# UC Irvine
## UC Irvine Previously Published Works


### Title
Structure and sequence of the Cu,Zn Sod gene in the Mediterranean fruit fly, Ceratitis capitata: Intron insertion/deletion and evolution of the gene


### Permalink
https://escholarship.org/uc/item/4p19m8d3



### Journal
Molecular Phylogenetics and Evolution, 1(1)

### ISSN
1055-7903



### Authors
Kwiatowski, Jan
Skarecky, Douglas
Ayala, Francisco J



### Publication Date
1992-03-01

### DOI
10.1016/1055-7903(92)90037-h



### Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

MOLECULAR PHYLOGENETICS AND EVOLUTION
Vol. 1, No. 1, March, pp. 72–82, 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[1]

JAN KWIATOWSKI,*,† DOUGLAS SKARECKY,* and FRANCISCO J. AYALA*,[2]

*Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717; and †Institute of Botany, Warsaw University, 00-478 Warsaw, Poland

Received December 10, 1991, revised April 14, 1992

**We have cloned a 4-kb region encompassing the *Cu,Zn superoxide dismutase* (*Sod*) gene from a genomic library of the Mediterranean fruit fly, *Ceratitis capitata*, using a cDNA probe from *Drosophila melanogaster*. The coding sequence of 462 bases is equally as long as that in *Drosophila* species. The rate of amino acid replacement over the past 100 million years is approximately the same in the Diptera and in mammals, thus excluding the hypothesis (proposed to account for an apparent acceleration in rate of evolution of *Sod* over geological time) that the evolution of the SOD protein is much higher in the mammals than in other organisms. The coding region is interrupted by two introns in *Ceratitis*, whereas only one occurs in *Drosophila*. Phylogenetic comparisons indicate that the second intron was present in the common dipteran ancestor, but was lost shortly after the divergence of the *Drosophila* lineage from other Diptera. Analysis of the exon/intron structure of *Sod* in various animal phyla, plants, and fungi indicates that intron insertions as well as deletions have occurred in the evolution of the *Sod* gene.** © 1992 Academic Press, Inc.

## INTRODUCTION

The superoxide dismutases (SOD) are abundant enzymes in aerobic organisms, with highly specific superoxide dismutation activity that protects the cell against the harmfulness of free oxygen radicals (Fridovich, 1986). These enzymes have active centers that contain either iron or manganese, or both copper and zinc (Fridovich, 1986). They have a complex evolutionary history (Kwiatowski, 1987).

The Cu,Zn SOD is a well-studied protein, found in eukaryotes but also in some bacteria (Steinman, 1988). The amino acid sequence is known in many organisms, plant, animals, fungi, and bacteria (Kwiatowski *et al.*, 1991a). The three-dimensional structure has been determined at a 2-Å resolution in cows (Tainer *et al.*, 1982); it is conserved in humans (Getzoff *et al.*, 1989) and presumably in bacteria (Bannister and Parker, 1985). The amino acids essential for catalytic action (Tainer *et al.*, 1983), as well as those for protein structure, are strongly conserved (Getzoff *et al.*, 1989). The exon/intron structure of the gene is known in humans (Levanon *et al.*, 1985), several *Drosophila* species (Kwiatowski and Ayala, 1989; Kwiatowski *et al.*, 1989a,b), the fluke (Simurda *et al.*, 1988), maize (R. Cannon, personal communication), and *Neurospora* (Chary *et al.*, 1990). Humans and *Drosophila* share a common intron site, but the location of all other introns varies in different groups of organisms.

The Mediterranean fruit fly, *Ceratitis capitata*, is a cosmopolitan pest of great agricultural interest because it feeds on fleshy, succulent fruits. *Ceratitis* belongs to the family Tephritidae, which is closely related to the Drosophilidae. The phylogenetic divergence of these two families is dated about 80 million years (my) ago, which corresponds to the radiation of modern mammal families, for which the possibility of a greatly accelerated rate of SOD evolution has been raised (Ayala, 1986).

## MATERIALS AND METHODS

*Cloning of the Sod gene.* A genomic library of *C. capitata* cloned into a *Bam*HI site of λ EMBL4 (titer $2 \times 10^7$ pfu/ml), constructed by M. Rina and B. Savakis at the IMBB Research Center of Crete, Greece, was made available by Dr. D. King of Harvard University. Phages ($2.5 \times 10^5$) were plated and transferred to nylon filters (colony/plaque screen, NEN–Dupont). The filters were hybridized under low stringency conditions (Church and Gilbert, 1984) with $^{32}$P-labeled cDNA of *D. melanogaster Cu,Zn Sod* (obtained as an insert in plasmid pUC13 from Dr. Gordon M. Tener; see Seto *et al.*, 1987b). The filters were incubated in the hybridiza-

Nucleotide sequence data from this article have been deposited with the GenBank Data Library under Accession No. M76975.

[1] This research was supported by National Institutes of Health Grant GM42397 and by Department of Energy Grant ER60713.

[2] To whom correspondence and reprint requests should be addressed.



1055-7903/92 $5.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

tion buffer (7% SDS, 1% BSA, 250 mM NaCl, 10% PEG 8000, 100 mM EDTA, and 500 mM sodium phosphate, pH 7.2) at 55°C for several hours and then hybridized for 2 days in the same buffer with radioactive probe and 200 μg/ml of salmon sperm DNA. The filters were washed successively in 2× and 1× SSC at 55°C for 2 h each, covered with plastic film, and autoradiographed for 72 h at −70°C with two intensifying screens. Positive clones were further purified in two subsequent steps of rescreening.

DNA from the purified clones was obtained by the plate lysate method (Maniatis *et al.*, 1982), digested with appropriate restriction enzymes, electrophoresed, and subjected to Southern blot transfer (Southern, 1975; Reed and Mann, 1985) in order to identify the DNA fragments carrying the *Sod* gene. Appropriate fragments were subcloned into pGEM4Z or pUC19 vectors for further analysis (see Fig. 1). PCR amplification and subsequent cloning of the product were performed as described in Kwiatowski *et al.* (1991b). The PCR primers are shown in Fig. 2.

*DNA sequence analysis.* The plasmid DNA was purified by small-scale alkaline lysis (Maniatis *et al.*, 1982; Ausubel *et al.*, 1987), analyzed by restriction digests, and subcloned into the pUC19 vector for sequencing. The DNA sequences of the restricted fragments were determined by the dideoxynucleotide chain-termination method (Sanger *et al.*, 1977) of subclones of pUC19 and/or pGEM4Z plasmids. Before sequencing, 2–5 μg of double-stranded DNA from a minipreparation was denatured by heating at 95°C for 5 min in 10 μl of a solution with the annealing buffer (Sequenase kit) and 2 μl of a 1 μM solution of the appropriate sequencing primers and incubated at 37°C for 20 min and at room temperature for another 10 min at least. The sequencing reactions followed the manufacturer's (U.S. Biochemical Corp.) protocol using modified phage T7 DNA polymerase (Sequenase) (Tabor and Richardson, 1987) with [$^{32}$P]dATP (Amersham; 3000 Ci/mmol) as the radioactive component. After electrophoresis, 6% gels were dried at 80°C without fixing and autoradiographed overnight.

*Computer-assisted sequence analysis.* DNA and protein sequences were analyzed using the University of Wisconsin Genetics Computer Group Sequence software package, Version 5.2 (Devereu *et al.*, 1984), and the Ligase package written by Mr. Robert Tyler in our laboratory. Phylogenetic analyses were made with the PHYLIP 3.2 package (Felsenstein, 1989).

## RESULTS

*Cloning and sequencing of the Cu,Zn Sod gene.* Figure 1 summarizes the cloning and sequencing strategies used for isolating and characterizing the *Cu,Zn Sod* gene of the Mediterranean fruit fly, *C. capitata*

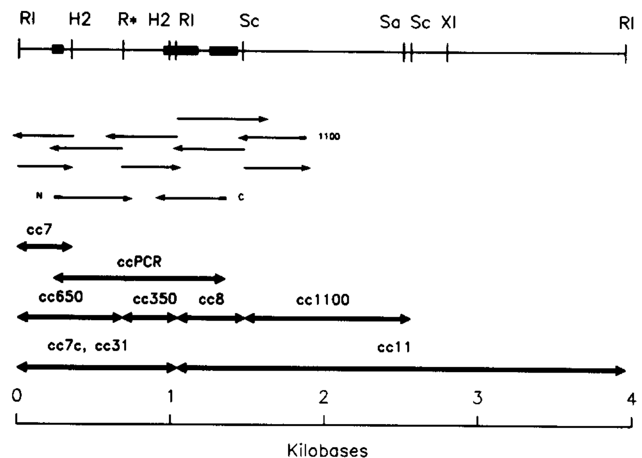


**FIG. 1.** Partial restriction map and sequencing strategy of the *Cu,Zn Sod* gene region of the Mediterranean fruit fly, *C. capitata*. Top: Restriction map of a 4-kb fragment with exons shown as black rectangles; the symbols for the restriction sites represent *Eco*RI (RI), *Eco*RI* (R*), *Hin*dII (H2), *Sal*I (Sa), *Sca*I (Sc), and *Xba*I (XI). Various subclones are indicated as thick double-headed arrows. The thin arrows represent the direction and extent of sequencing; three custom sequencing primers are shown as short bars labeled 1100, N, and C.

(Tephritidae). Ten colonies with positive hybridization signals were obtained when a genomic phage library of *C. capitata* was screened with a *Cu,Zn Sod* probe from *Drosophila melanogaster*. The DNA of all positive colonies gave identical patterns of hybridization bands upon digestion with various restriction endonucleases (data not shown), indicating that all 10 colonies had DNA fragments encompassing the same *Sod* gene. The partial restriction map of a 4-kb fragment comprising the gene is shown at the top of Fig. 1.

Two *Eco*RI fragments, 1.0 and 3.0 kb long (bottom, Fig. 1), were subcloned for sequencing purposes. We ascertained that these two fragments are contiguous by sequencing through the *Eco*RI site in the middle of the clone labeled ccPCR in Fig. 1, which was obtained by PCR amplification of the phage DNA using as primers carboxy- and amino-end oligomers of the *Sod* coding sequence that are shared by a number of *Drosophila* species (unpublished data from our laboratory). Nucleotide sequencing was carried out on four subclones of smaller fragments (cc650, cc350, cc8, and cc1100, bounded by *Eco*RI, *Eco*RI*, *Hin*dII, and *Sca*I sites; see Fig. 1) by means of (1) standard M13 vector sequencing primers, (2) our PCR primers, and (3) a primer specifically synthesized to sequence the polyadenylation-signal region, which is included in fragment cc1100 within a long segment bounded by *Sca*I and *Sal*I that lacks any sites for the endonucleases used.

No nucleotide-site differences were observed in three independent sequences of the coding region obtained from two different phage clones and from a PCR clone

from a third phage. We have, therefore, no evidence for the presence of more than one copy of the *Cu,Zn Sod* gene in the *Ceratitis* genome, a possibility suggested by Fernandez-Sousa and Michelson's (1976) observation of two activity bands upon electrofocussing of a *Ceratitis* protein extract.

*Gene structure.* The coding region was defined by comparing the *Sod* sequence of *Ceratitis* with those of several *Drosophila* species (Seto *et al.*, 1987a,b; Kwiatowski and Ayala, 1989; Kwiatowski *et al.*, 1989a,b), all of which show sharply defined boundaries between highly conserved and poorly conserved regions at the intron splice sites and at the start and stop codons. Evolutionary comparisons between species are also effective for identifying putative promoter and enhancer sequences (Blackman and Meselson, 1986; Efstratiadis *et al.*, 1980).

Figure 2 gives the nucleotide sequence of the *Ceratitis* gene, as well as the inferred amino acid sequence.

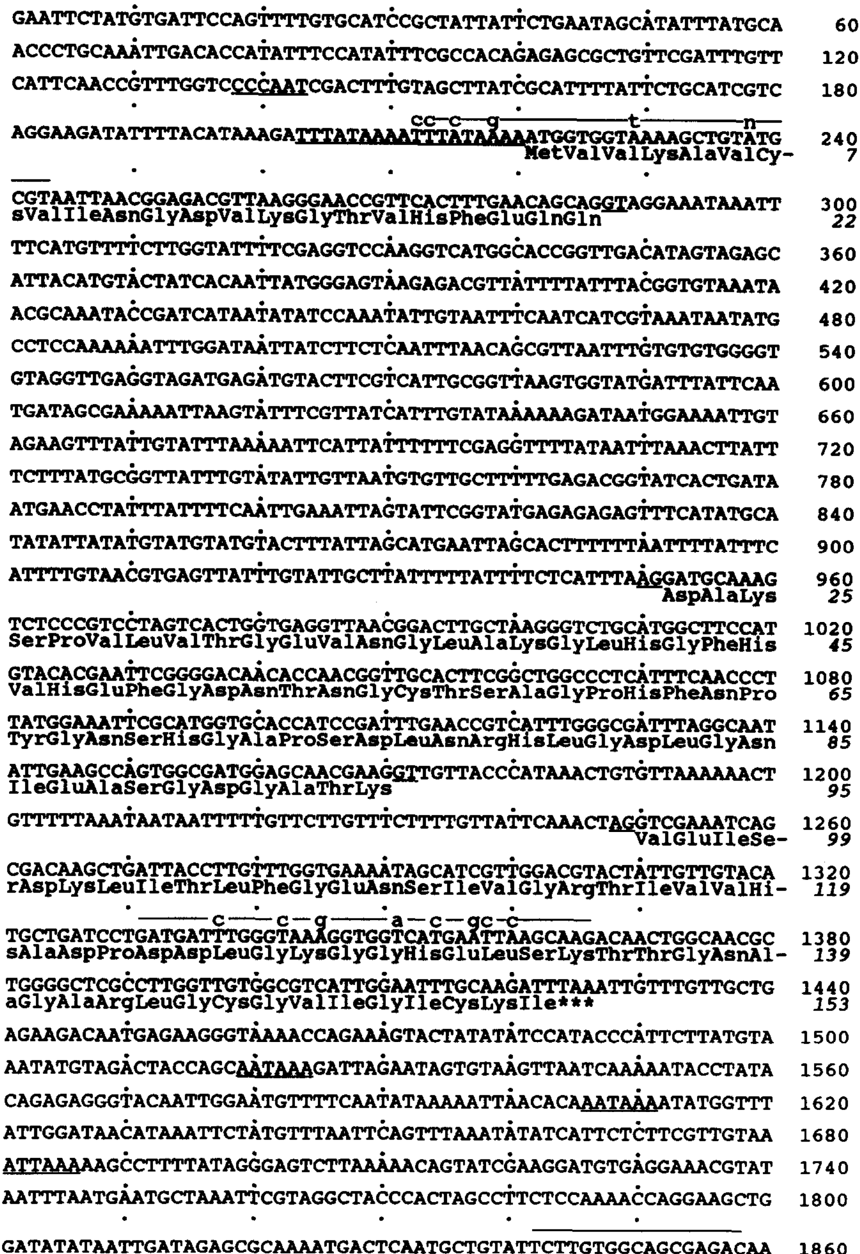

FIG. 2. Nucleotide sequence of the *Cu,Zn Sod* gene in *C. capitata*. The inferred amino acid sequence of the polypeptide is shown below the corresponding codons, starting with the initial methionine; the stop codon is indicated with a triple asterisk; and the amino acids are numbered in italics at the right margin. Putative regulatory sequences are underlined. Lines above the sequence represent three custom oligonucleotide sequencing primers, with mismatches shown in lowercase; two of these primers (represented by N and C in Fig. 1) were also used for PCR amplification of phage DNA.

The 462-nucleotide-long coding sequence exhibits high similarity with the equally long coding sequence of three *Drosophila* species, *melanogaster* (Kwiatowski *et al.*, 1989b), *simulans* (Kwiatowski *et al.*, 1989a), and *virilis* (Kwiatowski and Ayala, 1989).

The exon/intron structure is different in *Ceratitis*, where two introns exist, and in the *Drosophila* species, which have only one. The first intron is most probably homologous in *Ceratitis* and the *Drosophila* species, since in all of them it is located between the 22nd and 23rd codons (Fig. 3). This intron is 665 bp long in *Ceratitis* but varies in *Drosophila* between 321 bp (*D. subobscura*, our unpublished results) and 725 bp (*D. melanogaster*, Kwiatowski *et al.*, 1989b). A possibly homologous intron is found at the same location in the human gene, in which four introns exist in all (Levanon *et al.*, 1985). The second intron of *Ceratitis* is rather short, 79 bp in length, and is located between codons 95 and 96, which occur within an evolutionarily highly variable region of the polypeptide (see Ayala, 1986, and below). No intron occurs at this location in humans or in *Neurospora* (Chary *et al.*, 1990), fluke (Simurda *et al.*, 1988), or maize (Cannon, personal communication). However, a homologous intron, placed between codons 95 and 96, exists in a closely related dipteran species, *Chymomyza amoena* (Kwiatowski *et al.*, 1992). The presence of the second intron in the two genera, *Ceratitis* and *Chymomyza*, and its absence from the *Drosophila* species suggest that it was present in the ancestral dipteran lineage but was lost early in the evolution of the *Drosophila* genus, before the split of the two subgenera, *Drosophila* and *Sophophora* (see Discussion below).

The two introns are bounded in all the *Cu,Zn Sod* dipteran genes by the consensus pairs 5′-GT. . .AG-3′ (Breathnach and Chambon, 1981). The relative size of the two *Sod* introns is consistent with the observation that in dipteran genomes the first intron from the 5′ end is usually larger than introns toward the carboxyl end of the protein (Cherbas *et al.*, 1986; Hawkins, 1988; Riley, 1989).

*C+G content and codon use.* The C+G content is 47% in the coding region of the *Sod* gene of *C. capitata*; it is 28 and 23% for introns I and II, respectively. All organisms, except vertebrates and fungi, have higher

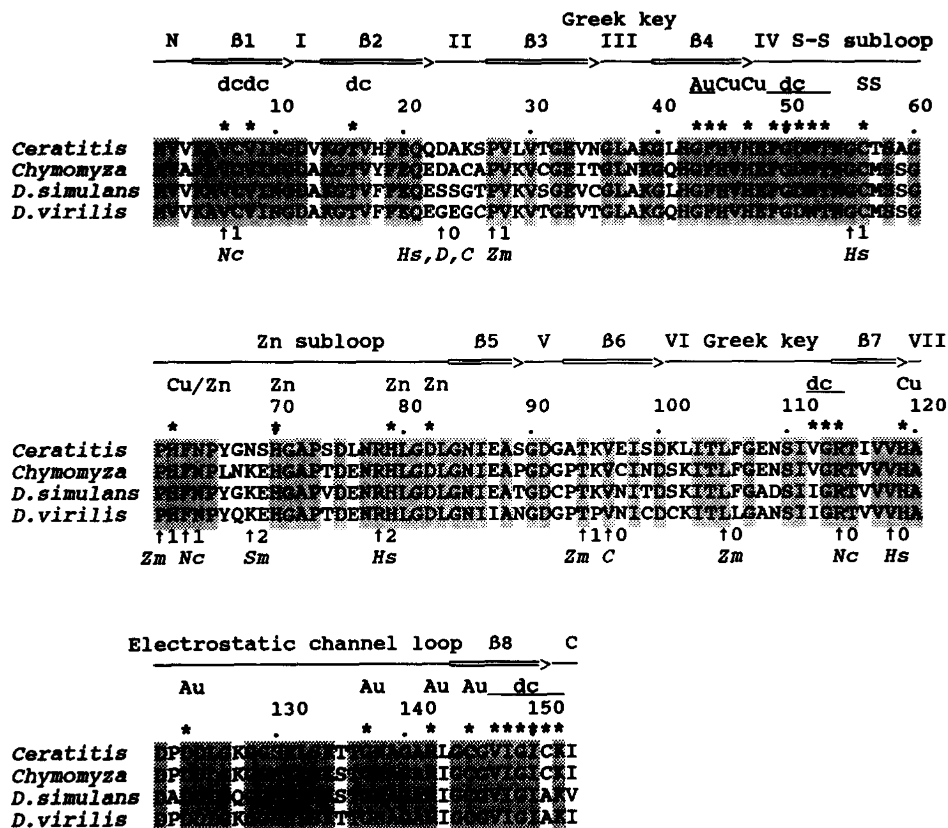


FIG. 3. Amino acid sequence alignment of the Cu,Zn SOD polypeptide in four dipteran species: *C. capitata*, *Chymomyza amoena*, *D. simulans*, and *D. virilis*. Structural features of the protein are represented above the sequences: arrows represent the eight β strands; lines represent the random-coil loops (I–VII). Residues with significant functions are marked with asterisks and their functions indicated by two-letter symbols as follows: active site metal-ligand (Cu or/and Zn); active site auxiliary function (Au): disulfide bridge (SS); dimer contact (dc) (Getzoff *et al.*, 1983). The symbols below the sequences refer to the introns: position (small arrows), site within the coding triplet (numbers), and species, humans (*Hs*), fluke (*Sm*), maize (*Zm*), *Neurospora* (*Nc*), *Drosophila* species (*D*), other dipterans (*C*). See text for sources.

A + T content in the introns than in the surrounding coding sequences (Csank *et al.*, 1990), a difference that has been shown to be important for splicing in plants (Goodall and Filipowicz, 1989) and might have similar significance in dipterans.

The C + G content of the *Sod* coding region is lower in *C. capitata* (47%) than in *D. melanogaster* (60%), which exhibits a strong bias toward codons ending with these bases (74.7% versus 41.6% in *Ceratitis*). The bias toward excess C + G use in the third position is also present in other *melanogaster* genes: the average is 63.3% in 412 genes (calculated from data in GenBank 63, produced by J. Michael Cherry of Harvard University). In four other *Ceratitis* genes for which the sequence is known (two chorion and two vitellogenin genes), the average C + G third-position content is 46.6% (calculated from data in Rina and Savakis, 1991). The use of codons ending with G drops from 46% in *D. melanogaster Sod* to about 15% in any of the *Ceratitis* genes. It can be seen in Table 1 that codon preferences for the *Sod* gene are quite different in *D. melanogaster* and in *Ceratitis*, in which they are much more similar to the codon preferences found in either the chorion or the vitellogenin genes (Konsolaki *et al.*, 1990; Rina and Savakis, 1991).

*Evolution of Cu,Zn Sod.* Figure 3 displays the amino acid sequence of the Cu,Zn SOD polypeptide in four dipteran species, inferred from the nucleotide sequence of the corresponding *Sod* genes. (The polypeptide sequences of *D. simulans* and *D. melanogaster* are identical and hence only one of these two is shown.)

**TABLE 1**

**Codon Usage in the *Cu,Zn Sod* and the *s36* Chorion gene of *C. capitata* and the *Cu,Zn Sod* of *D. melanogaster***

| Amino acid | Codon | *Ceratitis Sod* Number | *Ceratitis Sod* Frequency | *Ceratitis s36*[a] | *Drosophila Sod*[b] | Amino acid | Codon | *Ceratitis Sod* Number | *Ceratitis Sod* Frequency | *Ceratitis s36*[a] | *Drosophila Sod*[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gly | GGG | 2 | 0.08 | 0.00 | 0.04 | Trp | TGG | 0 | 0.00 | 0.00 | 0.00 |
| Gly | GGA | 7 | 0.28 | 0.07 | 0.16 | End | TGA | 0 | 0.00 | 0.00 | 0.00 |
| Gly | GGT | 9 | 0.36 | 0.52 | 0.24 | Cys | TGT | 1 | 0.25 | 1.00 | 0.00 |
| Gly | GGC | 7 | 0.28 | 0.41 | 0.56 | Cys | TGC | 3 | 0.75 | 0.00 | 1.00 |
| Glu | GAG | 1 | 0.14 | 0.13 | 0.88 | End | TAG | 0 | 0.00 | 0.00 | 0.00 |
| Glu | GAA | 6 | 0.86 | 0.88 | 0.13 | End | TAA | 1 | 1.00 | 1.00 | 1.00 |
| Asp | GAT | 7 | 0.70 | 0.80 | 0.50 | Tyr | TAT | 1 | 1.00 | 0.48 | 1.00 |
| Asp | GAC | 3 | 0.30 | 0.20 | 0.50 | Tyr | TAC | 0 | 0.00 | 0.52 | 0.00 |
| Val | GTG | 1 | 0.07 | 0.30 | 0.31 | Leu | TTG | 4 | 0.36 | 0.42 | 0.00 |
| Val | GTA | 5 | 0.33 | 0.22 | 0.06 | Leu | TTA | 2 | 0.18 | 0.11 | 0.00 |
| Val | GTT | 5 | 0.33 | 0.30 | 0.25 | Phe | TTT | 2 | 0.40 | 0.25 | 0.00 |
| Val | GTC | 4 | 0.27 | 0.17 | 0.38 | Phe | TTC | 3 | 0.60 | 0.75 | 1.00 |
| Ala | GCG | 0 | 0.00 | 0.07 | 0.00 | Ser | TCG | 2 | 0.25 | 0.11 | 0.11 |
| Ala | GCA | 3 | 0.30 | 0.20 | 0.00 | Ser | TCA | 0 | 0.00 | 0.18 | 0.11 |
| Ala | GCT | 6 | 0.60 | 0.59 | 0.27 | Ser | TCT | 1 | 0.13 | 0.36 | 0.00 |
| Ala | GCC | 1 | 0.10 | 0.13 | 0.73 | Ser | TCC | 1 | 0.13 | 0.14 | 0.33 |
| Arg | AGG | 0 | 0.00 | 0.00 | 0.00 | Arg | CGG | 0 | 0.00 | 0.00 | 0.00 |
| Arg | AGA | 0 | 0.00 | 0.00 | 0.00 | Arg | CGA | 0 | 0.00 | 0.00 | 0.00 |
| Ser | AGT | 1 | 0.13 | 0.07 | 0.00 | Arg | CGT | 2 | 0.67 | 0.25 | 0.33 |
| Ser | AGC | 3 | 0.38 | 0.14 | 0.44 | Arg | CGC | 1 | 0.33 | 0.75 | 0.67 |
| Lys | AAG | 7 | 0.78 | 0.60 | 0.89 | Gln | CAG | 2 | 1.00 | 0.33 | 1.00 |
| Lys | AAA | 2 | 0.22 | 0.40 | 0.11 | Gln | CAA | 0 | 0.00 | 0.67 | 0.00 |
| Asn | AAT | 3 | 0.30 | 0.33 | 0.38 | His | CAT | 7 | 0.78 | 1.00 | 0.13 |
| Asn | AAC | 7 | 0.70 | 0.67 | 0.63 | His | CAC | 2 | 0.22 | 0.00 | 0.88 |
| Met | ATG | 1 | 1.00 | 1.00 | 1.00 | Leu | CTG | 2 | 0.18 | 0.05 | 0.71 |
| Ile | ATA | 0 | 0.00 | 0.06 | 0.00 | Leu | CTA | 1 | 0.09 | 0.05 | 0.00 |
| Ile | ATT | 7 | 0.78 | 0.31 | 0.56 | Leu | CTT | 2 | 0.18 | 0.05 | 0.14 |
| Ile | ATC | 2 | 0.22 | 0.63 | 0.44 | Leu | CTC | 0 | 0.00 | 0.32 | 0.14 |
| Thr | ACG | 1 | 0.11 | 0.14 | 0.44 | Pro | CCG | 0 | 0.00 | 0.03 | 0.40 |
| Thr | ACA | 1 | 0.11 | 0.43 | 0.00 | Pro | CCA | 1 | 0.20 | 0.47 | 0.00 |
| Thr | ACT | 4 | 0.44 | 0.29 | 0.00 | Pro | CCT | 3 | 0.60 | 0.33 | 0.00 |
| Thr | ACC | 3 | 0.33 | 0.14 | 0.56 | Pro | CCC | 1 | 0.20 | 0.17 | 0.60 |

*Note.* For each codon, the number and frequency of use are given for the *Ceratitis Sod* gene; only the frequency is given for the other two genes.

[a] Data from Konsolaki *et al.* (1990).

[b] Data from Kwiatowski *et al.* (1989b).

The coding region translates in all species into 153 amino acids, which makes the sequence alignment obvious. The first and last amino acids are, however, removed from the mature protein in *D. melanogaster*, the amino acid sequence of which was obtained in our laboratory (Lee *et al.*, 1985). It seems likely that the final amino acid, even though post-translationally removed, is functionally significant since it is evolutionarily conserved (isoleucine is present in three species and has been replaced by the functionally equivalent valine in *D. simulans* and *D. melanogaster*).

Identical residues are present in all five species at 111 (72.5%) of the 153 sites. The conserved amino acids include all residues engaged in catalytic activity and all but two of the residues involved in maintaining protein structure. The two exceptions are residues (at sites 112 and 151) that have been postulated in mammals to be involved in dimer formation (Getzoff *et al.*, 1989). Table 2 gives the number of amino acid differences (above the diagonal) between the sequences of the five dipterans. The number of nucleotide differences between the coding sequences is also given (below the diagonal). The results are summarized in Table 3, where the average values have been obtained for comparisons between increasingly more inclusive sister lineages. The effect of superimposed and back mutations is corrected by the PAM method (Dayhoff, 1978, p. 375). The corrected values have been normalized relative to the *D.virilis*/*melanogaster*+*simulans* values.

Comparison between amino acid and nucleotide normalized values suggests that at the nucleotide level the *Chymomyza* lineage has evolved faster (or, alternatively, at the amino acid level slower) than the *Ceratitis* or the *Drosophila* lineages.

TABLE 2

**Amino Acid (above Diagonal) and Nucelotide (below) Differences between Three *Drosophila* and Two Other Fruit Fly Species**

| Species | *Ceratitis* | *Chymomyza* | *virilis* | *simulans* | *melanogaster* |
|---|---|---|---|---|---|
| *Ceratitis capitata* | — | 29 | 31 | 33 | 33 |
| *Chymomyza amoena* | 121 | — | 22 | 25 | 25 |
| *D. virilis* | 133 | 124 | — | 20 | 20 |
| *D. simulans* | 133 | 122 | 88 | — | 0 |
| *D. melanogaster* | 131 | 121 | 86 | 11 | — |

*Note.* Residues compared: 153 amino acids, 462 nucleotides.

## DISCUSSION

*Intron evolution.* The origin and evolution of introns have been the subjects of much debate. Although there is evidence that group I as well as group II introns can integrate into genes (Lambowitz, 1989), the prevailing view is that introns have existed from the start of protein evolution, and even that they may derive from the early "RNA world" that preceded the appearance of DNA organisms (Doolittle, 1987; Gilbert, 1987). The discovery that group I introns are shared by chloroplasts and cyanobacteria (Kuhsel *et al.*, 1990; Xu *et al.*, 1990) supports the existence of introns at least by the time of the eukaryote/prokaryote split. The early occurrence of introns is also supported by the intron distribution of some genes, such as triosephosphate isomerase (Gilbert *et al.*, 1986). Nevertheless, the distribution of introns in tubulin and actin genes, for example (Dibb and Newman, 1989), indicates that introns have not only been evolutionarily deleted but also inserted into coding sequences.

The evolution of *Cu,Zn Sod* manifests the occurrence of both insertion and deletion of introns at different positions in the coding sequence. Figure 3 indicates (arrows below the sequence) the positions of introns in five distantly related organisms: *Homo sapiens* (*Hs;* data of Levanon *et al.*, 1985), dipterans (*Drosophila* species, *D;* Kwiatowski *et al.*, 1989a,b; Kwiatowski and Ayala, 1989; *Ceratitis capitata* and *Chymomyza amoena, C;* this paper and Kwiatowski *et al.*, 1992), *Schistosoma mansoni* (*Sm;* Simurda *et al.*, 1988); *Zea*

TABLE 3

**Differentiation between the *Sod* Sequences of Organisms with Increasingly Remote Common Ancestors**

| Comparison | Amino acids[a] | | | Nucleotides[a] | | | Million years since divergence[d] |
|---|---|---|---|---|---|---|---|
| | Observed | PAM[b] | Normalized[c] | Observed | PAM | Normalized[c] | |
| *melanogaster*/*simulans* | 0 | — | — | 2.4 | 2.4 | 0.11 | — |
| *virilis*/*melanogaster* + *simulans* | 13.1 | 14.3 | 1.00 | 18.8 | 21.6 | 1.00 | 47.7 |
| *Chymomyza*/*Drosophila* | 15.7 | 17.5 | 1.22 | 26.5 | 32.4 | 1.50 | 58.3 |
| *Ceratitis*/all other | 20.6 | 24.0 | 1.68 | 28.0 | 34.8 | 1.61 | 80.0 |

[a] Averages per 100 residues.
[b] Corrected values calculated using Tables 36 and 37 of Dayhoff *et al.* (1978).
[c] PAM values normalized to the *melanogaster*/*virilis* value.
[d] Assuming a rate of 30 amino acid replacements per 100 residues per 100 million years.

*mays* (*Zm;* R. Cannon, personal communication), and *Neurospora crassa* (*Nc;* Chary *et al.*, 1990).

Evidence for intron insertion derives from the observation that all 13 intron sites are unique to each type of organism, except for the site between amino acids 22 and 23, which is shared by humans and dipterans. In order to avoid postulating that intron insertions have occurred, one would have to postulate an ancestral gene with numerous introns, which seems extremely unlikely on the face of the evidence at hand (plus the likelihood that more intron sites will yet be discovered as the variety of organisms studied increases). One must, of course, consider the possibility that some intron sites are homologous but have changed their position in the amino acid sequence through evolution. Some intron sites are closely located: *Neurospora* and maize at 61 and 63; *Ceratitis* and maize at 94 and 96. But it seems unlikely that these are homologous, i.e., that they are derived from a preexisting intron located at a particular position accompanied by the insertion/deletion in one or both organisms of one (or two) amino acid on either side of the intron. The evolutionary conservation of an intron site would seem particularly unlikely when it involves a reading-frame shift, as in the case of residue 95, flanked by dipteran and maize introns. Two evolutionarily independent intron insertions in the two lineages, or one deletion followed by one insertion, in one of the lineages, would seem more probable. In any case, intron insertions must have occurred at sites that are far removed from each other, unless we postulate the presence of very many introns in the ancestral gene.

The data presented in this paper support the evolutionary deletion of an intron site in the *Sod* gene. Figure 4 outlines the phylogeny of the three dipteran genera *Ceratitis, Chymomyza,* and *Drosophila.* All the 30 or more *Sod Drosophila* genes sequenced in our laboratory have only one intron, always between codons 22 and 23. The sequences published include *D. virilis* (Kwiatowski and Ayala, 1989), which belongs to the *Drosophila* subgenus, and *D. melanogaster* and *simulans* (Kwiatowski *et al.*, 1989a,b), which belong to the other major subgenus, *Sophophora.* All *Drosophila* species lack the second intron, between codons 95 and 96, found in the two other dipteran genera. The most parsimonious explanation of this state of affairs is that both introns were present in a common dipteran ancestor, but the second intron was deleted in the *Drosophila* lineage, shortly after its divergence from the *Chymomyza* lineage and, in any case, before the split of the two subgenera, *Sophophora* and *Drosophila.*

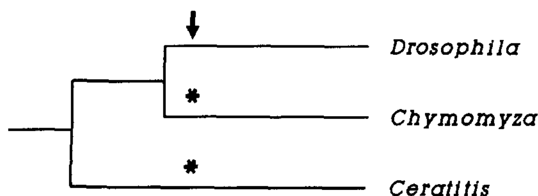


FIG. 4. Phylogeny of three dipteran genera. The absence of the *Cu,Zn Sod* second intron from *Drosophila* requires either a single deletion in the *Drosophila* lineage (↓) or two separate insertions, one in each of the other two lingeages (*).

An alternative explanation requires two evolutionary events: the insertion of an intron independently in the *Ceratitis* and the *Chymomyza* lineages. That this insertion would occur independently at precisely the same site, between residues 95 and 96, would seem *prima facie* unlikely. Introns, however, tend to be inserted in "proto-splice" sites, which persist in the DNA sequence as a consequence of functional constraints in the primary structure of the protein (Dibb and Newman, 1989). As shown in Fig. 5, the tetranucleotide AGGT, which is a possible proto-splice site, is preserved in some *Drosophila* species lacking the second intron. As a consequence of the 5′-GT. . .AG-3′ rule of intron flanking (Breathnach and Chambon, 1981), the AGGT tetranucleotide sequence becomes repeated at the intron/exon boundaries, which makes the repeated evolutionary insertion of an intron at this particular site plausible. The issue can be settled by sequencing the *Sod* gene in dipteran species from lineages that are outgroups of those represented in Fig. 4.

The fickleness of the intron/exon structure of the *Cu,Zn Sod* gene over evolutionary history (Fig. 3) contrasts with the enduring conservation of the number and location of introns in other genes, such as those coding for globin proteins (Hardison, 1991). Two introns located at equivalent positions occur in all functional vertebrate genes, although the original duplication from which derive the α-like and β-like hemoglobin genes occurred about 450 million years ago, and the hemoglobin/myoglobin duplication is at least 500 my old. Both introns have been deleted in the midge *Chironomus tummi* (Antoine and Niessing, 1984; see below), but they are present and at the same splice locations in the gene for plant globins (Brisson and Verma, 1982), which diverged from the other globins at the time of the plant/animal ancestral split, perhaps one billion years ago. The plant globin genes have, however, a third intron that splits the animal middle exon into two (Jensen *et al.*, 1981; Landsmann *et al.*, 1986; Bogusz *et al.*, 1988), and which is postulated to have existed in the ancestral globin gene but became lost in animal evolution (Gō, 1981).

```
Ceratitis         ...agcaacgaaggt...aggtcgaaatca...
Chymomyza         ...tcccacaaaggt...aggtgtgcatta...
D. melanogaster   ...ccccaccaag---------gtcaacatca...
D. simulans       ...ccccacaaag---------gtcaacatca...
D. virilis        ...tcccactccc---------gtgaacatct...
```

FIG. 5. Nucleotide sequence alignment at the boundaries of the second intron of the *Cu,Zn Sod* gene.

The intron/exon structure observed in *Ceratitis* is rather typical of dipteran genomes. Usually, as noted above, whenever more than one intron occurs in a gene, the first one from the 5′ end is longer: in *C. capitata* the lengths of the two introns are 665 and 79 bp. In contrast to the vertebrates, which typically have large genomes and genes punctuated by large introns, dipterans have compact genomes with few and small introns (Hawkins, 1988). In view of these observations, and of the evidence herein presented favoring the evolutionary deletion of the second intron in the *Drosophila* lineage, we suggest that there is a general trend in the evolution of the Diptera favoring intron size reduction and even elimination. This trend toward reduction and elimination of introns parallels the one postulated to have led long ago to the total elimination of introns in the eubacteria, and might be similarly motivated by a tendency toward the reduction of any unnecessary DNA load.

Comparison of the two very closely related (and morphologically nearly indistinguishable) species *D. melanogaster* and *D. simulans* indicates that nucleotide deletion/insertion is much more prevalent than substitution in the intron evolution of *Cu,Zn Sod* (Kwiatowski *et al.*, 1989a). Moreover, there is a highly conserved 26-bp-long stretch that starts 60 bp downstream from the 5′ end of the intron. The incidence of nucleotide substitutions between *D. melanogaster* and *D. virilis*, which diverged some 50 my ago, is only 12% within this stretch, whereas it is 19% in the coding sequence (Kwiatowski and Ayala, 1989). A trace of the conservation of this stretch can be detected in *C. capitata* (sites 321–345 in Fig. 2; see Kwiatowski *et al.*, 1992). This stretch is likely to have regulatory or some other function. Regulatory sequences have been identified in dipteran genes within the introns, most often within the first and longer one (Gasch *et al.*, 1989; Bruhat *et al.*, 1990).

*Protein structure and function vs sequence evolution.* The structure of the Cu,Zn SOD protein has been ascertained in mammals (Tainer *et al.*, 1982, 1983; Getzoff *et al.*, 1989). Figure 3 shows the relevant features, superimposed on the amino acid sequence. The polypeptide can be divided into eight regions made up of eight anti-parallel β-sheet structures that constitute the backbone of the molecule and are connected by seven hinge random-coil segments. Two of the random coils form the active site region: they bind the two metals copper and zinc and create the electrostatic channel that attracts the substrate molecules (Tainer *et al.*, 1982; Getzoff *et al.*, 1983).

The variable positions in the dipteran SODs are mostly located in and around hinge regions without known functional significance. A good example is hinge region II, where at each of the four sites (23–26) there occur polar and nonpolar amino acids (as well as charged ones at two positions, 23 and 25) in just the four sequences given in Fig. 3. The two fairly variable regions of the molecule (sites 22–41 and 87–112) are defined by residues forming two antiparallel β strands (β3 and β6) that are adjacent in the three-dimensional structure, plus the four loops flanking them (II, III, V, and VI). These regions, which are obviously variable already in the dipteran sequences, are also conspicuously variable when other species, including mammals, fungi, and plants, are also compared (Lee *et al.*, 1985; Ayala, 1986; Kwiatowski *et al.*, 1991a). No functionally significant residues are found in these two regions, with the possible exception of leucine 105, which stabilizes the β-barrel structure (Getzoff *et al.*, 1989). One of the two dipteran introns is located in each of these two regions. The three largely conserved regions of the SOD polypeptide are made up of the remaining β strands (β1, β2, β4, β5, β7, and β8), which contain various functionally significant residues, plus the two large loops that surround the active site (Tainer *et al.*, 1982).

On the three-dimensional structure, the variable portions of the molecule mentioned above, made up of β-hairpin and random coil secondary-structure elements, appear as a patch on the surface of the protein—a patch devoid of any function other than contributing to the structural backbone of the protein. On the contrary, the remainder of the protein sequence is engaged in essential catalytic or structural function. Thus, the position of variable residues is not dependent on the secondary structure of the protein per se, but is rather determined by which residues do or do not contribute to the enzymatic activity or structural integrity of the molecule, which is in turn precisely determined by the three-dimensional structure of the protein.

A similar state of affairs occurs in cytochrome c, one of the best studied proteins from both the evolutionary and the structural perspective. An analysis of 60 eukaryotic sequences (Dickerson and Timkovich, 1975) evinces the presence of some "hypervariable" regions including a total of 11 sites (out of 105), at each of which 8–10 different amino acids are found at any one position among the 60 species. Although some hypervariable sites are located proximal to each other on the primary sequence, the 11 sites are spread throughout the sequence (positions 3, 54, 58, 60, 62, 65, 88, 89, 92, 100, and 103; see Table 5 and pp. 423–429 in Dickerson and Timkovich, 1975). On the three-dimensional structure, these sites lie mostly on the surface in the back side of the molecule. Margoliash *et al.* (1972) found that when the three-dimensional location of all variable (not just the hypervariable) positions in cytochrome c is examined within a group of related organisms, such as mammals, insects, or fungi, the variable sites cluster as patches on the surface of the molecule at locations that vary from group to group. In the mam-

mals, for example, the variant residues occur on the surface in a band in front, above, and behind the heme. Although individual cytochromes may differ in as many as 63 amino acids (Dickerson and Timkovich, 1975, p. 411), the set of variable sites becomes disproportionally smaller as more and more closely related organisms are examined. Fitch and Markowitz (1970) have developed the concept of "concomitantly variable codons" (or "covarions") to refer to the inferred number of residues that can accept replacements in any one species, which in the case of cytochrome c is estimated as 10 (see also Margoliash *et al.*, 1972; Dickerson and Timkovich, 1975, pp. 449–450). The concept of covarions may help in understanding the apparently erratic rate of evolution of SOD (Lee *et al.*, 1985; Ayala, 1986; Kwiatowski *et al.*, 1991a). In terms of the covarion hypothesis, the rate of SOD evolution within a group of organisms, such as the mammals or the Drosophilidae, is high (30 replacements per 100 amino acids per 100 million years) because there are a relatively large number of covarions (about 20% of all sites), but when organisms from different groups are compared, such as mammals with Drosophilidae, the rate becomes much lower (less than 10 replacements per 100 amino acids per 100 million years; see Lee *et al.*, 1985; Ayala, 1986) because of the constraints imposed by the functional and structural requirements of the SOD protein.

*Rate of SOD evolution in dipterans and mammals.* Lee *et al.* (1985) and Ayala (1986) observed an apparent acceleration in the rate of evolution of Cu,Zn SOD in recent geological time. The observed rate of amino acid divergence (using PAM values to correct for superimposed and back replacements) was estimated at about 30 per 100 residues per 100 million years when comparisons were made between species diverged within the last 100 my (Lee *et al.*; Ayala, 1986), a rate five times greater than the one observed when comparisons were made between animals and fungi, which diverged at least 10 times earlier. The discrepancy could possibly be due to a greatly increased rate of evolution in the mammals since only mammal data were available for comparisons between recently diverged species.

Table 3 shows the estimated divergence time between the dipteran lineages using the mammal rate of evolution (30 amino acid replacements per 100 residues per 100 my). The estimated times of divergence (80, 58, and 48 my) for the successive splits of the *Ceratitis, Chymomyza,* and *virilis* lineages relative to those of the *melanogaster/simulans* lineage are reasonably similar to those estimated on the basis of other data. Collier and MacIntyre (1977), for example, give 52 my for the *virilis/melanogaster* divergence on the basis of immunological data, and Throckmorton (1975) argues on the basis of biogeographic considerations that that split occurred more than 36 my ago. We, therefore, conclude that the rate of evolution of SOD is approximately the same in mammals and in dipterans (see also Kwiatowski *et al.*, 1991a).

## ACKNOWLEDGMENTS

We thank Drs. G. Tener for the *Drosophila melanogaster Cu,Zn Sod* cDNA clone; D. King for the *C. capitata* genomic library; R. Cannon for information on intron distribution in maize SOD prior to publication; and J. Felsenstein, G. D. F. Wilson, and R. Tyler for computer software.

## REFERENCES

Antoine, M., and Niessing, J. (1984). Intron-less globin genes in the insect *Chironomus thummi thummi. Nature* **310:** 795–798.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, G., and Struhl, K. (1987). "Current Protocols in Molecular Biology," Wiley, New York.

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

Bannister, J. V., and Parker, M. W. (1985). The presence of a copper/zinc superoxide dismutase in the bacterium *Photobacterium leiognathi:* A likely case of gene transfer from eukaryotes to prokaryotes. *Proc. Natl. Acad. Sci. USA* **82:** 149–152.

Blackman, R. K., and Meselson, M. (1986). Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila hsp82* gene. *J. Mol. Biol.* **188:** 499–515.

Bogusz, D., Appleby, C. A., Landsmann, J., Dennis, E. S., Trinick, M. J., and Peacock, W. J. (1988). Functioning haemoglobin genes in non-nodulating plants. *Nature* **331:** 178–183.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349–383.

Brisson, N., and Verma, D. P. S. (1982). Soybean leghemoglobin gene family: Normal, pseudo, and truncated genes. *Proc. Natl. Acad. Sci. USA* **79:** 4055–4059.

Bruhat, A., Tourmente, S., Chapel, S., Sobrier, M. L., Couderc, J. L., and Dastugue, B. (1990). Regulatory elements in the first intron contribute to transcriptional regulation of the beta 3 tubulin gene by 20-hydroxyecdysone in *Drosophila* Kc cells. *Nucleic Acids Res.* **18:** 2861–2867.

Chary, P., Hallewell, R. A., and Natvig, D. O. (1990). Structure, exon pattern, and chromosome mapping of the gene for cytosolic copper-zinc superoxide dismutase (*sod-1*) from *Neurospora crassa. J. Biol. Chem.* **265:** 18,961–18,967.

Cherbas, L., Schulz, R. A., Koehler, M. M. D., Savakis, C., and Cherbas, P. (1986). Structure of the *Eip28/29* gene, and Ecdysone-inducible gene from *Drosophila. J. Mol. Biol.* **189:** 617–631.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

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

Csank, C., Taylor, F. M., and Martindale, D. W. (1990). Nuclear pre-mRNA introns: Analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes. *Nucleic Acids Res.* **18:** 5133–5141.

Dayhoff, M. D. (1978). "Atlas of Protein Sequence and Structure," National Biomedical Research Foundation, Washington, DC.

Devereu, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Dibb, N. J., and Newman, A. J. (1989). Evidence that introns arose at proto-splice sites. *EMBO J.* **8:** 2015–2021.

Dickerson, R. E., and Timkovich, R. (1975). Cytochromes c. *In* "The Enzymes" (P. D. Boyer, Ed.), Vol. XI, 3rd ed., pp. 397–547, Academic Press, New York.

Doolittle, W. F. (1987). What introns have to tell us: Hierarchy in genome evolution. *Cold Spring Harbor Symp. Quant. Biol.* **52:** 907–913.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. W., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. *Cell* **21:** 653–668.

Felsenstein, J. (1989). "PHYLIP 3.2 Manual," University of California Herbarium, Berkeley.

Fernandez-Sousa, J. M., and Michelson, A. M. (1976). Variation of superoxide dismutase during development of the fruit fly *Ceratitis capitata*. *Biochem. Biophys. Res. Commun.* **73:** 217–223.

Fitch, W. M., and Markowitz, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixations of mutations in evolution. *Biochem. Genet.* **4:** 579–593.

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

Gasch, A., Hinz, U., and Renkawitz-Pohl, R. (1989). Intron and upstream sequences regulate expression of the *Drosophila b*3-tubulin gene in the visceral and somatic musculature, respectively. *Proc. Natl. Acad. Sci. USA* **86:** 3215–3218.

Getzoff, E. D., Tainer, J. A., Stempien, M. M., Bell, G. I., and Hallewell, R. A. (1989). Evolution of CuZn superoxide dismutase and the Greek key beta-barel structural motif. *Proteins* **5:** 322–336.

Getzoff, E. D., Tainer, J. A., Weiner, P. K., Kollman, P. A., Richardson, J. S., and Richardson, D. C. (1983). Electrostatic recognition between superoxide and copper, zinc superoxide dismutase. *Nature* **306:** 287–290.

Gilbert, W. (1987). The exon theory of genes. *Cold Spring Harbor Symp. Quant. Biol.* **52:** 901–905.

Gilbert, W., Marchionni, M., and McKnight, G. (1986). On the antiquity of introns. *Cell* **46:** 151–154.

Gō, M. (1981). Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* **291:** 90–92.

Goodall, G. J., and Filipowicz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58:** 473–483.

Hardison, R. C. (1991). Evolution of Globin Gene Families. *In* "Evolution at the Molecular Level" (R. K. Selander, A. G. Clark, and T. S. Whittam, Eds.), pp. 272–289, Sinauer, Sunderland, MA.

Hawkins, J. D. (1988). A survey on intron and exon lengths. *Nucleic Acids Res.* **16:** 9893–9908.

Jensen, E. O., Paludan, K., Hyldig-Nielsen, J. J. H., Jorgensen, P., and Marcker, K. A. (1981). The structure of a chromosomal leghaemoglobin gene from soybean. *Nature* **291:** 677–679.

Konsolaki, M., Komitopoulou, K., Tolias, P. P., King, D. L., Swimmer, C., and Kafatos, F. C. (1990). The chorion genes of the medfly, *Ceratitis capitata,* I: Structural and regulatory conservation of the s36 gene relative to two Drosophila species. *Nucleic Acids Res.* **18:** 1731–1737.

Kuhsel, M. G., Strickland, R., and Palmer, J. K. (1990). An ancient group-I intron shared by eubacteria and chloroplasts. *Science* **250:** 1570–1573.

Kwiatowski, J. (1987). Comparison of chloroplast and cytosolic Cu/Zn superoxide dismutase isozymes from tomato in relation to superoxide dismutase evolution. *Isozymes: Curr. Top. Biol. Med. Res.* **15:** 121–139.

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

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

Kwiatowski, J., Hudson, R. R., and Ayala, F. J. (1991a). The rate of Cu,Zn Superoxide dismutase evolution. *Free Radical Res. Commun.* **12–13:** 363–370.

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

Kwiatowski, J., Skarecky, D., Burgos, M., and Ayala, F. J. (1992). 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:** in press.

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

Lambowitz, A. M. (1989). Infectious introns. *Cell* **56:** 323–326.

Landsmann, J., Dennis, E. S., Higgins, T. J. V., Appleby, C. A., Kortt, A. A., and Peacock, W. J. (1986). Common evolutionary origin of legume and non-legume plant haemoglobins. *Nature* **324:** 166–168.

Lee, Y. M., Friedman, D. J., and Ayala, F. J. (1985). Superoxide dismutase: An evolutionary puzzle. *Proc. Natl. Acad. Sci. USA* **82:** 824–828.

Levanon, D., Lieman-Hurwitz, J., Dafni, N., Wigderson, M., Sherman, L., Bernstein, Y., Laver-Rudich, Z., Danciger, E., Stein, O., and Groner, Y. (1985). Architecture and anatomy of the chromosomal locus in human chromosome 21 encoding the Cu/Zn superoxide dismutase. *EMBO J.* **4:** 77–84.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Margoliash, E., Fitch, W. M., Markowitz, E., and Dickerson, R. E. (1972). Functional limits of cytochrome c variability. *In* "Oxidation Reduction Enzymes" (A. Akeson, and A. Ehrenberg, Eds.), pp. 5–17, Pergamon Press, Oxford/New York.

Reed, K. C., and Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* **13:** 7207–7221.

Riley, M. (1989). Nucleotide sequence of the *Xdh* region in *Drosophila pseudoobscura* and an analysis of the evolution of synonymous codons. *Mol. Biol. Evol.* **6:** 33–52.

Rina, M., and Savakis, C. (1991). A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata:* Sequence and structural conservation in dipteran yolk proteins and their genes. *Genetics* **127:** 769–780.

Sanger, F., Miklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Seto, N. O. L., Hayashi, S., and Tener, G. M. (1987a). Cloning, sequence analysis and chromosomal localization of the Cu–Zn superoxide dismutase gene of *Drosophila melanogaster*. *Gene* **75:** 85–92.

Seto, N. O. L., Hayashi, S., and Tener, G. M. (1987b). Drosophila Cu–Zn superoxide dismutase cDNA sequence. *Nucleic Acids Res.* **15:** 5483.

Simurda, M. C., van Keulen, H., Rekosh, D. M., and LoVerde,

P. T. (1988). *Schistosoma mansoni:* Identification and analysis of an mRNA and a gene encoding superoxide dismutase (Cu/Zn). *Exp. Parasitol.* **67:** 73–84.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Steinman, H. M. (1988). Bacterial superoxide dismutases. *Basic Life Sci.* **49:** 641–646.

Tabor, S., and Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84:** 4767–4771.

Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S., and Richardson, D. C. (1982). Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J. Mol. Biol.* **160:** 181–217.

Tainer, J. A., Getzoff, E. D., Richardson, J. S., and Richardson, D. C. (1983). Structure and mechanism of copper, zinc superoxide dismutase. *Nature* **306:** 284–287.

Throckmorton, L. H. (1975). The phylogeny, ecology and geography of Drosophila. *In* "Handbook of Genetics" (R. C. King, Ed.) pp. 421–469, Plenum, New York.

Xu, M.-Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A., and Shub, D. A. (1990). Bacterial origin of a chloroplast intron—Conserved self-splicing group-I introns in cyanobacteria. *Science* **250:** 1566–1570.