the amino acid sequences of the nine *Drosophila* Cu,Zn SODs, translated from the nucleotide sequences, plus *C. capitata*, *Xenopus laevis*, and *Bos taurus*, is given in Fig. 4. The two exons consist of 462 bp, of which 46 bp (9.9%) is variable. Of the 46 nucleotide variable positions, 38 (82.6%) are in the third-codon position, 3 (6.5%) are in the second-codon position, and 5 (10.9%) are in the first-codon position. Six amino acid sites (13%) contain nonsynonymous substitutions (three in the first-codon and three in the second-codon position), while the remaining 40 substitutions (86.9%) are synonymous. The alignment of the amino acid sequences of the eight species of the *melanogaster* subgroup plus the one of (used as an outgroup species) *D. willistoni* shows high conservation, having only six different amino acids in a total number of 151 positions. The average G+C content for the eight species is 60.2% for the SOD

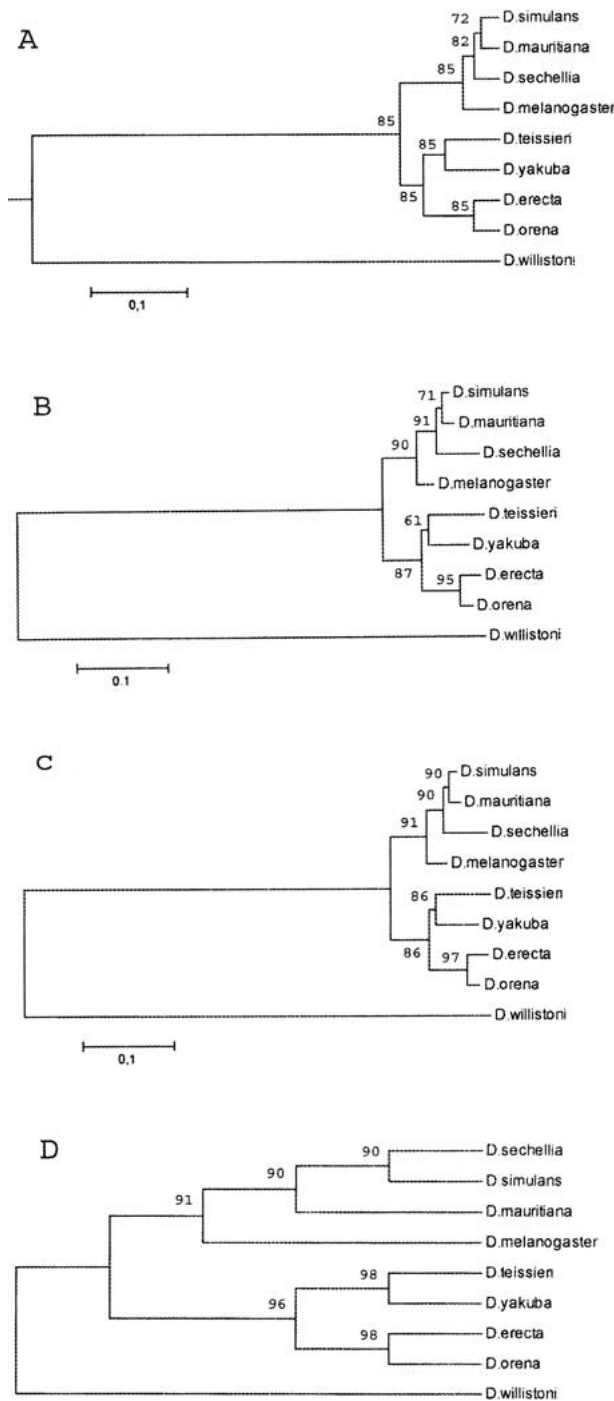
coding region and even higher for the third-codon position (76.3%). The average transition bias for the SOD coding region is 1.8. None of these substitutions occur in functionally important regions of the enzyme.

To determine the importance of the amino acid substitutions that occur between the *melanogaster* subgroup SOD and other species in relation to the protein function, we aligned the *Drosophila* SOD amino acid sequences with the crystallographically determined 3D structure of the Cu,Zn SOD sequence of the frog *Xenopus laevis* (Falconi et al. 1991) and the bovine *Bos taurus* SOD (Tainer et al. 1982). They have, respectively, 63 and 60% identical amino acids with the *Drosophila melanogaster* sequence. In addition, we used the amino acid sequence of another dipteran species that is not classified in *Drosophilidae*, *C. capitata* to estimate the conservative amino acid residues and speculate about their potential structural and functional role in the dipteran's enzyme activity. The percentage identical amino acids of *C. capitata* and *D. melanogaster* was found to be 78.3%.

#### Discussion

The Cu,Zn superoxide dismutase is an abundant enzyme in eukaryotic organisms and exhibits highly specific activity, thus protecting aerobic cells against toxic free oxygen radicals (Fridovich 1986). The Cu,Zn SOD gene of the species analyzed in this study consists of a 462-bp coding region that is interrupted by one intron ranging from 708 bp in *D. yakuba* to 738 bp in *D. teissieri*. As expected, the second SOD intron (<100 bp) that is present in *Scaptodrosophila* (as well as in *Ceratitis capitata* and *Chymomyza*) is absent in the species under investigation. This finding is in accordance with the absence of a second SOD intron from other species of the subgenus *Sophophora* examined previously (Kwiatowski et al. 1994).

The 3D structure of the enzyme from *Xenopus laevis* and *Bos taurus* has been elucidated to 2-Å resolution (Tainer et al. 1982; Falconi et al. 1991). The enzyme crystallizes in space group C2, with two dimeric enzyme molecules per asymmetric unit. The



**Fig. 3.** Phylogenetic trees of eight species of the *melanogaster* subgroup, based on the Cu,Zn SOD intron sequences. **A** UPGMA. **B** Neighbor joining. **C** Minimum evolution. **D** Maximum parsimony. Evolutionary distance was calculated by Tamura's method. *D. willistoni* was taken as an outgroup. Numbers represent the bootstrap values.

functional and structural role of each domain has been well analyzed. Since SOD is a highly conserved enzyme overall and since the critical residues in the enzyme active site were found to be invariant in many known sequences, including *Drosophila melanogaster*, *Ceratitis capitata*, *Gallus gallus*, *Oryctolagus cunicu-*

*lus*, *Rattus norvegicus*, *Ovis aries*, *Sus scrofa*, *Bos taurus*, *Xiphias gladius*, *Prionace glauca*, and *Homo sapiens* (unpublished data), the determined 3D structure of the Cu,Zn SOD of the frog and the cow was thought to fulfill the expectations to be considered representative of the entire class of Cu,Zn SODs. These data seem to be very useful in our attempt to define the respective domains of the *Drosophila* species. Moreover, comparisons among the available sequences of *Drosophila*, frog, and cow as well as the topography of the conserved amino acids may reveal selective or functional constraints that control the evolutionary history of SOD. Each subunit is composed of eight antiparallel  $\beta$ -strands, forming a flattened cylinder of Greek key topology containing 46% of the residues, and three external loops with 48% of the residues. Cu and Zn are held in the active site 6.3 Å apart by interaction with the imidazole ring of His 61 which forms a bridge between the two metals. Important roles for SOD function and structure are played by histidine residues 44, 46, 69, 78, and 118, as well as Asp 81, contributing to metal binding, and glycine residues that are essential to the formation of the  $\beta$ -strand barrel.

The only insertions-deletions occur outside the  $\beta$ -strand regions forming the core of the protein, in loops I and II, and they are limited to two amino acids in length. The N-terminal methionine and the C-terminal valine are removed in the mature protein (Lee et al. 1985). The amino acids at the protein active site are all conserved (Fig. 4). Thus, His44, His46, His61, and His118 and His61, His69, His78, and Asp81, the crucial amino acids for Cu and Zn binding, respectively, are conserved for all species examined.

The conservation rate is also high around the active site, at the positions where amino acids interact with the amino acids covalently linked to the metals. Residues that hydrogen-bond to the Cu-linked amino acids, such as His41, Gly42, Thr114, Val116, Asp122, and Gly139, are all conserved. Furthermore Asn63, Lys67, Agr77, and His78, which are found in the Zn-binding region of loop IV (from residue 52 to residue 83) and making hydrogen bonds to Zn-linked amino acids, are also conserved. The only exception is Lys68. At position 68 the lysine found in bovine SOD changes to asparagine in the frog and to glutamic acid in *Drosophila*. The threonine 133 in bovine is found as leucine and lysine in frog and *Drosophila*, respectively. Both are located on the surface of the protein, far away from the active center or the subunit interface. The change from positive charge (Lys) to glutamic acid and to asparagine seems not to cause a significant change in the overall charge distribution as calculated using DelPhi (Honig and Nichols 1995). The degree of homology of SOD of the nine *Drosophila* species is 97% for this disulfide and zinc ligand-containing region.

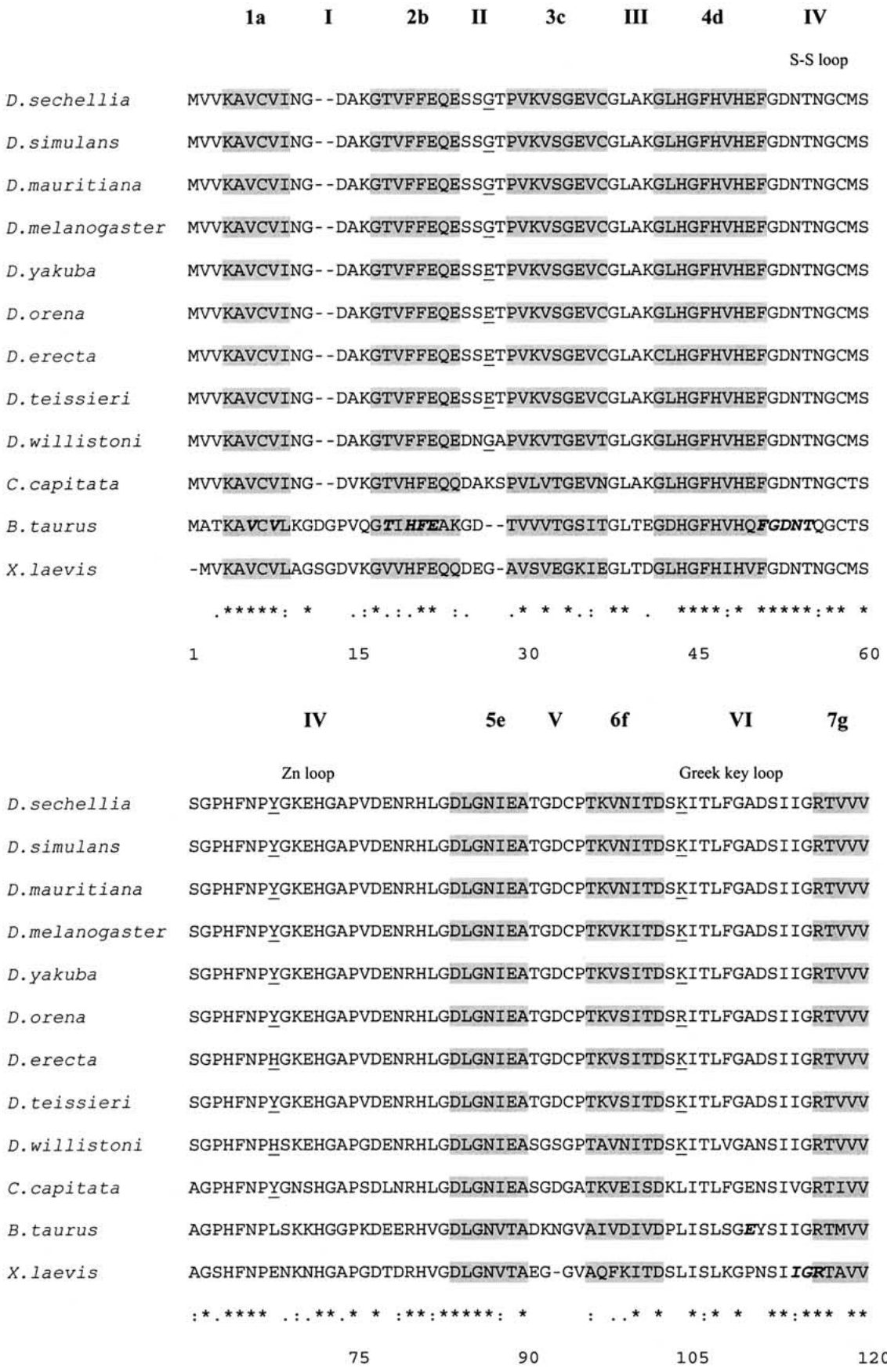


Fig. 4. Multiple alignment of the amino acid sequences of 12 Cu,Zn SODs, from eight species of the *D. melanogaster* subgroup plus *D. willistoni*, *C. capitata*, *Xenopus laevis*, and *Bos taurus*. The numbering uses as reference the bovine amino acid sequence. The shaded boxes refer to the  $\beta$ -strand conformation of the bovine structure. The underlined and the bold-face italic letters refer to the

amino acids that play an important role in the formation of the upper rim of the electrostatic channel and in the contact between the two subunits, respectively. The symbols at the bottom correspond to the definitions of the CLUSTAL program: (\*) fully conserved; (:) highly conserved; (.) conserved substitution.



	7g	VII	8h
	electrostatic channel loop		
<i>D. sechellia</i>	HADADDLGQGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. simulans</i>	HADADDLGQGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. mauritiana</i>	HADADDLGQGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. melanogaster</i>	HADADDLGQGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. yakuba</i>	HADADDLGKGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. orena</i>	HADADDLGKGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. erecta</i>	HADADDLGKGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. teissieri</i>	HADADDLGQGGHELKSTGNAGARIGCGVIGIAKV		
<i>D. willistoni</i>	HADPDDLKGGHELKSTGNAGARIGCGVIGIAKI		
<i>C. capitata</i>	HADPDDLKGGHELKSTGNAGARLGGCVIGICKI		
<i>B. taurus</i>	HEKPDDLGRGGNEESTKTGNAGSRLACGVIGIAK-		
<i>X. laevis</i>	HEKADDLGKGGNDESLKTGNAGGRLACGVIGYSP-		
	* ..*****:**: * ..*****.*:***** .		
	135		150

Fig. 4. Continued.

The amino acids that define the active site's electrostatic channel (Getzoff et al. 1983) are highly conserved (Fig. 4). The amino acid substitutions at positions Glu130, Leu131, Ser134, Thr135, Ala119, Asp120, His129, and Lys133 support the classification of *Drosophila* electrostatic shell in the fourth category of the seven arrangements given by Bordo et al. (1994). All amino acids at the above-mentioned positions are conserved in the *Drosophila* species and are in agreement with the total arrangement of the charges in the electrostatic channel.

The high content of glycines that contribute to the active-site stability as well as to the formation of the characteristic  $\beta$ -strand barrel structure represent another conserved protein feature. Most glycines are conserved in all examined species not only in abundance but also in the positions where they are located. Twenty-four glycine residues were found in identical positions, of a total of 25, in the nine *Drosophila* species examined. The few exceptions noticed in the overall alignment in Fig. 4 (Gly39, Gly89, Gly90, Gly91) are located in external loops that contribute little to the total conformation of the enzyme.

There is also conservation in the amino acid positions that participate in subunit contact such as Lys151 and Asp50, Val146 and Ile111, Gly49 and Gly112, and, finally, Gly148. Opposite substitutions

occur at positions 107 (Pro, Glu, Ala) (which is in contact with the Arg113 at the opposite subunit), 17 (Val, Thr), 19 (His, Phe), 149 (Tyr, Ile), 150 (Ser, Ala), and 151 (Pro, Lys). The substitutions Val17Thr, His19Phe, Tyr149Ile, Ser150Ala, and Pro151Lys, although at the edge of the subunit interface, seem not to contribute significantly to the subunit association. The Pro at position 107 occupies the  $i + 2$  position of a tight external  $\beta$ -bend. The substitution by Ala or Glu may not necessarily affect the conformation and certainly does not disrupt any local interactions. The substitutions found in the amino acid sequences of the *Drosophila* species with respect to the frog and bovine 3D structures seem not to make important contributions to the 3D structure and not to affect the enzyme's activity, as they are localized at the surface of the protein in parts not associated with enzyme's function. None of these are found in any essential position of the active center or at the electrostatic channel.

These observations are in agreement with the expectations. In general, the enzyme is well conserved over long time spans; thus, about 60% of the amino acid residues were found to remain identical between organisms from different kingdoms, such as humans and yeasts. The substitutions that have been established and maintained among the various species

have passed through selective processes that depend on the crucial role of the substituted amino acids (Ayala 1986).

We inferred the phylogenetic relationships of species of the *melanogaster* subgroup by using only the intron sequence. The coding region was found to be highly conserved and thus not informative for our phylogenetic analysis. In contrast, Kwiatowski et al. (1994) used exclusively the coding sequence of *Sod* for phylogenetic analysis but the circumstances were quite different from those in our study. Concerning the analysis of distantly related *Drosophila* species, the noncoding regions were so highly diverse that their alignment raised many uncertainties in several cases. In the case of closely related species such as *D. mauritiana*, *D. simulans*, and *D. sechellia*, we found that the *Sod* coding sequence does not contain enough phylogenetic information to resolve the branching pattern (data not shown). In addition, the selection of a proper species as an outgroup was of pivotal importance for our study. *D. willistoni*, which belongs to the subgenus *Sophophora*, is an almost closely related but well-defined outgroup and the split between *D. willistoni* and *D. melanogaster* cluster seems to have occurred a long time ago (about 36 Mya).

The *D. simulans* complex has been the subject of many efforts to infer phylogeny, and indeed, all three possible pairs of species have been proposed as the most closely related species pair, including *simulans/sechellia*, *simulans/mauritiana*, and also the *sechellia/mauritiana* pairing, which seems more unlikely compared to the two other alternative pairs. This difficult phylogenetical problem was approached by constructing a cluster of diagrams for 14 genes, thus grouping the species by various combinations (Kliman et al. 2000), in an attempt to clarify whether the next most similar sequences to the ones of *D. sechellia* are from *D. simulans* or *D. mauritiana*. Hey and Kliman (1993) suggested that the origin of *D. sechellia* arose prior to the splitting that gave rise to the *D. simulans* and *D. mauritiana* species. Therefore Kliman et al. (2000) presented an analysis involving 14 genes, in the framework of the divergence population genetics (DPG) approach, with the plurality of the cluster diagrams favoring the aforementioned explanation. This conclusion is also supported by a recent study by Ting et al. (2000) using the *OdsH* locus. Taking into consideration all these data as well as the peculiarity of the analysis used in the phylogenetic model presented by Kliman et al. (2000), which does not employ any assumptions of instantaneous splitting among distinct homogeneous entities, we may conclude that *D. simulans* and *D. mauritiana* species are more closely related to each other than either are to those of *D. sechellia*. What appears to be the most unlikely pairing according to

the cluster analyses presented and to biogeographic grounds (*D. sechellia* and *D. mauritiana*) was the favored topology in the studies by Caccone et al. (1988, 1996), suggesting that *D. mauritiana* and *D. sechellia* are the most closely related species, followed by *D. simulans* and *D. melanogaster*. According to the phylogenetic trees constructed with our sequences, all trees suggest that *D. sechellia* arose prior to the species pair *D. mauritiana/D. simulans*. *D. erecta* and *D. orena* cluster together, and, in parallel, *D. yakuba* and *D. teissieri* form another species pair. This topology is supported by the UPGMA, NJ, ME, and MP trees, with an exception in the clustering of *D. sechellia* in the last tree. The UPGMA and NJ trees are based on molecular distances. The MP tree is based on character-state differences. The ME trees with JC distances, Kimura's (1980) two-parameter distances, Tajima and Nei's (1984), and Tamura's (1992) distance were constructed and the topology remained the same. These findings are in accordance with the generally accepted phylogeny of the subgroup *melanogaster*.

It should be mentioned here that many questions have been raised dealing with the ability of the pure molecular information to provide definitive answers and lead to important conclusions concerning phylogenetic matters. Furthermore, a certain debate does exist dealing with the extent to which the genealogical reconstruction of single-copy genes accurately reflects "phylogeny." In fact, several sets of molecular data provide information not always compatible with other taxonomic knowledge or consistent with each other (Kwiatowski et al. 1994). From this point of view, nobody can speculate that molecular data are necessarily better than morphological data. Molecular phylogenies are also known to be affected by various sources of errors (Nei 1991). Moreover, phylogenetic trees that have been constructed from different parts of DNA are, in several occasions, inconsistent with each other. Some sequence data such as those for mitochondrial DNA and nuclear rRNA seem to be less informative than other sequence data (i.e., *Adh*) except for some particular purposes (Russo et al. 1995). For example, it is well known that the low level of polymorphism in ribosomal genes can be attributed to the action of concerted evolution. The ribosomal genes that exist in many copies [400–600 in *Xenopus laevis* and 130–250 in *D. melanogaster* (Nei and Koehn 1983)] look almost-identical to each other, thus lacking the plurality of information that could be used for phylogenetic work. It should always be kept in mind that studies based on data for a single gene may have also been subjected to some undetected peculiarities of the gene which distort the phylogenetic trees. Therefore, it is preferable to examine the phylogeny by using several sets of DNA sequences, corresponding to

various genes. But, unambiguously, the DNA sequences of any organism always carry the record of its evolutionary history. Taking into consideration the topology of the phylogenetic trees constructed by Kwiatowski et al. (1994), based on Cu,Zn SOD, these trees were also inconsistent with trees based on the *Adh* gene (Russo et al. 1995) as well as others. This is why the SOD tree was disregarded in the past. Notably, in the present study, we showed that the *Sod* gene can be considered as a reliable gene contributing to the phylogenetic analysis of the *melanogaster* species subgroup. According to Zuckerkandl and Pauling (1962) the rate of amino acid substitutions in proteins may be constant over evolutionary time, a hypothesis called the "molecular evolutionary clock." In the past, Lee et al. (1984) reported that SOD is not an acceptable evolutionary clock. The rates of amino acid substitutions grossly departed from constancy, yielding contradictory results concerning the rate of evolution in studies involving various organisms (human, horse, cow, *Drosophila*, yeast). However, further investigations revealed that the rate of amino acid substitutions in the Cu,Zn SOD of diverse organisms has been fairly constant during the last 60 million years (MY). This rate is approximately 15 aa/100 aa/100 MY for PAM-corrected data (PAM is the estimated percentage of amino acid differences corrected for superimposed and back replacements) (Kwiatowski et al. 1991a, 1992, 1994). Nevertheless, investigators should always take into account the warning of Lee et al. (1985), who concluded that suggestions or estimations about evolutionary events based on the primary sequence of a gene or protein may be subject to considerable errors.

## References

- Asada K, Kanematsu S, Okaka S, Hayakawa T (1980) In: Banister JV, Hill HAO (eds) Chemical and biochemical aspects of superoxide dismutase. Elsevier/North Holland, New York, pp 136–153
- Ayala FJ (1986) On the virtues and pitfalls of the molecular evolutionary clock. *J Hered* 77:226–235
- Benov LT, Fridovich I (1994) *Escherichia coli* expresses a copper- and zinc-containing superoxide dismutase. *J Biol Chem* 269 (41):25310–25314
- Beverly SM, Wilson AC (1982) Molecular evolution in *Drosophila* and higher Diptera. I. Micro-complement fixation studies of a larval hemolymph protein. *J Mol Evol* 18:251–264
- Bordo D, Djinovic K, Bolognesi M (1994) Conserved patterns in the Cu,Zn superoxide dismutase family. *J Mol Biol* 238:366–386
- Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci USA* 75:4853–4857
- Caccone A, Amato GD, Powell JR (1988) Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics* 118:671–683
- Caccone A, Moriyama EN, Gleason JM, Nigro L, Powell JR (1996) A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data. *Mol Biol Evol* 13:1224–1232
- Cavener DR (1987) Comparison of the consensus flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acid Res* 15:1353–1361
- Falconi M, Rotilio G, Desideri A (1991) Modelling the three-dimensional structure and electrostatic potential field of the two Cu,Zn superoxide dismutase variants from *Xenopus laevis*. *Proteins* 10(2):149–155
- Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17:368–376
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791
- Felsenstein J (1995) PHYLIP: Phylogeny inference package, version 3.5c, University of Washington, Seattle
- Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* 155:279–284
- Fitch WM (1971) Towards defining the course of evolution: Minimum change for a specific tree topology. *Sys Zool* 20: 406–416
- Fleming JE, Miquel J, Cottrel SF, Yenogoyan LS, Economos AC (1982) Is cell aging caused by respiration-dependent injury to the mitochondrial genome? *Gerontology* 28:44–53
- Fridovich I (1986) Superoxide dismutases. *Adv Enzymol* 58:61–97
- Getzoff ED, Tainer JA, Weiner P, Kollman PA, Richardson J, Richards DC (1983) Electrostatic recognition between superoxide and copper, zinc superoxide dismutase. *Nature* 306:287–290
- Grimaldi DA (1990) A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bull Am Mus Nat Hist* 197:1–139
- Halliwell B, Gutteridge GMC (1984) Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochem J* 219:1–14
- Harman D (1982) The free radical theory of aging. In: Pryor WA (ed) Free radicals in biology, vol V. Academic Press, New York, pp 255–275
- Herny JM, Raina AK, Ridgway RL (1990) Isolation of high molecular weight DNA from insects. *Anal Biochem* 185:147–150
- Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* 268:1144–1149
- Katoh T, Tamura K, Aotsuka T (2000) Phylogenetic position of the subgenus *Lordiphoza* of the genus *Drosophila* (Diptera: Drosophilidae) inferred from alcohol dehydrogenase (*Adh*) gene sequences. *J Mol Evol* 51:122–130
- Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156(4):1913–1931
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2: Molecular evolutionary genetics analysis software (submitted for publication)
- Kwiatowski J, Patel M, Ayala F (1989a) *Drosophila melanogaster* Cu,Zn superoxide dismutase gene sequence. *Nucleic Acid Res* 17:1264
- Kwiatowski J, Gonzalez F, Ayala F (1989b) *Drosophila simulans* Cu,Zn superoxide dismutase gene sequence. *Nucleic Acid Res* 17:6735
- 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 F (1991b) High fidelity of polymerase chain reaction. *Mol Biol Evol* 8(6):884–887
- Kwiatowski J, Skarecky D, Ayala FJ (1992) 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 Evol* 1:72–82

- Kwiatowski J, Skarecky D, Bailey K, Ayala FJ (1994) Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu,Zn Sod gene. *J Mol Evol* 38:443–454
- Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol Biol* 22:159–226
- Lee YM, Friedman DJ, Ayala F (1985) Superoxide dismutase: An evolutionary puzzle. *Proc Natl Acad Sci USA* 82:824–828
- McCord JM, Fridovich I (1969) Superoxide dismutase. An enzymic function for erythrocyte (hemocuprein). *J Biol Chem* 244:6049–6055
- McPherson M, Quirke P, Taylor GR (1991) PCR. A practical approach. IRL Press, Oxford, New York, Tokyo
- Nei M (1991) Relative efficiencies of different tree-making methods for molecular data. In: Miyamoto MM, Cracraft J (eds) *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York, pp 90–128
- Nei M, Koehn RK (1983) *Evolution of genes and proteins*. Sinauer Associates, Sunderland, MA
- Pinol J, Fancino O, Fontdevila A, Cabre O (1988) Rapid isolation of *Drosophila* DNA. *Nucleic Acids Res* 16:1726
- Puget K, Michelson AM (1974) Isolation of a new copper-containing superoxide dismutase bacteriocuprein. *Biochem Biophys Res Commun* 58:830–838
- Russo CAM, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of Drosophilid species. *Mol Biol Evol* 12(3):391–404
- Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol* 9:945–967
- Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol* 10:1073–1095
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning, a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74(12):5463–5467
- Seto N, Hayashi S, Tener GM (1989) Cloning, sequence analysis and chromosomal localization of the Cu-Zn superoxide dismutase gene of *Drosophila melanogaster*. *Gene* 75:85–92
- 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*, *Drosophila melanogaster* and *Homo sapiens*: A review of the considerable within-species diversity. *Nucleic Acids Res* 16:8207–8211
- Sneath PHA, Sokal RR (1973) *Numerical taxonomy*. Freeman, San Francisco
- Solignac M, Monnerot M (1986) Race formation, speciation and introgression within *Drosophila simulans*, *D. mauritiana* and *D. sechellia* inferred from mitochondrial DNA analysis. *Evolution* 40:531–539
- Solignac M, Monnerot M, Mounolou JC (1986) Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. *J Mol Evol* 23:31–40
- Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC (1982) Determination and analysis of the 2 Å structure of Cu-Zn SOD. *J Mol Biol* 160:181–217
- Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1:269–285
- Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content bias. *Mol Biol Evol* 9:678–687
- Throckmorton LH (1975) The phylogeny, ecology, and geography of *Drosophila*. In: King RC (ed) *Handbook of genetics*. Plenum Press, New York, pp 421–469
- Ting CT, Tsaur SC, Wu CI (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *odysseus*. *Proc Natl Acad Sci USA* 97:5313–5316
- Tompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680
- Vontas J, Fotinos S, Loukas M, Tsakas S (1999) The selective effect of paraquat (herbicide), ozone and UV radiation on the viability and recombinogenicity of *D. melanogaster* in relation to its SOD polymorphism. *Contributions to the Zoogeography and Ecology of the Eastern Mediterranean Region*. Vol 1, pp 299–304
- Wheeler MR (1986) Additions to the catalog of the world's *Drosophilidae*. In: Ashburner M, Carson HL, Thompson JN (eds) *The genetics and biology of Drosophila*, vol 3a. Academic Press, New York, pp 395–409
- Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13(5):155–556
- Zuckerandl E, Pauling L (1962) Molecular disease, evolution and genic heterogeneity. In: Kasha M, Pullman B (eds) *Horizons in biochemistry*. Academic, New York, pp 189–225