Phylogenetic Relationships Among Prokaryotic and Eukaryotic Catalases

Martin G. Klotz, Glen R. Klassen,† and Peter C. Loewen†*

*Department of Biology, University of Colorado–Denver; and †Department of Microbiology, University of Manitoba, Winnipeg

Seventy-four catalase protein sequences, including 29 bacterial, 8 fungal, 7 animal, and 30 plant sequences, were compiled, and 70 were used for phylogenetic reconstruction. The core of the resulting tree revealed unique, separate groups of plant and animal catalases, two groups of fungal catalases, and three groups of bacterial catalases. The only overlap of kingdoms occurred within one branch and involved fungal and bacterial large-subunit enzymes. The other fungal branch was closely linked to the group of animal enzymes. Group I bacterial catalases were more closely related to the plant enzymes and contained such diverse taxa as the Gram-positive *Listeria seeligeri, Deinococcoccus radiodurans,* and γ-proteobacteria. Group III bacterial sequences were more closely related to fungal and animal sequences and included enzymes from a broad range of bacteria including high- and low-GC Gram positives, proteobacteria, and a bacteroides species. Group II was composed of large-subunit catalases from diverse sources including Gram positives (low-GC *Bacilli* and high-GC *Mycobacteria*), proteobacteria, and species of the filamentous fungus *Aspergillus.* These data can be interpreted in terms of two gene duplication events that produced a minimum of three catalase gene family members that subsequently evolved in response to environmental demands. Horizontal gene transfer may have been responsible for the group II mixture of bacterial and fungal large-subunit catalases.

Introduction

The decomposition of hydrogen peroxide by catalase has been the object of study for over 150 years, and purified catalases have been available for study for over 50 years. Despite extensive work characterizing the catalatic mechanism, only two catalases of prokaryotic origin, from *Micrococcus luteus* (Herbert and Pinsent 1948) and *Rhodobacter spheroides* (Clayton 1959), had been purified prior to 1979, and both appeared to be structurally similar to their eukaryotic counterparts. In 1979, two catalases from *E. coli* were purified and characterized, both of which differed from the typical catalase in several significant respects. Hydroperoxidase I (HPI) exhibited an organic peroxidatic activity in addition to the catalatic activity (Claiborne and Fridovich 1979), and HPII, while appearing to be a monofunctional catalase, was larger and contained a different heme (Claiborne, Malinowski, and Fridovich 1979; Loewen and Switala 1986). Since then, catalases from a diverse collection of bacteria have been purified and characterized, providing examples that resemble HPI, HPII, and the typical catalases. In addition, nonheme catalases have been purified from two organisms, further increasing the diversity of catalatic enzymes (Kono and Fridovich 1983; Allgood and Perry 1986).

A significant number of catalase sequences have recently been reported, bringing the total number to 74, compared to the 20 that were included in the original phylogenetic comparison of catalase sequences published just 4 years ago (von Ossowski, Hausner, and Loewen 1993). This includes 23 new bacterial catalase sequences, 6 new fungal sequences, 2 new animal sequences, and 23 new plant sequences. A second group of proteins with associated catalase and peroxidase activities, the catalase-peroxidases, have also been the focus of attention, and there are now 10 such sequences available which exhibit sequence similarity to plant peroxidases rather than to catalases. Guan and Scandalios (1996) recently described a phylogenetic reconstruction of 16 plant catalase sequences and discussed possible evolutionary relationships to bacterial and other eukaryotic catalases. In addition, Mayfield and Duvall (1996) described anomalous groupings of catalase gene sequences using a subset of nine bacterial and five eukaryotic sequences. Both of these reports utilized small subsets of the total number of catalase sequences available, leaving in question the overall phylogenetic relationships within the larger family. In order to present a more complete picture of catalase sequence relationships, this paper reports a phylogenetic comparison of 70 of the 74 available catalase sequences.

Materials and Methods

Organisms and Sequence Data

The database sources (accession numbers) of the catalase sequences that were included in the analysis are summarized in table 1, along with a classification of the various bacteria.

Phylogenetic Analyses

The alignments of the amino acid sequences were created using the multiple-alignment feature of ClustalW (Thompson, Higgins, and Gibson 1994) with maximum fixed-gap and gap extension penalties to reduce the influence of gaps in the subsequent analyses. Generally, higher confidence values were observed in the constructed trees as a result. Estimates of the evolutionary relationships of the sequences were determined using both parsimony and distance methods. Phylogenies based on unrooted parsimony and distance criteria were generated using programs contained in the PHYLIP package (version 3.5; Felsenstein 1994). In or-

Key words: catalase, parallel evolution, horizontal gene transfer, phylogeny.

Address for correspondence and reprints: Peter C. Loewen, Department of Microbiology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. E-mail: peter_loewen@umanitoba.ca.

Mol. Biol. Evol. 14(9):951–958. 1997

Table 1
Catalase and Catalase Peroxidase Sequences Used

Species	Accession No.[a]	Classification[b]
Bacterial catalases		
Bacillus firmus KatA	M74194	Gram positive, low GC
Bacillus subtilis KatA	M80796	Gram positive, low GC
Bacillus subtilis KatE	X85182	Gram positive, low GC
Bacteroides fragilis KatB	U18676	*Cytophaga/Flexibacter/ Bacteriodes*
Bordetella pertussis KatA	U07800	Proeobacteria-β
Brucella abortus KatA	U11439	Proteobacteria-α
Campylobacter jejuni KatA	X85130	Proteobacteria-ϵ
Deinococcus radiodurans KatA	D63898	*Thermus/Deinococcus*
Escherichia coli KatE	M55161	Proteobacteria-γ
Haemophilus influenzae HktE	U02682	Proteobacteria-γ
Helicobacter pylori Kat	U67458	Proteobacteria-ϵ
Lactobacillus sake KatA	M84015	Gram positive, low GC
Listeria seeligeri Kat	M75944	Gram positive, low GC
Methylobacterium extorquens KatA	L48340	Proteobacteria-γ
Micrococcus luteus CatA	P29422	Gram positive, high GC
Mycobacterium avium KatE	L41246	Gram positive, high GC
Neisseria gonorrhoeae KatA	U35457	Proteobacteria-β
Proteus mirabilis KatA	P42321	Proteobacteria-γ
Pseudomonas aeruginosa KatB	U34896	Proteobacteria-γ
Pseudomonas fluorescens CatB	U83329	Proteobacteria-γ
Pseudomonas fluorescens pKatA	U72068	Proteobacteria-γ
Pseudomonas putida CatA	U63511	Proteobacteria-γ
Pseudomonas putida CatC	U82622	Proteobacteria-γ
Pseudomonas syringae CatF	U03465	Proteobacteria-γ
Rhizobium meliloti KatA	U59271	Proteobacteria-α
Rhizobium sp. Cat	U56239	Proteobacteria-α
Streptomyces coelicolor CatA	X96981	Gram positive, high GC
Streptomyces violaceus Bca	X74791	Gram positive, high GC
Xanthomonas oryzae KatX	X97673	Proteobacteria-γ
Plant catalases		
Arabidopsis thaliana CatA	X64271	
Curcurbita pepo Cat1 (squash)	D55645	
Curcurbita pepo Cat2 (squash)	D55646	
Curcurbita pepo Cat3 (squash)	D55647	
Glycine max CatA (soybean)	Z12021	
Gossypium hirsutum Cat1 (cotton)	X52135	
Gossypium hirsutum Cat2 (cotton)	X56675	
Helianthus annuus CatA (sunflower)	L28740	
Hordeum vulgare Cat1 (barley)	U20777	
Hordeum vulgare Cat2 (barley)	U20778	
Ipomoea batatas CatA (sweet potato)	X05549	
Lycopersicon esculentum Cat1 (tomato)	M93719	
Nicotiana plumbaginifolia Cat1 (tobacco)	Z36975	
Nicotiana plumbaginifolia Cat2 (tobacco)	Z36976	
Nicotiana plumbaginifolia Cat3 (tobacco)	Z36977	
Nicotiana sylvestris Cat1 (tobacco)	U07626	
Nicotiana sylvestris Cat2 (tobacco)	U07627	
Oryza sativa CatA (rice)	X61626	
Oryza sativa CatB (rice)	D26484	
Phaseolus aureus CatA (mungbean)	D13557	
Pisum sativum CatA (pea)	X60169	
Ricinus communis Cat1 (castor bean)	D21161	
Ricinus communis Cat2 (castor bean)	D21162	
Secale cereale Cat (rye)	Z54143	
Solanum melongena Cat (eggplant)	X71653	
Solanum tuberosum Cat1 (potato)	U27082	
Triticum aestivum CatA (wheat)	X94352	
Zea mays Cat1 (maize)	X12538	
Zea mays Cat2 (maize)	X54819	
Zea mays Cat3 (maize)	L05934	
Animal catalases		
Bos taurus CatA (bovine)	P00432	
Caenorhabditis elegans Cat (nematode)	X82175	
Drosophila melanogaster CatA (fruit fly)	X52286	
Homo sapiens CatA (human)	X04076	
Mus musculus CatA (mouse)	M62897	
Onchocerca volvulus Cat (nematode)	X82176	
Rattus norvegicus CatA (rat)	M11670	

Table 1
Continued

Species	Accession No.[a]	Classification[b]
Fungal catalases		
Aspergillus nidulans CatA	U37803	
Aspergillus niger CatR	Z23138	
Botrytis cinerea CatA	Z54346	
Candida tropicalis CatA	M18832	
Hansenula polymorpha CatA	X56501	
Penicillium vitale CatA	P11934	
Saccharomyces cerevisiae (CatA)	X13028	
Saccharomyces cerevisiae (CatT)	X04625	

[a] The accession numbers with the prefix "P" were obtained from SwissProt, and all other sequences were obtained from GenBank.

[b] The bacterial classifications were taken from Olsen, Woese, and Overbeek (1994).

der to estimate the confidence in the potential monophyly of clusterings comprised of more than one species, bootstrap analysis was carried out. The SEQBOOT program was used to generate 200 bootstrap replicates, each of which was analyzed by the PROTPARS or PROTDIST programs. CONSENSE was used to construct a majority-rule consensus tree from the parsimony replicates from PROTPARS. The distance matrix replicates from PROTDIST were analyzed by the neighbor-joining method (Saitou and Nei 1987) in NEIGHBOR and the Fitch-Margoliash algorithm in FITCH.

Results

Sequence Alignments

The monofunctional catalases were aligned independently of the 10 catalase-peroxidases, which have very dissimilar sequences and are not considered further in this report. Alignment of 74 catalases revealed a core region of approximately 360 amino acids with significant sequence similarity among all catalases. Because of the large number of sequences involved, the alignments are not shown here, but an alignment of 20 sequences including representative enzymes from plant, animal, fungal, and bacterial sources has previously been published (von Ossowski, Hausner, and Loewen 1993). On either side of this conserved core region, the sequences are more divergent, although the subgroup of large-subunit catalases (more than 650 residues as compared to the more normal 480 residues in small-subunit catalases) retained some similarity in the extended carboxy domain.

Three of the sequences, those from *Micrococcus luteus, Nicotiana sylvestris,* and *Penicillium vitale,* were excluded from the phylogenetic reconstructions because of uncertainties or truncations in their sequences. For example, the *M. luteus* and *P. vitale* catalases have never been sequenced, and the reported sequences were derived from the crystal structure and analogy with the bovine enzyme, which would result in an inherent bias in the sequences toward the bovine sequence at any residues for which the electron density was not clear. The *N. sylvestris* sequence was excluded because of its apparent incompleteness at the N-terminus. Finally, we excluded the *Botrytis cinerea* sequence because we found that its alignment with other fungal sequences was not as expected, suggesting the presence of possible errors.

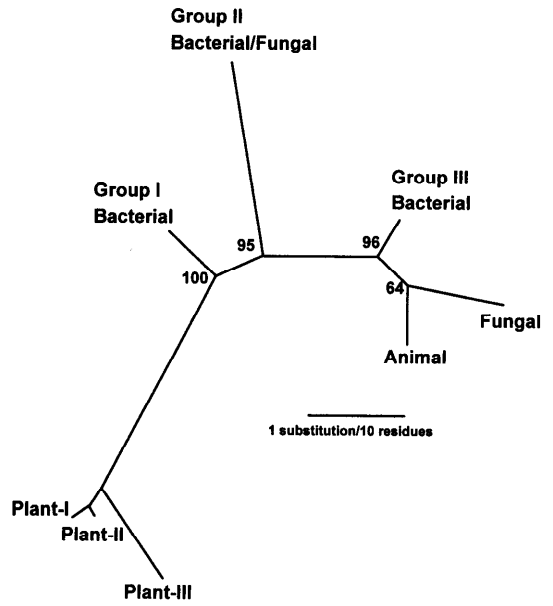

FIG. 1.—Unrooted phylogenetic tree based on the amino acid sequences of the 70 bacterial catalases constructed by parsimony and distance methods. The lengths of the branches are a reflection of the evolutionary distances as the average number of amino acid changes over a given length calculated by FITCH. The numbers at the nodes represent the proportion of bootstrap samplings that support the topology shown, and only values at the main nodes are shown. Two hundred bootstrap replicates were analyzed. The detailed branch patterns within the individual groups are described in figure 2.

Phylogenetic Relationships—Catalases

Phylogenetic reconstruction applying parsimony methods to the core sequences of 70 catalases resulted in an unrooted tree that was similar in its key features to the tree previously reported (von Ossowski, Hausner, and Loewen 1993). The diagram in figure 1 describes the core of the tree focusing on the key branch points that lead to the main groupings of species. Significantly, catalases from plants and animals group separately in unique branches on the tree. Fungal catalase sequences are found in two groups: one group that is close to the animal branch contains only fungal sequences, and a second group (labeled "Group II–Bacterial/Fungal" in fig. 1) contains a mixture of large-subunit bacterial and

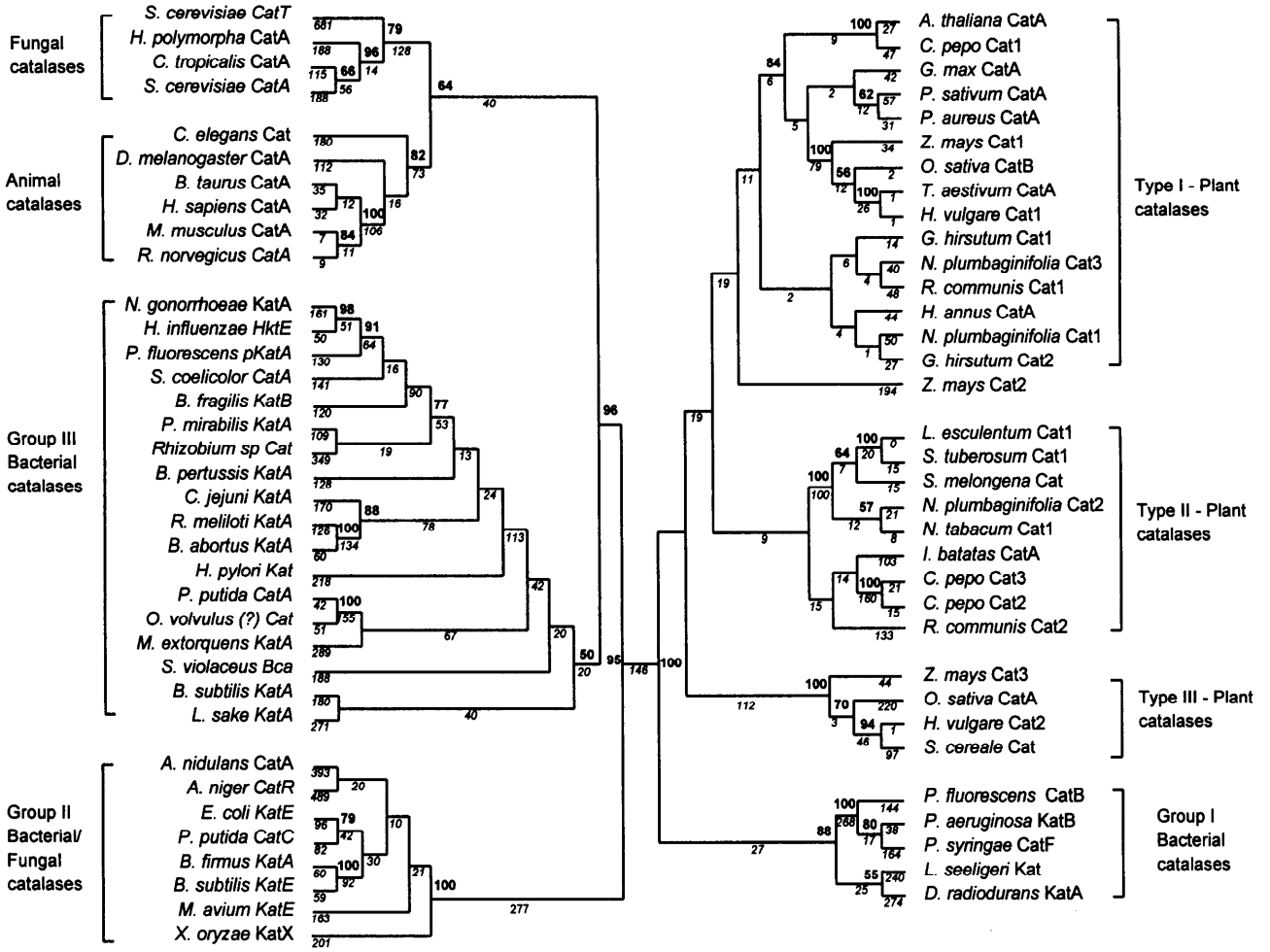

FIG. 2.—Unrooted phylogenetic tree based on the amino acid sequences of 70 catalases constructed by parsimony methods. Distance methods resulted in a similar tree, from which the lengths of the branches, in italics below the branch, were taken. The branch lengths reflect the evolutionary distances calculated as the average number of amino acid changes per one thousand residues. The numbers in bold at the nodes represent the proportion of bootstrap samplings that support the topology shown, and only those values greater than 50% are shown. Two hundred bootstrap replicates were analyzed.

fungal sequences. Bacterial catalases are found in three distinct groups. Groups I and III contain exclusively small-subunit prokaryotic sequences, while group II, as already mentioned, contains a mixture of bacterial and fungal large-subunit sequences. The bootstrap confidence levels of these main branches are very high, in the 95%–100% range, with the exception of that of the fungal/animal separation, which is only 64%. Distance methods (neighbor-joining and Fitch) gave rise to a similar tree topology, from which the branch lengths shown in figure 1 are derived. Similar trees were also derived using the full lengths of the catalase sequences (data not shown). The topology and composition of the individual branches are shown in detail in figure 2.

The animal and fungal catalases share one branch of the tree in figure 1, indicating a closer relationship among the catalases from fungal and animal sources than with enzymes from any other source. One surprising result was the absence from the animal group of the catalase ascribed to *O. volvulus,* a nematode, and its location with bacterial catalases in group III closely linked to the *P. putida* CatA (fig. 2). The nematode sequence was derived from isolated mRNA, and it seems likely that the mRNA ascribed to *O. volvulus* may have originated from a bacterial parasite in the worm.

The 29 plant catalases are all found in a single branch of the core tree (fig. 1) and are subdivided into three subgroups (fig. 2), the compositions of which are consistent with those reported earlier by Guan and Scandalios (1996) using only 16 sequences. For example, type I plants now include 15 sequences, up from 9; type III plants include 4 sequences; and the most significant increase lies in type II plants, where there are now 9 sequences as compared to 2 in the original comparison. *Zea mays* Cat2 remains a puzzle because it falls between type I and type II catalases as previously noted by Guan and Scandalios (1996), and analysis of its gene sequence (see below) suggests that it might be more closely related to type III sequences. Similar trees were obtained from both parsimony and distance methods, with many of the internal nodes being supported by low bootstrap confidence values, presumably arising from the relatively short evolutionary distances or small residue varia-

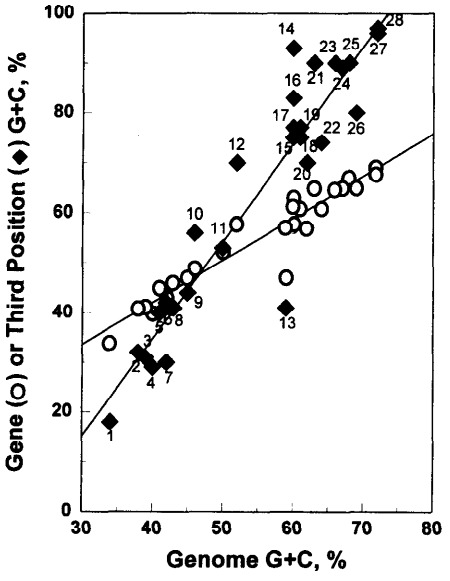

FIG. 3.—Correlation of GC content between the total genomic DNA and either the bacterial catalase gene (○) or the third codon positions (♦) of the catalase genes. The (♦) symbol for each catalase is numbered, and the corresponding (○) symbol for the gene composition can be found either above or below at the same genome G+C value. The straight lines were determined by linear regression. The numbers refer to the following catalase genes: 1. *C. jejuni katA.* 2. *L. seeligeri kat.* 3. *B. firmus katA.* 4. *P. mirabilis kat.* 5. *B. fragilis katB.* 6. *H. pylori kat.* 7. *H. influenzae hktE.* 8. *B. subtilis katA.* 9. *L. sake katA.* 10. *B. subtilis katE.* 11. *E. coli katE.* 12. *N. gonorrhoeae katB.* 13. *P. fluorescens pkatA.* 14. *P. syringae catF.* 15. *P. fluorescens catB.* 16. *R. meliloti katA.* 17. *Rhizobium sp. cat.* 18. *P. putida catC.* 19. *P. putida catA.* 20. *B. abortus katA.* 21. *M. extorquens katA.* 22. *X. oryzae katX.* 23. *D. radiodurans kat.* 24. *P. aeruginosa katB.* 25. *M. avium katE.* 26. *B. pertussis katA.* 27. *S. violaceus bca.* 28. *S. coelicolor catA.*

tions among the plant sequences as compared to other kingdoms.

Mutational Bias

The catalase gene analysis of Mayfield and Duvall (1996) was based on an analysis of DNA sequences (including third codon positions) and did not consider the effect of mutational drift in the third position of the codon on their results. The third positions would have become saturated with multiple substitutions, thereby retaining little phylogenetic information. Furthermore, the data in figure 3 reveal a linear relationship between the G+C content of both the catalase gene and the third codon positions in the gene and the G+C content of the genome, but with different slopes for the two lines. There is an obvious bias to either higher A+T or higher G+C content of the third codon positions in AT-rich or GC-rich organisms, respectively, an observation reported earlier by Muto and Ozawa (1987). This bias is explained in terms of the lower selection pressures on the third position having allowed mutational pressure or genetic drift to introduce more G+C in GC-rich organisms and more A+T in AT-rich organisms. An important conclusion from these data is that species-specific codon bias will affect any phylogenetic comparisons based solely on catalase DNA sequences.

A similar comparison of eukaryotic catalase gene sequences revealed the same bias (data not shown), although there was more scatter among the data. As noted previously by Guan and Scandalios (1996), there was a very striking divergence to a very high G+C content of the gene and third codon position for all type III plant catalases. This was also true for the gene for *Z. mays* Cat2 leading to uncertainty as to its proper grouping. The gene for *D. melanogaster* catalase is the only other eukaryotic gene in this comparision to show a marked deviation to higher G+C (57% in the gene and 77% at the third codon position compared to a genome content of 40% G+C), a phenomenon that has been addressed earlier (Shield et al. 1988).

Application of the principle of G+C content analysis should also allow the detection of horizontal gene transfer, as in the case of the plasmid-encoded pKatA from *P. fluorescens,* where the *katA* and third-codon-position G+C contents are significantly lower than the *P. fluorescens* genome G+C content (fig. 3), suggesting a nonpseudomonad source for *katA.* Similarly, the inclusion of the plasmid-encoded KatP found in enterohemorrhagic *E. coli* O157:H7 (X89017, H. Schmidt, personal communication) in a phylogenetic comparison of catalase-peroxidases (data not shown) reveals that it is not derived from the *E. coli* genome.

Discussion

The basic structure of the phylogenetic tree based on 70 catalase sequences is similar to that of the tree reported earlier using only 20 sequences (von Ossowski, Hausner, and Loewen 1993), in that it contains separate branches for sequences from plant, fungal, and animal sources. However, the additional prokaryotic sequences impart two unique features to the revised core tree as outlined in figure 1. The first is the multiplicity of branches containing prokaryotic catalases as compared to the single branches for plants and animals, and the second is the mixture of bacterial and fungal sequences in group II, the only branch in the core tree to contain sequences from more than one kingdom.

On first examination, the species origins of the catalases in the three bacterial groups are unusual in several respects. Group I sequences are derived from nonpathogenic or opportunistic bacteria that are widespread in nature and include representatives from such diverse taxa as the γ-proteobacterial fluorescent pseudomonads (*P. aeruginosa, P. fluorescens,* and *P. syringae*), the Gram-positive *L. seeligeri,* and a deinococcus, *D. radiodurans.* Group III sequences include catalases predominantly from proteobacteria, but also from a bacteroides species and three Gram-positive bacteria, many of which lead a parasitic life in association with a host. It is interesting to note that the bacteria harboring a group III catalase and having a restricted environment usually contain only a single catalase isozyme *(L. sake* [Knauf, Vogel, and Hammes 1992], *H. pylori* [Odenbreit, Wieland, and Haas 1996], *B. abortus* [Sha, Stabel, and Mayfield 1994], *C. jejuni* [Grant and Park 1995], *B. pertussis* [DeShazer, Wood, and Friedman 1994], *P. mirabilis* [Jouve, Tessier, and Pelmont 1983], *B. fragilis* [Rocha and Smith 1995], *H. influenzae* [Fleischmann et al.

1995], *N. gonorrhoeae* [Archibald and Duong 1986]), in contrast to bacteria from the same group with a diverse habitat, which contain more than one catalase isoenzyme from separate groups, including *B. subtilis* (fig. 2), *P. putida* (fig. 2), *M. extorquens* (M. E. Lidstrom, personal communication), *Rhizobium* sp. (Crockford, Davis, and Williams 1995), *R. meliloti* (Herouart et al. 1996), and *S. coelicolor* (Kim et al. 1994). The makeup of catalase group II is puzzling because it contains enzymes from both high- and low-GC Gram-positive bacteria, γ-proteobacteria, and the fungi *A. nidulans* and *A. niger*. The catalase from the fungus *P. vitale* also falls in this group but was excluded from the comparison because of uncertainties about its sequence. The unifying property of this group of catalases is the fact that they are all large-subunit enzymes (~80 kDa) as compared to other catalases (50–65 kDa).

The tree generated from a comparison of all catalase isozymes is a composite tree, in that sequences from the same species occur on more than one branch. As such, the tree as a whole cannot be compared with the 16S rRNA species tree (Olsen, Woese, and Overbeek 1994) to test for congruence, but if each group is viewed as a separate species tree, profitable comparisons can be drawn. Analysis of group III enzymes reveals the Gram-positive bacteria separated from the proteobacteria and a bacteroides species with only *S. coelicolor* CatA diverging from the expected 16S rRNA phylogeny. Even in the other two groups of bacterial catalases, which are too small to draw firm conclusions, there appear to be the expected groupings. We interpret this as an expression of individual evolution of each of the three members of the bacterial catalase gene family.

From these observations, a model to explain the pattern of evolution evident in catalase sequences can be proposed. The progenitor catalase presumably arose in bacteria for protection as oxygen became more common in the environment, and, indeed, to date there has been no report of a true catalase from the domain archaea; only catalase-peroxidases have been reported. Multiple duplications, in parallel or serial, of this progenitor catalase gene and the parallel evolution of the emerging variants can explain the generation of the family of catalase genes consisting of the three groups shown in figure 1. To include all organismal domains in an explanation of the evolution of the catalase gene family, it is sufficient to propose only two major gene duplications. Because of the uncertainty in organization, the duplications have been represented as a single trifurcation in figure 4. Gene family member 1 includes plant catalases and group I bacterial catalases, with subsequent duplications in the plant branch to give rise to the three types of plant catalases; gene family member 2 includes group II bacterial and fungal catalases; and gene family member 3 includes group III bacterial, animal, and fungal catalases. Within each gene family member, parallel evolution and drift subsequently occurred to give rise to the patterns seen in figure 2. The overall pattern of catalase evolution as outlined in figure 4 is strongly supported by the existence of two different catalase family members in the same organism. For example, both *P. putida* and *B. subtilis* contain enzymes from both group II and group III. Other organisms with catalases from more than one member of the catalase family include *K. pneumoniae* (Goldberg and Hochman 1989), *B. firmus* (Hicks 1995), and *Rhizobium* (hybridization data, not shown; Crockford, Davis, and Williams 1995), but a lack of sequence information precludes assignments to specific groups. At least two gene duplication scenarios are possible within the trifurcation shown in figure 4. Selection of one or the other scenario of gene duplication events may eventually be possible as more catalase sequences become available, particularly if examples of family members 1 and 3 or 1 and 2 are ever found in the same organism.

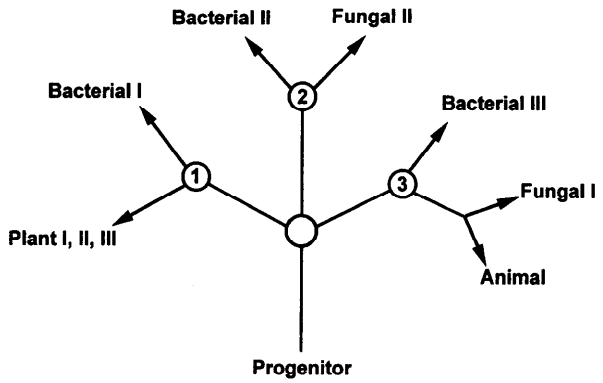

FIG. 4.—A mechanism for the generation of three catalase gene families from a common progenitor catalase gene. The apparent trifurcation represented by the unnumbered circle would require a minimum of two gene duplication events to give rise to the three hypothetical gene family members represented by the circled numbers 1, 2, and 3.

All of the enzymes in group II are large-subunit enzymes, having subunits with a mass greater than 80 kDa, as compared to the normal catalase subunit of 50–65 kDa. A number of these enzymes from both bacterial and fungal sources have been extensively characterized, both kinetically and physically, including determination of the crystal structures of the enzymes from *P. vitale* and *E. coli*. The larger size of these catalases is the result of an extended N-terminal sequence of 70 residues and a C-terminal domain of about 150 residues that are little more than simple additions to the core catalase structure, which remains very similar regardless of subunit size. It is not yet possible to say whether evolution proceeded from small- to large-subunit enzymes or from large- to small-subunit enzymes, but the very simple structural differences between the two suggest that either route would have been feasible, involving either gene fusion or gene cleavage events to add or remove the additional sequences. The extended sequences of the large-subunit enzymes significantly stabilize the core catalase sequence, resulting in an enzyme that remains active at 70°C and in the presence of various denaturing reagents, properties which may have been beneficial at some stage of evolution. Regardless of their position in the evolutionary scheme, the large-subunit enzymes evolved to protect organisms under stressful conditions, consistent with KatE from *E. coli* (Loewen,

Switala, and Triggs-Raine 1985) and CatC from *P. putida* being expressed as part of the stationary phase response.

The most striking feature of group II catalases is the presence of enzymes from different evolutionary domains, and two mechanisms may be considered for explanation: endosymbiosis and horizontal gene transfer. In the context of both mechanisms, it is important to note that the two fungi, *A. niger* and *A. nidulans,* share a common habitat with pseudomonads (*P. putida* and *X. oryzae*) and bacilli (*B. subtilis* and *B. firmus*), which are soil organisms and are often found in association with plant hosts. For example, it is known that plant growth promotion by fluorescent pseudomonads can be achieved by bacterial colonization and growth inhibition of the plant pathogenic fungus *Phytophthora parasitica* (Yang, Menge, and Cooksey 1994). The major location of eukaryotic catalases is the peroxisome, a single membrane-lined organelle that lacks its own DNA or ribosomes but contains a great diversity of protein (Subramani 1993). Apocatalase subunits are exported from the cytoplasm into the peroxisome by a receptor-mediated transport process, followed by association with heme and assembly of the functional protein. If bacteria are ancestors to the peroxisome, this could serve as an explanation for the similarity between the large-subunit bacterial and fungal catalases in group II. However, it seems unlikely that an organelle would lose control over the biosynthesis of one of its key enzymes, and it is hard to imagine that only a few species of the ascomycetes would contain peroxisomes of a different evolutionary origin from those in other fungi, animals, and plants. Therefore, horizontal gene transfer seems the more likely explanation for the group II separation. Transfer of the large-subunit encoding gene could have occurred both between bacterial species and between bacterial and fungal species, but whether the exchange was unidirectional from bacteria to fungi or bidirectional cannot be stated. The horizontal transfer of catalase genes between prokaryotes and eukaryotes in both directions has been proposed as a common occurrence on the basis of DNA sequence comparisons of nine catalase genes (Mayfield and Duvall 1996). However, the limited data set and the use of DNA sequences without consideration being given to drift or bias in the third codon positions, which is evident in the catalase gene sequences analyzed in figure 3, raises doubts about the likelihood of such frequent exchanges. The question of how introns in the eukaryotic gene are treated in prokaryotes following a putative transfer is another problem in rationalizing eukaryote-to-prokaryote transfers that was not addressed. It should be stressed that there is no evidence for lateral gene transfer of catalases between domains anywhere other than among the group II sequences. However, it should be noted that within the prokaryotes there are two anomalous groupings of *S. coelicolor* (high-GC Gram-positive) CatA and *Bacteroides fragilis* KatB with Proteobacteria suggestive of possible interspecies transfer. The *sodC* gene, which encodes periplasmic Cu-Zn superoxide dismutase in various bacteria, has been identified as another candidate for horizontal gene transfer among species (Imlay and Imlay 1996). Furthermore, the plasmid-encoded *P. fluorescens* KatA is clearly derived from a species other than a pseudomonad (fig. 3). A comparison of trees based on RecA and 16S rRNA (Lloyd and Sharp 1993; Eisen 1995), however, suggests that horizontal gene transfer even among bacteria is relatively rare.

The evolutionary selection of one gene family member over another in a given group of organisms suggests that there should be differences in the kinetic or physical properties of the enzymes from the different catalase groups that would provide selective advantages in the different habitats and give rise to the observed groupings. There are clear catalytic differences between the large- and small-subunit enzymes, but it is not clear what differences there are between group I and III small-subunit enzymes that would allow their selection as discrete family members. On the other hand, one might predict that the selective forces experienced by the group of intracellular pathogens within group III, such as bursts of H_2O_2 in the intracellular milieu, may be quite different from those encountered by soil bacteria. Unfortunately, many of the catalases in this comparison have not been purified, and a detailed, side-by-side comparison using common criteria for assay of the isolated enzymes does not exist, and might be considered the next logical step in this study.

Acknowledgments

This work was supported by a grant from the Natural Sciences and Engineering Research Council to P.C.L. (OGP9600) and a Faculty Grant from the University of Colorado (95-46422) to M.G.K. We would like to thank Drs. A. J. Anderson, R. Haas, S. Mongolsuk, and H. Schmidt for providing us with sequences prior to publication.

LITERATURE CITED

ALLGOOD, G. S., and J. J. PERRY. 1986. Characterization of a manganese-containing catalase from the obligate thermophile *Thermoleophilum album.* J. Bacteriol. **168**:563–567.

ARCHIBALD, F. S., and M.-N. DUONG. 1986. Superoxide dismutase and oxygen toxicity defenses in the genus *Neisseria.* Infect. Immun. **51**:631–641.

CLAIBORNE, A., and I. FRIDOVICH. 1979. Purification of the o-dianisidine peroxidase from *Escherichia coli* B. Physicochemical characterization and analysis of its dual catalatic and peroxidatic activities. J. Biol. Chem. **254**:4245–4252.

CLAIBORNE, A., D. P. MALINOWSKI, and I. FRIDOVICH. 1979. Purification and characterization of hydroperoxidase II of *Escherichia coli.* J. Biol. Chem. **254**:11664–11668.

CLAYTON, R. K. 1959. Purified catalase from *Rhodopseudomonas spheroides.* Biochim. Biophys. Acta **36**:40–47.

CROCKFORD, A. J., G. A. DAVIS, and H. D. WILLIAMS. 1995. Evidence for cell density-dependent regulation of catalase activity in *Rhizobium leguminosarum* bv. *phaseoli.* Microbiology **141**:843–851.

DESHAZER, D., G. E. WOOD, and R. L. FRIEDMAN. 1994. Molecular characterization of catalase from *Bordetella pertussis*: identification of the *katA* promoter in an upstream insertion sequence. Molec. Microbiol. **14**:123–130.

EISEN, J. A. 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J. Mol. Evol. **41**:1105–1123.

FELSENSTEIN, J. 1994. PHYLIP version 3.5 user manual. University of Washington, Seattle.

FLEISCHMANN, R. D., M. D. ADAMS, O. WHITE et al. (37 co-authors). 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**:496–512.

GOLDBERG, I., and A. HOCHMAN. 1989. Purification and characterization of a novel type of catalase from the bacterium *Klebsiella pneumoniae.* Biochim. Biophys. Acta **991**:330–336.

GRANT, K. A., and S. F. PARK. 1995. Molecular characterization of *katA* from *Campylobacter jejuni* and generation of a catalase-deficient mutant of *Campylobacter coli* by interspecific allelic exchange. Microbiology **141**:1369–1376.

GUAN, L., and J. G. SCANDALIOS. 1996. Molecular evolution of maize catalases and their relationship to other eukaryotic and prokaryotic catalases. J. Mol. Evol. **42**:570–579.

HERBERT, D., and J. PINSENT. 1948. Crystalline bacterial catalase. Biochem. J. **43**:193–202.

HEROUART, D., S. SIGAUD, S. MOREAU, P. FRENDO, D. TOUATI, and A. PUPPO. 1996. Cloning and characterization of the *katA* gene of *Rhizobium meliloti* encoding a hydrogen peroxide-inducible catalase. J. Bacteriol. **178**:6802–6809.

HICKS, D. B. 1995. Purification of three catalase isozymes from facultatively alkaliphilic *Bacillus firmus* OF4. Biochim. Biophys. Acta **1229**:347–355.

IMLAY, K. R. C., and J. A. IMLAY. 1996. Cloning and analysis of *sodC,* encoding the copper-zinc superoxide dismutase of *Escherichia coli.* J. Bacteriol. **178**:2564–2571.

JOUVE, H. M., S. TESSIER, and J. PELMONT. 1983. Purification and properties of the *Proteus mirabilis* catalase. Can. J. Biochem. Cell Biol. **61**:8–14.

KIM, H., J.-S. LEE, Y. C. HAH, and J.-H. ROE. 1994. Characterization of the major catalase from *Streptomyces coelicolor* ATCC10147. Microbiology **140**:3391–3397.

KNAUF, H. J., R. F. VOGEL, and W. P. HAMMES. 1992. Cloning, sequence, and phenotypic expression of *katA,* which encodes the catalase of *Lactobacillus sake* LTH677. Appl. Environ. Microbiol. **68**:832–839.

KONO, Y., and I. FRIDOVICH. 1983. Isolation and characterization of the pseudocatalase of *Lactobacillus plantarum.* A new manganese containing enzyme. J. Biol. Chem. **258**: 6015–6019.

LLOYD, A. T., and P. M. SHARP. 1993. Evolution of the *recA* gene and the molecular phylogeny of bacteria. J. Mol. Evol. **37**:399–407.

LOEWEN, P. C., and J. SWITALA. 1986. Purification and characterization of catalase HPII from *Escherichia coli* K12. Biochem. Cell Biol. **64**:638–646.

LOEWEN, P. C., J. SWITALA, and B. L. TRIGGS-RAINE. 1985. Catalases HPI and HPII in *Escherichia coli* are induced independently. Arch. Biochem. Biophys. **243**:144–149.

MAYFIELD, J. E., and M. R. DUVALL. 1996. Anomalous phylogenies based on bacterial catalase gene sequences. J. Mol. Evol. **42**:469–471.

MUTO, A., and S. OZAWA. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84**:166–169.

ODENBREIT, S., B. WIELAND, and R. HAAS. 1996. Cloning and genetic characterization of *Helicobacter pylori* catalase and construction of a catalase-deficient mutant strain. J. Bacteriol. **178**:6960–6967.

OLSEN, G. J., C. R. WOESE, and R. OVERBEEK. 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176**:1–6.

ROCHA, E. R., and C. J. SMITH. 1995. Biochemical and genetic analyses of a catalase from the anaerobic bacterium *Bacteroides fragilis.* J. Bacteriol. **177**:3111–3119.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SHA, Z., T. J. STABEL, and J. E. MAYFIELD. 1994. *Brucella abortus* catalase is a periplasmic protein lacking a standard signal sequence. J. Bacteriol. **176**:7375–7377.

SHIELD, D. C., P. M. SHARP, D. G. HIGGINS, and F. WRIGHT. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5**:704–716.

SUBRAMANI, S. 1993. Protein import into peroxisomes and the biogenesis of the organelle. Annu. Rev. Cell Biol. **9**:445–478.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VON OSSOWSKI, I., G. HAUSNER, and P. C. LOEWEN. 1993. Molecular evolutionary analysis based on the amino acid sequence of catalase. J. Mol. Evol. **37**:71–76.

YANG, C.-H., J. A. MENGE, and D. A. COOKSEY. 1994. Mutations affecting hyphal colonization and pyoverdine production in pseudomonads antagonistic toward *Phytophthora parasitica.* Appl. Environ. Microbiol. **60**:473–481.

PAUL M. SHARP, reviewing editor

Accepted May 30, 1997