Insect Molecular Biology (2004) **13**(6), 587–594

Molecular phylogenetic evidence for an extracellular Cu Zn superoxide dismutase gene in insects

J. D. Parker, K. M. Parker and L. Keller

Department of Ecology and Evolution, University of Lausanne, Switzerland

Abstract

Representatives of three ancient gene families of the antioxidant enzyme superoxide dismutase (SOD) can be found in most metazoans. In mammals and *Caenorhabditis elegans*, there is at least one gene each of the cytoplasmic, mitochondrial and extracellular lineages of SOD genes. The cytoplasmic SOD was one of the first enzymes to be implicated in ageing due to its protection against damaging oxygen free radicals. In contrast to other metazoans, insects were thought to lack a gene for the extracellular SOD. We have cloned and sequenced an SOD mRNA in the ant *Lasius niger* that appears to belong to this extracellular family. Subsequent searches and analyses of SOD gene sequences in insect databases revealed that insects do indeed express all three SOD genes including the extracellular form. We conclude that insects as well as other metazoans appear to have the full repertoire of the three families of SOD.

Keywords: extracellular superoxide dismutase, EC SOD, SOD3, SOD5, phylogeny.

Introduction

The superoxide dismutases (SODs) are one of the first enzymatic defences against reactive oxygen species (ROS) which are primarily generated by aerobic metabolism. The SOD enzymes work by converting negatively charged oxygen molecules to hydrogen peroxide that is subsequently destroyed by catalase. Not only are ROS directly dangerous to cellular components (Fridovich, 1987), but evidence has also accumulated implicating the damaging effects of ROS as one of the central intermediaries of ageing (Perez-Campo *et al*., 1998; Finkel & Holbrook, 2000; Arking *et al*., 2002; Barja, 2002; Sohal *et al*., 2002; Hasty *et al*., 2003; Hekimi & Guarente, 2003). This important role of SOD as a protective enzyme results in it being present in both eukaryotic and prokaryotic organisms. Hence, it is not surprising that the major families of SOD genes evolved and diverged before the animal body plans (Smith & Doolittle, 1992; Bordo *et al*., 1994; Zelko *et al*., 2002).

There are three basic families of animal SOD enzymes with two possible metal configurations in the active sites using either Mn, or Cu and Zn ions (Bordo *et al*., 1994). The Mn SOD is restricted to the inner matrix of the mitochondria (Okado-Matsumoto & Fridovich, 2001) and is very different in structure from the two Cu Zn SODs. Of the two Cu Zn SODs, one gene product is primarily localized in cytoplasm (Crapo *et al*., 1992) with some also occurring in the mitochondrial intermembrane space (Okado-Matsumoto & Fridovich, 2001). This gene is usually referred to as the cytoplasmic Cu Zn SOD. The other Cu Zn SOD is predominantly extracellular although it has also been shown to be membrane bound (Marklund, 1984; Folz *et al*., 1997; Fujii *et al*., 1998). When this second SOD, termed extracellular Cu Zn SOD, is secreted rather than routed to the membrane, it can be reabsorbed and transported inside the nucleus where it protects genomic DNA (Ookawara *et al*., 2002) and slows telomere shortening (von Zglinicki, 2002; Serra *et al*., 2003). The extracellular Cu Zn SOD gene is distinguished from the cytoplasmic Cu Zn SOD gene by having an N-terminal signal cleavage peptide that routes the extracellular Cu Zn SOD for secretion (Tibell *et al*., 1996; Folz *et al*., 1997; Fujii *et al*., 1998).

In this study, we direct our attention to the cytoplasmic Cu Zn SOD and extracellular Cu Zn SOD. The Mn SOD will be designated as mitochondrial Mn SOD. The nomenclature of the SODs is confusing because of inconsistent numbering among species which was based upon the historical order in which the genes were described. These species specific numbers (i.e. SOD1, SOD2, … SOD5) will be denoted in parentheses here.

Received 21 April 2004; accepted following revision 4 June 2004. Correspondence: Joel D. Parker, Department of Ecology and Evolution, Biology Building, University of Lausanne, CH-1015, Lausanne, Switzerland. Tel.: +41 021692 41 93; fax: +41 021692 41 65; e-mail: joel.parker@i.e.-zea.unil.ch

In contrast to other metazoans, insects were generally thought to have only the mitochondrial and cytoplasmic SODs (Banks *et al*., 1995; Fujii *et al*., 1998; Kirby *et al*., 2002), although the basis for this conclusion has never been firmly established in any study (Fridovich, 1987; Bordo *et al*., 1994). To date, the Mediterranean fruit fly, *Ceratitis capitata*, is the only insect that has been found with more than one Cu Zn SOD (Banks *et al*., 1995). Although the extra *C. capitata* Cu Zn SOD contains all of the functional sites and general structure of cytoplasmic SOD, it is extremely divergent both at the nucleotide and protein sequence level from the cytoplasmic Cu Zn SOD in *C. capitata* and other insects. The additional *C. capitata* Cu Zn SOD also appears to lack the secretion signal cleavage peptide sequence that is characteristic of extracellular Cu Zn SOD. Hence, Banks *et al.* speculated that this could be either a new family of SOD or an ancient gene duplication (Banks *et al*., 1995), but nothing has been published on these hypotheses since.

While studying the mechanistic basis underlying the extreme queen lifespan in ants we cloned two expressed Cu Zn SOD genes in the ant *Lasius niger*. The first *L. niger* Cu Zn SOD was clearly homologous to the other insect cytoplasmic Cu Zn SOD genes (Parker *et al*., 2004). Here we show that this second Cu Zn SOD from *L. niger* is more divergent from other insect cytoplasmic Cu Zn SODs than from the second Cu Zn SOD that was discovered in *C. capitata*. However, unlike the second Cu Zn SOD in *C. capitata*, the second *L. niger* Cu Zn SOD appears to contain a signal peptide cleavage sequence much like the extracellular Cu Zn SODs. To ask whether the novel *L. niger* Cu Zn SOD (SOD3) is an undiscovered insect extracellular SOD, we undertook a search of the model insect sequence databases for other Cu Zn SODs to compare with the well characterized extracellular SODs of *Mus musculus* and *Caenorhabditis elegans* and to our new *L. niger* Cu Zn SOD. We found that insects do indeed have a second uncharacterized expressed extracellular SOD that clusters with the extracellular SOD in mammals and worms. Phylogenetic analyses further reveal that the previously discovered second Cu Zn SOD in *C. capitata* is a truncated homologue of this insect extracellular SOD that may have recently lost its signal peptide. Finally, a third Cu Zn SOD in *C. elegans* was found that likely resulted from a gene duplication of the cytoplasmic SOD. Thus, all animals, including insects, probably have conserved representatives of the extracellular, cytoplasmic, and mitochondrial lineages of SOD.

Results

We cloned a novel Cu Zn SOD from *Lasius niger* by rapid amplification of cDNA ends (RACE). Analysis of the 5′- and 3′ RACE products allowed us to determine that this Cu Zn SOD (SOD3) is coded by a 1028 base transcript (GenBank no. AY672457). A conceptual translation gave a protein sequence with all of the functionally important metal binding sites and the two Cu Zn SOD signature sequences (Fig. 1). The protein sequence only shared 45% identity with the other described *L. niger* cytoplasmic Cu Zn SOD (SOD1) (Parker *et al*., 2004). The novel Cu Zn SOD also has an additional twenty-one amino acid N-terminal peptide that appears to be an extracellular signal cleavage sequence (Fig. 1 and Table 1).

Database searches of insects revealed the presence of two groups of Cu Zn SODs in three insect model systems (*D. melanogaster*, *A. mellifera* and *A. gambiae*). All of the gene sequences shown in Fig. 1 are from cDNA with the exception of the *A. gambiae* sequence which was inferred from genomic sequence. At least one of each SOD group was annotated as the cytoplasmic SOD (SOD1 in all three cases) in their respective databases. The other Cu Zn SODs were computer annotated as possible Cu Zn SODs based on the presence of the Cu Zn SOD signature sequences. We call these (SOD3) in the figures and tables as they are the third SOD to be identified in these organisms (the mitochondrial SODs being the second). In all three cases (*D. melanogaster* SOD3, *A. mellifera* SOD3 and *A. gambiae* SOD3), the novel Cu Zn SOD has an additional extracellular signal cleavage peptide (Fig. 1 and Table 1).

The previously discovered second *C. capitata* Cu Zn SOD does not appear to have the N-terminal extracellular signal cleavage sequence. However, reanalysis of the originally published *C. capitata* mRNA sequence reveals a 5′ open reading frame continuous with the identified open reading frame that codes for eighteen amino acids ahead of the presumed start site (underlined italics in Fig. 1). This eighteen amino acid predicted peptide sequence is 50% identical to the signal cleavage peptide of *D. melanogaster*. When this additional sequence is included in the TargetP analysis, the second *C. capitata* Cu Zn SOD (SOD3) is predicted to be cleaved and extracellular like the others (Table 1).

In all three of our phylogenetic analyses, the cytoplasmic and extracellular SOD families resolved into two distinct lineages. The best tree recovered using the program MrBayes (using a Bayesian method to search for the best tree) is shown in Fig. 2. The other two phylogenetic methods (using quartet puzzling and star decomposition methods) produced similar trees supporting the two families.

The three analyses also yielded identical phylogenies within the cytoplasmic SOD clade. Surprisingly, the two *C. elegans* cytoplasmic SODs (SOD1 and SOD5) clustered with *M. musculus* (SOD1) rather than with the other invertebrates. Within the insect clade, *D. melanogaster* clustered with *C. capitata* and *A. mellifera* with *L. niger* (Fig. 2). The *A. gambiae* cytoplasmic Cu Zn SOD grouped with the *A. mellifera* and *L. niger* cytoplasmic Cu Zn SODs.

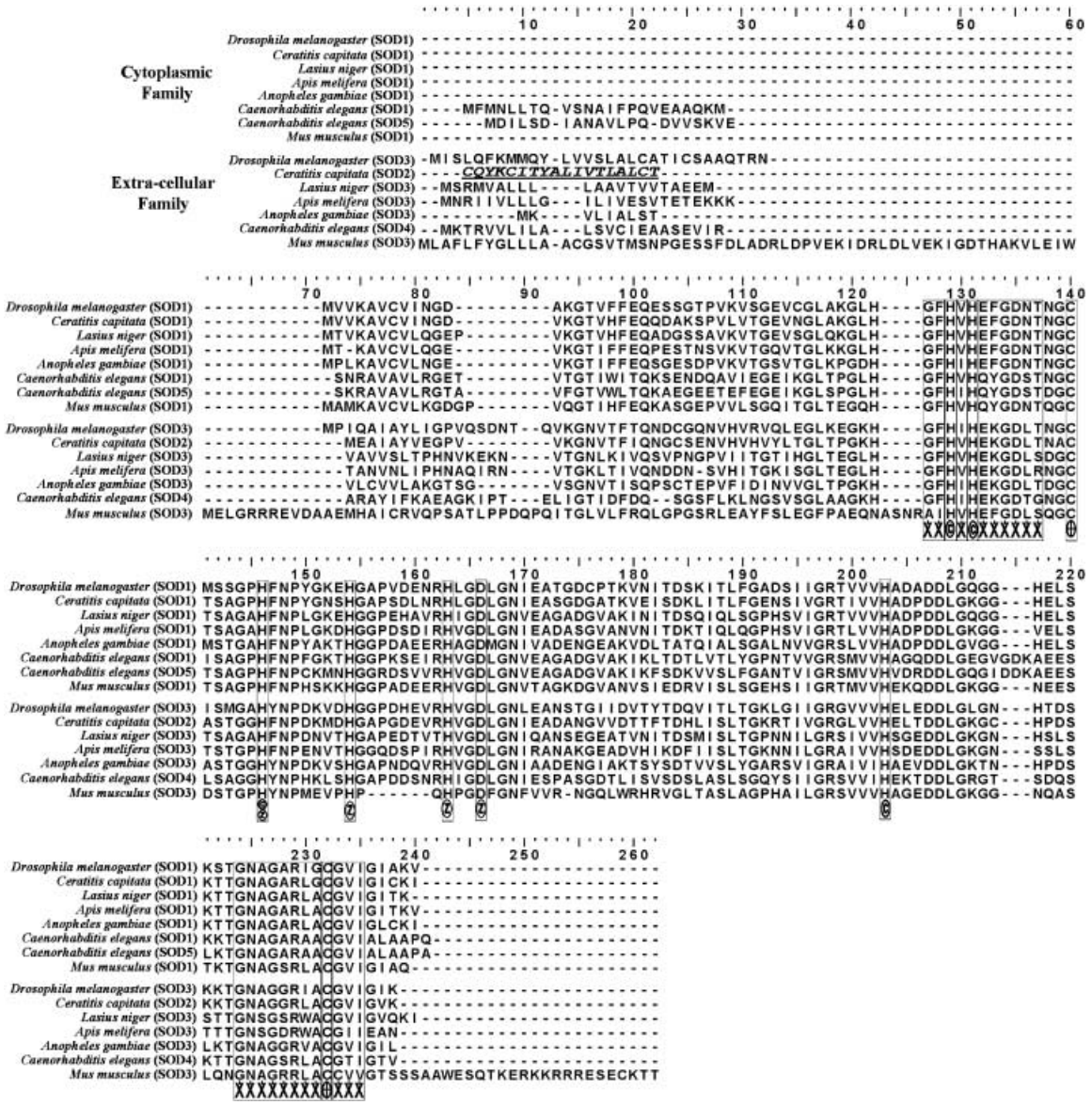

Figure 1. Alignment of the SOD gene sequences used in this study. Conserved functionally important sites are in blocks: circled C = Copper binding, circled Z = Zinc binding, circled C and circled Z = Copper and Zinc binding, and a circled plus sign = disulphide bond. The SOD signature sequences are denoted by a dotted X. The hypothetical leader sequence derived from the immediately upstream in frame translation for *Ceratitis capitata* (SOD2) is in underlined italics.

In the case of *C. elegans*, we unexpectedly found three Cu Zn SODs when only two were expected (one known cytoplasmic (SOD1) and one known extracellular Cu Zn SOD (SOD4)). The new *C. elegans* Cu Zn SOD (SOD5) is much more similar to the cytoplasmic SOD (SOD1) than to extracellular *C. elegans* SOD (SOD4) (Fujii *et al.*, 1998) (Figs 1 and 2). Like the new SODs in insects, the undescribed *C. elegans* Cu Zn SOD (SOD5) was computer annotated in the database as a likely Cu Zn SOD. The novel *C. elegans* Cu Zn SOD (SOD5) is most likely another cytoplasmic SOD based on its greater sequence similarity to the cytoplasmic Cu Zn SOD than to the extracellular SOD. Further supporting this conclusion, the two similar *C. elegans* Cu Zn SODs (SOD1 and SOD5) share two distinct and unique characteristics that strongly suggest that they arose from a gene duplication event. They both have extra N-terminal end peptide runs and share an insertion of three base pairs (Fig. 1). The functional significance of these extra peptide runs and insertions could not be determined by our analysis.

The only differences among the results of the three phylogenetic analyses are seen in the more divergent extracellular SOD family. There seems to be greater divergence within this family that makes it difficult to resolve the

Gene sequence	Database no.	Localization (reliability score)	Signal cleavage
Cytoplasmic SOD family			
Anopheles gambiae (SOD1)	XP311594	– (2)	None
Apis mellifera (SOD1)	AAP93581	– (2)	None
Caenorhabditis elegans (SOD1)	NP495431	– (4)	None
Caenorhabditis elegans (SOD5)	NP494779	– (2)	None
Ceratitis capitata (SOD1)	P28755	– (2)	None
Drosophila melanogaster (SOD1)	NP476735	– (2)	None
Lasius niger (SOD1)	AY309973	– (2)	None
Mus musculus (SOD1)	XM-128337	– (2)	None
Extracellular SOD family			
Anopheles gambiae (SOD3)	XP314137	Secreted (1)	1–16
Apis mellifera (SOD3)	Contig584	Secreted (2)	1–16
Caenorhabditis elegans (SOD4)	NP499091	Secreted (1)	1–17
Ceratitis capitata (SOD2)	S57193	– (1)	None
[plus leader sequence]		[Secreted (1)]	[1–23]
Drosophila melanogaster (SOD3)	AAL25378	Secreted (1)	1–25
Lasius niger (SOD3)	AY672457	Secreted (1)	1–18
Mus musculus (SOD3)	NP035565	Secreted (2)	1–20

Table 1. Predicted localization of the Cu Zn Sod proteins from this study. TargetP classifies the proteins as secreted, routed to mitochondria or other ('–' in the table) and predicts whether there is a signal cleavage sequence and where the cleavage will take place (Emanuelsson *et al.*, 2000). The reliability scores range from 1 to 5 with 1 being the most reliable

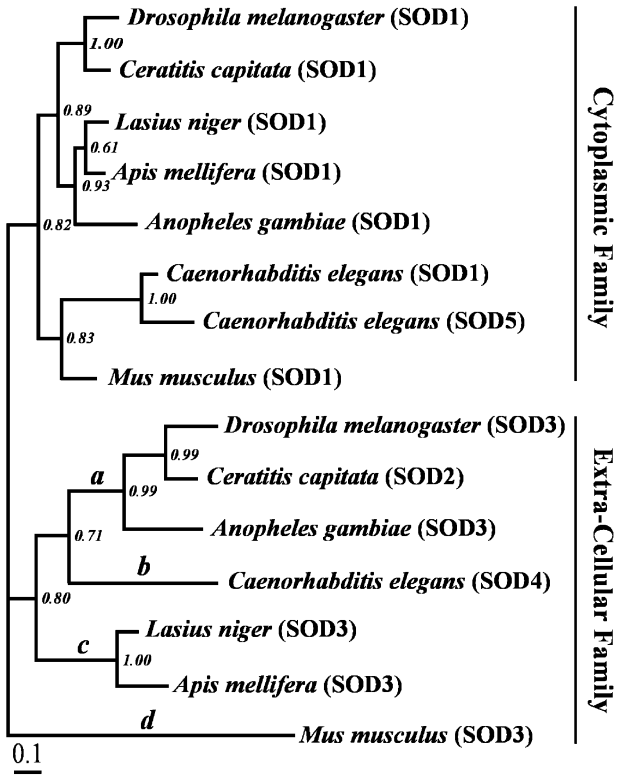

Figure 2. Majority rule consensus tree from the Bayesian analysis. Numbers in parentheses are clade credibility values which are equivalent to a posteriori probabilities. This was the most likely tree found (– ln 2764). The lowercase letters indicate the four supported extracellular branches described in the text.

phylogenetic relationships. All three methods agreed on the existence of four branches (a–d, in Fig. 2) and the topologies within those branches. However, the three phylogenetic methods differed in the relationship amongst these four branches. The best tree from the PAML (star decomposition) differed from the MrBayes (Bayesian analysis) presented in Fig. 2 by swapping the branch leading to *A. mellifera* and *L. niger* extracellular SOD (SOD3) and the *C. elegans* extracellular SOD (SOD4) branch (switching branches b and c in Fig. 2). The best tree from Tree-Puzzle (quartet puzzling) paired the *C. elegans* and *M. musculus* extracellular SODs (moving branch b on to branch d in Fig. 2). When trees from the three different programs were compared with TreePuzzle using the same model (quartet puzzling), the tree of Fig. 2 retained the highest maximum likelihood score although the other trees have very similar maximum likelihood scores (–ln 2764 for the tree in Fig. 2, –ln 2766 for the PAML tree and –ln 2769 for Tree-Puzzle). Thus, we can not draw any definitive conclusions on the relationships of these four clades (a–d) to one another based on this analysis.

Discussion

In the course of studying how some species of ant queens can live for nearly thirty years (Kutter & Stumper, 1969; Hölldobler & Wilson, 1990; Parker *et al.*, 2004), we discovered a second Cu Zn SOD in the ant *L. niger* that appears to be an insect homologue of the extracellular Cu Zn SOD found in non-insect animals. This new gene (SOD3) exhibits very little sequence homology to the *L. niger* cytoplasmic Cu Zn SOD (SOD1) and has an N-terminal signal cleavage sequence that is predicted to cause the protein to be secreted like the extracellular SODs in mammals and *C. elegans*. These two observations imply that we have discovered a previously uncharacterized insect extracellular SOD rather than a gene duplication of *Lasius niger* cytoplasmic SOD (SOD1). This discovery of extracellular Cu Zn SOD in insects contradicts the prevailing belief that insects possess only the cytoplasmic and mitochondrial SODs

(Banks *et al*., 1995; Fujii *et al*., 1998; Kirby *et al*., 2002). Furthermore, a thorough review of the literature revealed that any conclusion that insects did not have an extracellular SOD was based on the absence of any published description of insect extracellular SOD (Fridovich, 1987; Bordo *et al*., 1994), rather than on any specific study asking whether such a gene exists.

Phylogenetic analysis of Cu Zn SOD genes provides strong support for the existence of two distinct Cu Zn SOD families even when ignoring any contribution of the N-terminal sequence runs. We included Cu Zn SOD sequences from database searches of three insect genomes (*D. melanogaster*, *A. gambiae*, and *A. mellifera*), the previously identified two Cu Zn SODs in *C. capitata*, the two Cu Zn SODs in *L. niger*, and two control genomes (*C. elegans* and *Mus musculus*) which have well characterized cytoplasmic and extracellular Cu Zn SOD genes. One of the two families identified in our analysis includes the known cytoplasmic SODs of all species studied here while the other contains the new insect Cu Zn SODs identified here as well as the extracellular Cu Zn SODs of *M. musculus* and *C. elegans*. Thus, the two insect lineages appear to be representatives of the cytoplasmic and extracellular Cu Zn SOD lineages of metazoans. The deep separation of the cytoplasmic and extracellular lineages is also consistent with previous work showing an ancient origin of SOD gene families (Bordo *et al*., 1994; Zelko *et al*., 2002).

The conclusion that the two major families identified in the phylogenetic analysis are cytoplasmic and extracellular SOD is further supported by the presence or absence of N-terminal cleavage sequences and subcellular targeting predictions. Like the extracellular SODs in *C. elegans* and *M. musculus*, four out of the five novel Cu Zn SODs in insects contain signal peptides and are thus predicted to be secreted (Table 1). The *C. capitata* extracellular SOD (SOD2) is the only one that apparently lacks a signal peptide and is not predicted initially to be extracellular (Table 1). However, a closer inspection of the *C. capitata* sequence revealed that it has four potential ATG start sites upstream of the one that was inferred to be used (Banks *et al*., 1995), suggesting a possible alternative start site. Indeed, using the entire open reading frame, including those amino acids upstream of the predicted methionine start site, revealed a signal cleavage sequence (italicized and underlined in Fig. 1) that changed the routing prediction to secreted with high reliability. This eighteen amino acid run also has 50% protein identity to the extracellular SOD (SOD3) signal sequence of *D. melanogaster*. The high identity strongly suggests that this N-terminal extension of the open reading frame is either functional now or was in the recent past. Thus, actual analysis of the peptide needs to be carried out in order to verify the true start site and to answer the question of whether *C. capitata* extracellular SOD (SOD2) has recently lost its leader peptide or if there is an alternative start position. At the very least, we can conclude that the gene is a member of the insect extracellular Cu Zn SOD lineage.

In the cytoplasmic clade, *C. elegans* appears to have undergone a gene duplication event producing two cytoplasmic Cu Zn SODs (SOD1 and SOD5). This pair is among 333 likely pairs of duplicated genes in *C. elegans* (Lynch & Conery, 2000; Katju & Lynch, 2003). The second *C. elegans* cytoplasmic Cu Zn SOD (SOD5) has not been previously described, but was correctly annotated in the database as a probable Cu Zn SOD. The gene duplication hypothesis is supported by the presence of an N-terminal peptide run (positions 5–29) in both sequences and a shared insertion of three amino acids at positions 214–216. Neither of the extra N-terminal peptide sequences are predicted to be signal cleavage peptides, nor could any potential function be assigned to these N-terminal peptide runs. This is not the first SOD gene duplication described in *C. elegans*. *C. elegans* has many duplicated genes (Lynch & Conery, 2000; Katju & Lynch, 2003) including two mitochondrial Mn SODs (SOD2 and SOD3) as well (Hunter *et al*., 1997).

The cytoplasmic Cu Zn SOD sequences are much more conserved than the extracellular Cu Zn SODs across the species studied. In the case of our study, the extra variation resulted in difficulty resolving the relationship of four extracellular SOD branches. There are two potential biological implications of the greater variation of extracellular Cu Zn SOD than cytoplasmic Cu Zn SOD. Either the extracellular SODs diverged into subfamilies before the animal body plans evolved and the current pattern resulted from lineage sorting, or there is greater functional selection acting on extracellular Cu Zn SOD than on cytoplasmic Cu Zn SOD.

Drosophila extracellular Cu Zn SOD (SOD3) is not likely highly expressed in adults because it has never been directly observed against an SOD1 mutant background. Extracellular Cu Zn (SOD3) could be difficult to observe if it behaves like mitochondrial Mn SOD (SOD2) or perhaps is unstable under the gel conditions commonly used. Another possibility for the elusive nature of *Drosophila* extracellular Cu Zn SOD (SOD3) is that it is membrane bound like one of the RNA splice variants of extracellular SOD (SOD4) (Fuji *et al*., 1998). The most commonly used SOD assay for whole flies discards cellular debris including membranes in the first step (Mockett *et al*., 2002). Unfortunately there is no current method for distinguishing between cytoplasmic and extracellular Cu Zn enzyme activities if the two are not separated during the isolation procedure. Current biochemical methods are only able to distinguish between Mn SOD and Cu Zn SOD enzyme activities (Mockett *et al*., 2002). One must assume that any study assaying total body homogenates or mixes of cellular compartments is potentially measuring both cytoplasmic and extracellular Cu Zn SOD enzyme activities together.

Measurements of insect extracellular Cu Zn SOD activity alone may have already been accomplished in studies where cell and plasma Cu Zn SOD activities were measured separately. In a study of Cu Zn SOD activity in honeybee sperm, a much higher SOD activity was found in the seminal plasma than in the cellular pellet (Weirich *et al.*, 2002). The authors could not explain this measured activity outside of cells because no one expected that an extracellular Cu Zn SOD existed in insects.

The widespread existence of extracellular Cu Zn SOD in insects could have ramifications for previous work on longevity and antioxidants that used insects as model systems depending upon the normal level of expression and the inducibility of extracellular Cu Zn SOD. There are many conflicting and ambiguous results from experiments on the importance of Cu Zn SOD for insect longevity (Le Bourg, 2001; Sohal *et al.*, 2002; Orr & Sohal, 2003; Spencer *et al.*, 2003; Parker *et al.*, 2004). The possibility of coregulation and compensatory expression among the insect cytoplasmic and extracellular SODs could eventually help to explain some of these results. Compensatory gene expression of extracellular Cu Zn SOD (SOD3) at various developmental or even adult stages could have gone undetected in experimental lines that were thought to lack or vary in cytoplasmic Cu Zn SOD (SOD1). For example, fly lines homozygous for a low expression cytoplasmic Cu Zn allele, SOD-CA1, retained twice the predicted level of Cu Zn SOD activity (Graf & Ayala, 1986). This same study demonstrated an effect of the second chromosome on expression of cytoplasmic Cu Zn SOD (SOD1). Extracellular Cu Zn SOD (SOD3) is on chromosome 2. Finally, an extra band was observed on SOD activity gels in both long lived and normal lived control lines of Drosophila in another study (Hari *et al.*, 1998).

In conclusion, using gene sequence data and *in silico* analysis, we have found strong evidence for the existence of an extracellular SOD homologue in insects based on sequence homology, conservation of all key functional domains, presence of signal cleavage sequences, and phylogenetic reconstruction that groups the novel insect Cu Zn SOD with *M. musculus* and *C. elegans* extracellular Cu Zn SODs. Thus, our results overturn the dogma that there is no member of the extracellular SOD gene family in insects, and suggest a general pattern of SOD defence for metazoans that includes the full repertoire of mitochondrial, cytoplasmic and extracellular SODs.

Experimental procedures

Initial cloning and sequencing

Total RNA was prepared from a pool of two queens, 100 workers and twenty males using the Totally RNA kit™ (Ambion, Austin, TX, USA). cDNA was made with the 3′ reverse transcription step of the FirstChoice RLM-RACE kit™ also from Ambion. Codehop amplification (Rose *et al.*, 1998) was then performed using the primers SOD 3-A (5′-GGA TGT ACT TCT ACG GGA GCA CAY TWY AAY CC-3′), SOD 3-1(5′-GAA TCA GGA TGA TTT CCT TTT CCX ARR TCT C-3′), and Amplitaq Gold™ polymerase from Perkin Elmer (Boston, MA, USA) with a Mg^{2+} gradient of 3 mM, 3.5 mM and 4 mM. An initial 9 min denaturation at 94 °C was followed by seven cycles where annealing started at 40 °C and then increased to 60 °C at the rate of 0.2 °C/ s. This was followed by thirty cycles with 60 °C annealing. PCR at all magnesium concentrations produced a ladder of bands up to approximately 650 bp in size. All reactions were combined on a preparative agarose gel and the 240 bp band was extracted using a Qiaquick Gel Extraction kit (Valencia, CA. USA). Purified DNA was then subcloned with Invitrogen's PCR II TOPO TA Dual Promoter Vector™ (Carlsbad, CA, USA). One white colony was selected for sequencing by Microsynth (Balgach, Switerland). A subsequent BlastX search at the NCBI web site revealed that the clone shared significant homology with Cu Zn SODs across many species.

5′ and 3′ extension (RACE)

5′ and 3′ rapid amplification of cDNA ends (RACE) was carried out using the FirstChoice RLM-RACE kit from Ambion (Austin, TX, USA) with the same total RNA used above. Based on the sequence obtained above, 5′ racing primers were SOD 3-B (5′-TCC GGT CAG CGA GAT CAT-3′) for the outer primer and SOD 3-C (5′-CAG ATC GCC GAC ATG AGA) for the inner primer. 3′ racing primers were SOD 3-2 (5′-AGA CAT GTC GGC GAT CTG-3′) for the outer primer and SOD 3-3 (5′-ATG ATC TCG CTG ACC GGA-3′) for the inner primer. The PCR reaction was performed as described in the FirstChoice RLM-RACE kit using a Biometra gradient thermocycler (Goettingen, Germany) with annealing temperatures ranging from 50 °C to 65 °C for both the inner and outer PCRs. Ambion SuperTaq Polymerase™ was used for all reactions. The 5′ race reaction generated a PCR product of approximately 420 bp while the 3′ race PCR generated 650 bp and 425 bp bands. All three bands were subcloned into an Invitrogen TOPO TA PCR 2.1 vector. Sequencing was performed by Microsynth and the sequencing facility at the University of Lausanne. Alignments of the sequence revealed two full length transcripts differing only at their 3′ untranslated regions. All subclones were sequenced on both strands and the sequence representing the longest complete transcript was submitted to GenBank (accession no. AY672457).

Database searches

The insect genes were identified by searches of model system insect (*Drosophila melanogaster*, *Apis mellifera* and *Anopheles gambiae*) cDNA or genomic databases. We also included the two Cu Zn SODs from *Lasius niger* and the two from *Ceratitis capitata*. We used one mammal (*Mus musculus*) and one worm (*Caenorhabditis elegans*) as our two cases with well characterized extracellular Cu Zn SODs to maximize representation of model system metazoans while keeping the number of taxa to a minimum to retain maximum power in the phylogenetic analyses. All of the potential SOD genes that we selected for this study contain the SOD recognition sequences (Prosite PS00087 and PS00323 (Parker & Blake, 1988; Bannister *et al.*, 1991; Smith & Doolittle, 1992), the cysteine cross-linking sites and all metal binding sites (Bordo *et al.*, 1994). All sequences were from cDNA clones from expression libraries or other specific studies demonstrating gene

expression with the exception of the *A. gambiae*, where the cDNA sequence was inferred from the genomic sequence.

Sequence analysis

Protein sequences were aligned with Clustal X (Thompson *et al.*, 1994). In the case of *A. gambiae*, the Cu Zn SOD sequences had at least one extra exon sequence that was attached by the initial computer automated analysis. Hence, *A. gambiae*'s Cu Zn SOD sequences were trimmed of the extra exon sequence and the last few amino acids retranslated from what was previously identified as the last intron sequence giving the sequences used in Fig. 1. The *A. gambiae* cytoplasmic SOD also had an apparent incorrect start site prediction which was deleted up to the first methionine in the final version of the inferred sequence used here. The final sequences (Fig. 1) were then examined with the program TargetP (Emanuelsson *et al.*, 2000) to look for signal peptide cleavage sites and to assess the likely subcellular localization of the various gene products.

Phylogenetic analysis

Protein sequences rather than DNA sequences were used for phylogenetic analysis because the SODs are an ancient gene family. Protein sequences are more likely to reflect the correct older phylogenetic and/or functional relationships that interest us here because protein mutations are less frequent and less likely to saturate (Nei & Kumar, 2000). Only portions of the sequence alignments without gaps were used in the phylogenetic analysis because of likely differing rates of evolution among functional regions of the enzyme. This also eliminates the influence of the signal sequences on the analysis, which might unfairly bias the results towards finding two gene families by overweighting the presence or absence of a signal cleavage peptide. Various models were explored within each analysis package to determine the model that gave the highest likelihood score. The maximum likelihood trees were constructed using three different programs: MrBayes 3.0b4 (mixed models, one invariant plus eight variable categories, mixed model, 2 000 000 generations, 2000 burnout (Huelsenbeck & Ronquist, 2001)), PAML 3.13d (star decomposition, WAG matrix, one invariant plus eight variable gamma categories (Yang, 1997)), and Tree-Puzzle (quartet puzzling, WAG matrix, one invariant plus eight variable gamma categories (Strimmer & von Haeseler, 1996)). The best trees from all three analyses were then compared with Tree-Puzzle to find the one with the highest likelihood score under the same conditions. The tree topologies could not be tested by the usual HK methods because these trees were not chosen a priori (Goldman *et al.*, 2000).

Acknowledgements

We thank Robin Mockett for helpful comments on the manuscript. This work was supported with grants to Laurent Keller from The AETAS Foundation (Geneva), The Fondation A. R. & J. Leenards (Lausanne), and The Swiss National Science Foundation.

References

Arking, R., Buck, S., Hwangbo, D. and Lane, M. (2002) Metabolic alterations and shifts in energy allocations are corequisites for the expression of extended longevity genes in *Drosophila*. *Ann NY Acad Sci* **959**: 251–262.

Banks, G.K., Robinson, A.S., Kwiatowski, J., Ayala, F.J., Scott, M.J. and Kriticou, D. (1995) A second superoxide dismutase in the medfly, *Ceratitis capitata*. *Genetics* **140**: 697–702.

Bannister, W.H., Bannister, J.V., Barra, D., Bond, J. and Bossa, F. (1991) Evolutionary aspects of superoxide dismutase, the copper/zinc enzyme. *Free Radic Res Commun* **12–13**: 349–361.

Barja, G. (2002) Endogenous oxidative stress: relationship to aging, longevity and caloric restriction. *Ageing Res Rev* **1**: 397–411.

Bordo, D., Djinovic, K. and Bolognesi, M. (1994) Conserved patterns in the Cu, Zn superoxide dismutase family. *J Mol Biol* **238**: 366–386.

Crapo, J.D., Oury, T.D., Rabouille, C., Slot, J.W. and Chang, L. (1992) Copper, zinc superoxide dismutase is primarily a cytosolic protein in human cells. *Proc Natl Acad Sci USA* **89**: 10405–10409.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heinje, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**: 1005–1016.

Finkel, T. and Holbrook, N.J. (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* **408**: 239–247.

Folz, R.J., Guan, J.G., Seldin, M.F., Oury, T.D., Enghild, J.J. and Crapo, J.D. (1997) Mouse extracellular superoxide dismutase: primary structure, tissue-specific gene expression, chromosomal localization, and lung in situ hybridization. *Am J Resp Cell Mol* **17**: 393–403.

Fridovich, I. (1987) Superoxide dismutases. *Adv Enzymol Relat Areas Mol Biol* **58**: 66–97.

Fujii, M., Ishii, N., Joguchi, A., Yasuda, K. and Ayusawa, D. (1998) A novel superoxide dismutase gene encoding membrane-bound and extracellular isoforms by alternative splicing in *Caenorhabditis elegans*. *DNA Res* **5**: 25–30.

Goldman, N., Anderson, J.P. and Rodrigo, A.G. (2000) Likelihood-based tests of topologies in phylogenetics. *Syst Biol* **49**: 652–670.

Graf, J.-D. and Ayala, F.J. (1986) Genetic variation for superoxide dismutase level in *Drosophila melanogaster*. *Biochem Genet* **24**: 153–163.

Hari, R., Burde, V. and Arking, R. (1998) Immunological confirmation of elevated levels of Cu Zn superoxide dismutase protein in an artificially selected long-lived strain of *Drosophila melanogaster*. *Exp Gerontol* **33**: 227–237.

Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H. and Vijg, J. (2003) Aging and genome maintenance: lessons from the mouse. *Science* **299**: 1355–1359.

Hekimi, S. and Guarente, L. (2003) Genetics and the specificity of the aging process. *Science* **299**: 1351–1354.

Hölldobler, B. and Wilson, E.O. (1990) *The Ants*. Springer-Verlag, Berlin.

Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Baysian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Hunter, T., Bannister, W.H. and Hunter, G.J. (1997) Cloning, expression and characterization of two manganese superoxide dismutases from *Caenorhabditis elegans*. *J Biol Chem* **272**: 28652–28659.

Katju, V. and Lynch, M. (2003) The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. *Genetics* **165**: 1793–1803.

Kirby, K., Hu, J., Hilliker, A.J. and Phillips, J.P. (2002) RNA interference-mediated silencing of Sod2 in *Drosophila* leads to early adult-onset mortality and elevated endogenous oxidative stress. *Proc Natl Acad Sci USA* **99**: 16162–16167.

Kutter, H. and Stumper, R. (1969) Hermann Appel, ein leidegeadelter entomologe (1892–1966). In *Proceedings of the Sixth International Congress of the IUSSI (Bern)*, pp. 275–279. IUSSI, Bern, Switzerland.

Le Bourg, E. (2001) Oxidative stress, aging and longevity in *Drosophila melanogaster* (minireview). *FEBS Lett* **498**: 183–186.

Lynch, M. and Conery, J. (2000) The evolutionary fate of and consequences of duplicate genes. *Science* **290**: 1151–1155.

Marklund, S.L. (1984) Extracellular superoxide dismutase and other superoxide dismutase isoenzymes in tissues from nine mammalian species. *Biochem J* **222**: 649–655.

Mockett, R.J., Bayne, A.-C.V., Sohal, B.H. and Sohal, R.S. (2002) Biochemical assay of superoxide dismutase activity in Drosophila. *Methods Enzymol* **349**: 287–292.

Nei, M. and Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

Okado-Matsumoto, A. and Fridovich, I. (2001) Subcellular distribution of superoxide dismutases (SOD) in rat liver. *J Biol Chem* **276**: 38388–38393.

Ookawara, T., Kizake, T., Takayama, E., Imazeki, N., Matsubara, O., Ikeda, Y., Suzuki, K., Ji, L.L., Tadakuma, T., Taniguchi, N. and Ohno, H. (2002) Nuclear translocation of extracellular superoxide dismutase. *Biochem Biophys Res Commun* **296**: 54–61.

Orr, W.C. and Sohal, R.S. (2003) Does overexpression of Cu, Zn-SOD extend life span in *Drosophila melanogaster*? *Exp Gerontol* **38**: 227–230.

Parker, M.W. and Blake, C.C.F. (1988) Iron- and manganese-containing superoxide dismutases can be distinguished by analysis of their primary structures. *FEBS Lett* **229**: 377–382.

Parker, J.D., Parker, K.M., Sohal, B.H., Sohal, R.S. and Keller, L. (2004) Decreased expression of Cu-Zn Superoxide Dismutase 1 in ants with extreme lifespan. *Proc Natl Acad Sci USA* **101**: 3486–3489.

Perez-Campo, R., López-Torres, M., Cadenas, E., Rojas, C. and Barja, G. (1998) The rate of free radical production as a determinant of the rate of aging: evidence from the comparative approach. *J Comp Physiol B* **168**: 149–158.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M. and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**: 1628–1635.

Serra, V., von Zglinicki, T., Lorenz, M. and Saretzki, G. (2003) Extracellular superoxide dismutase is a major antioxidant in human fibroblasts and slows telomere shortening. *J Biol Chem* **278**: 6824–6830.

Smith, M.W. and Doolittle, R.F. (1992) A comparison of evolutionary rates of the 2 major kinds of superoxide-dismutase. *J Mol Evol* **34**: 175–184.

Sohal, R.S., Mockett, R.J. and Orr, W.C. (2002) Mechanisms of aging: An appraisal of the oxidative stress hypothesis. *Free Radic Biol Med* **33**: 575–586.

Spencer, C.C., Howell, C.E., Wright, A.R. and Promislow, D.E.L. (2003) Testing an 'aging gene' in long-lived *Drosophila* strains: increased longevity depends on sex and genetic background. *Aging Cell* **2**: 123–130.

Strimmer, K. and von Haeseler, A. (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13**: 964–969.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (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**: 4673–4680.

Tibell, L.A.E., Skärfstad, E. and Jonsson, B. (1996) Determination of the structural role of the N-terminal domain of human extracellular superoxide dismutase by use of protein fusions. *Biochim Biophys Acta* **1292**: 47–52.

Weirich, G.F., Collins, A.M. and Williams, V.P. (2002) Antioxidant enzymes in the honey bee, *Apis mellifera*. *Apidologie* **33**: 3–14.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**: 555–556.

Zelko, I.N., Mariani, T.J. and Folz, R.J. (2002) Superoxide Dismutase multigene family: a comparison of the Cu-Zn-SOD (SOD1), Mn.SOD (SOD2) and EC-SOD (SOD3) gene structures, evolution and expression. *Free Radic Biol Med* **33**: 337–349.

von Zglinicki, T. (2002) Oxidative stress shortens telomeres. *Trends Biochem Sci* **27**: 339–344.