J Mol Evol (1996) 42:469–471

JOURNAL OF MOLECULAR EVOLUTION

Letter to the Editor

Anomalous Phylogenies Based on Bacterial Catalase Gene Sequences

J.E. Mayfield,[1] M.R. Duvall[2]

[1] Department of Zoology and Genetics, 2106 Molecular Biology, Iowa State University, Ames, IA 50011, USA
[2] Department of Biology and Microbiology, South Dakota State University, Brookings, SD 57007, USA

Received: 2 November 1995

Abstract. Phylogenies based on nine prokaryotic catalase sequences demonstrate no relationship to phylogenies based on rDNA sequences or other known criteria. When this observation is considered together with the monophyletic relationship observed for eukaryotic catalase sequences, it seems likely that the catalase gene sequence has migrated repeatedly from eukaryotes to prokaryotes.

Key words: Catalase — Phylogeny — Horizontal transfer — DNA sequence

Introduction

Catalase enzyme activity has long been considered a virtual necessity for aerobic life. In recent years, it has become clear that there are at least three gene families, unrelated at the sequence level, that encode enzymes with catalase activity. ''True catalase'' (all eukaryotic catalases, *E. coli katE*) (Schonbaum and Chance 1976; Loewen et al. 1985a) is found in essentially all eukaryotes and in many prokaryotes, while ''catalase/peroxidase'' (*E. coli katG*) (Loewen et al. 1985b) and ''Mn-catalase'' (Penner-Hahn 1995) have been reported only from prokaryotes. It was previously reported (Von Ossowski et al. 1993) that 16 eukaryotic catalase sequences were monophyletic while four bacterial sequences were not. Recently, there have been a number of new bacterial catalase sequences reported. We report here that phylogenies based on nine prokaryotic and five eukaryotic sequences demonstrate no relationship to phylogenies based on rDNA gene sequences or based on any other known criteria (Fig. 1B). Thus, trees based on bacterial catalase sequences do not demonstrate even such basic groupings as prokaryote and eukaryote, or *Firmicutes* (Gram positive) and *Gracilicutes* [*Proteobacter* (Stackebrandt et al. 1988)] (*Bergey's Manual of Systematic Bacteriology* 1989). This broad conclusion is very robust and strongly suggests that horizontal transfer of the catalase gene has occurred widely among the eubacteria. Taken together with the monophyletic pattern observed in the eukaryotes, the observed relationships suggest that the ancestral roots of the sequence lie in the eukaryotic lineage and that the sequence has repeatedly migrated horizontally into the eubacterial realm.

Methods

Catalase Sequences. Catalase coding sequences were retrieved from GenBank and translated, and both nucleotide and amino acid sequences were aligned with the ''Pileup'' program provided in the GCG (Genetics Computer Group Sequence Analysis Software Package, University of Wisconsin) sequence analysis software package, version 8.0 (Devereaux et al. 1984). The default values for GapWeight = 5.0 and GapLengthWeight = 0.3 for the DNA sequence alignments, and 3.0 and to 0.1, respectively, for amino acid sequence alignments, were used. The ''Pileup'' algorithm is based on the method of Feng and

Correspondence to: J. E. Mayfield

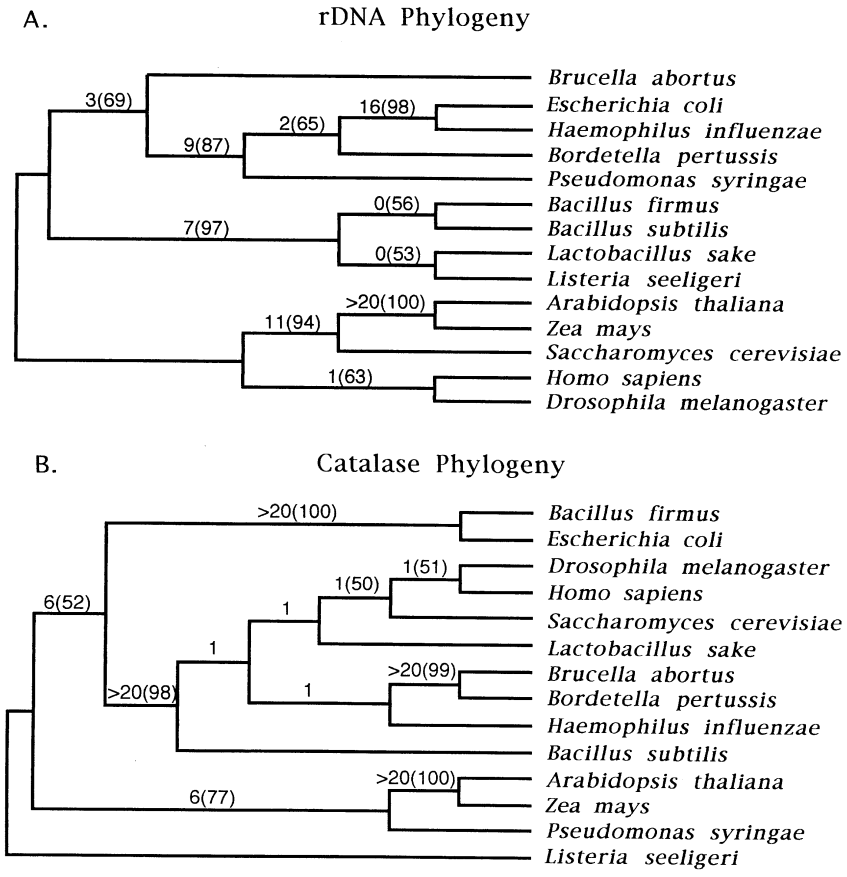

Fig. 1. **A** Bootstrap consensus tree produced on analysis of rDNA sequences. Note that this tree is identical to the strict consensus of the two equally optimal trees of length 1,242 each except that the relative position of species within the *Firmicutes* clade is unresolved in the latter. *Numbers* along the branches are decay values. *Parenthetical numbers* are bootstrap values of 50% or greater. Consistency index (CI), excluding uninformative characters: CI = 0.666; retention index (RI): RI = 0.765. **B** Single shortest tree produced on analysis of catalase nucleotide sequences. Support values are shown as for the upper tree. CI = 0.465; RI = 0.377.

Doolittle (Feng and Doolittle 1987). Following the initial alignment, a more highly conserved core region was defined and the core sequence was realigned by the same algorithm. Table 1 lists the sequences analyzed and defines the core sequences for each sequence (starting and ending nucleotide numbers are indicated in parenthesis). Nucleotide sequences of small-subunit rDNA for the same species were aligned using the same method.

Phylogenetic Analysis. Aligned catalase nucleotide sequences from the highly conserved core region were analyzed for 14 species using the Phylogenetic Analysis Using Parsimony package (PAUP) version 3.1.1 (Swofford 1993). A heuristic search was conducted with 100 randomly determined input orders of data, with global (TBR) branch swapping on all of the most parsimonious trees (MULPARS invoked) to obtain an optimal tree: 203 suboptimal trees, up to 20 steps longer than the shortest tree, were also obtained during this heuristic search to provide decay values (Bremer 1988; Donoghue et al. 1992). A bootstrap analysis (Felsenstein 1985) was also performed with 1,000 subsamplings of the original data matrix, 4 replicate searches per subsample, and with the TBR and MULPARS options. Aligned amino acid sequences for the same conserved core region and the same species were analyzed in a parallel search.

Aligned nucleotide sequences from the small-subunit rDNA (16S or 18S) were analyzed by the same methods for the same species; 725 suboptimal trees up to 20 steps longer than the two equally optimal trees were obtained in the decay analysis.

Table 1. Sequences analyzed/core sequences defined

Species	Core nucleotides	Accession #
Bacillus firmus	(82–1,183)	L02551
Bacillus subtilis	(19–1,105)	M80796
Bordetella pertussis	(28–1,114)	U07800
Brucella abortus	(22–1,105)	U11439
Escherichia coli	(241–1,342)	M55161
Haemophilus influenzae	(46–1,132)	U02682
Lactobacillus sake	(16–1,102)	M84015
Listeria seeligeri	(22–1,108)	M75944
Pseudomonas syringae	(91–1,168)	U03465
Arabidopsis thaliana	(52–1,137)	X64271
Drosophila melanogaster	(76–1,164)	X52286
Homo sapiens	(82–1,171)	X04076
Saccharomyces cerevisiae	(67–1,161)	X13028
Zea mays	(52–1,138)	M33104

Results

Catalase sequences from all reported sources exhibit a highly conserved core sequence and highly variable N-terminal and C-terminal sequences. Pairwise alignments show that within the core sequence region, prokaryotic

nucleotide sequence identities range from 70% (with one gap) when *Brucella abortus* and *Bordetella pertussis* are compared to 51% (with three gaps) when *B. abortus* and *Bacillus firmus* are compared. For amino acid sequences, *B. firmus* and *Escherichia coli* exhibit the highest sequence identity within the core region, 75% (with no gaps), and *E. coli* and *Pseudomonas syringae* the lowest, 47% sequence identity (with four gaps).

Phylogenetic trees derived from rDNA and catalase nucleotide sequences were compared for the same set of species (Fig. 1). Note the contrasting topologies and associated measures of support of these two trees. In the rDNA tree, a strongly supported (decay: 7; bootstrap: 97) clade of *Firmicutes* is distinguished from a clade of exclusively Gram-negative species. In this tree, eukaryotes can be specified as an outgroup having a single common ancestry. The topology of the catalase tree is markedly different and anomalous. For example, a clade consisting of *Bacillus firmus*, a Gram-positive species, and *Escherichia coli*, a Gram-negative species, is very strongly supported (decay: >20; bootstrap: 100). Note that the second *Bacillus* species, *B. subtilis*, occupies a second clade consisting of a mixture of eukaryotes and Gram-positive and Gram-negative species. Eukaryotes cannot be forced into a monophyletic outgroup arrangement in the optimal catalase tree since they occupy derived positions of nonsibling clades. Phylogenetic analysis of catalase amino acid sequences gives a similar anomalous tree topology (results not shown).

Discussion

The horizontal transfer of useful genes between distantly related organisms represents an evolutionary opportunity. The apparent advantage to an organism which could acquire useful gene functions from other organisms seems to be so great that it is a wonder phylogenies based on single genes contain any useful information about the long-term evolutionary history of organisms. Yet, clearly they often do, especially when genes essential to the normal metabolism of the cell are examined (Woese 1987; Avise 1994). Catalase activity is often described as essential for aerobic life, but only recently has it been realized that there are at least three unrelated sequence families that code for enzymes with catalase activity. These are ''true catalase'' (Schonbaum and Chance 1976), ''catalase/peroxidase'' (Triggs-Raine et al. 1988), and ''Mn-catalase'' (Penner-Hahn 1995). All three sequence families are found in the prokaryotic world, while only ''true catalases'' have been found in eukaryotes. There are currently three known organisms with Mn-catalase activity (Shank et al. 1994), and none of these DNA sequences has been reported. Catalase/peroxidase is very widespread, and there are currently six catalase/peroxidase sequences reported in GenBank. Distance analysis of these sequences is in reasonable agreement with accepted phylogenies (data not shown). This observation taken together with the nonsensical relationships reported in this paper for the ''true catalases'' may indicate that the catalase/peroxidase gene family represents an ancient eubacterial mechanism for protecting cells from hydrogen peroxide while the true catalases originated early in the eukaryotic lineage and spread horizontally into the Eubacteria after the appearance of the plant and animal kingdoms. If this conclusion is accepted, the analysis reported here requires that the migration of the catalase sequence from eukaryotes to prokaryotes have occurred multiple times in evolutionary history. At this time, we offer no additional evidence to support this hypothesis, but simply suggest that it is the simplest explanation for the bizarre relationships demonstrated in this paper.

References

Avise JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York, pp 352–354

Bergey's manual of systematic bacteriology, 8th ed (1989) Williams & Wilkins, Baltimore, pp 31–35

Bremer K (1988) The limits of amino acid sequence data in Angiosperm phylogenetic reconstruction. Evolution 42:795–803

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

Donoghue M, Olmstead J, Smith J, Palmer J (1992) Phylogenetic relationships of Dipsacales based on rbcL sequences. Ann Missouri Bot Gard 79:333–345

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

Feng D, Doolittle RF (1987) Progressive sequence alignments as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Loewen PC, Switala J, Triggs-Raine BL (1985a) Catalases HPI and HPII in *Escherichia coli* are induced independently. Arch Biochem Biophys 243:144–149

Loewen PC, Triggs BL, George CS, Hrabarchuk BK (1985b) Genetic mapping of *katG*, a locus that affects synthesis of the bifunctional catalase-peroxidase hydroperoxidase I in *Escherichia coli*. J Bacteriol 162:661–667

Penner-Hahn JE (1995) Structural properties of the manganese site in the manganese catalases. In: Pecoraro VL (ed) Manganese redox enzymes. Verlag Chemie, New York, pp 29–45

Schonbaum GR, Chance B (1976) Catalase. In: Boyer PD (ed) The Enzymes, vol XII, pt C, 3rd ed. Academic Press, New York, pp 363–408

Shank M, Barynin V, Dismukes GC (1994) Protein coordination to manganese determines the high catalytic rate of dimanganese catalases. Comparison to functional catalase mimics. Biochem 33: 15433–15436

Stackebrandt E, Murray RGE, Truper HG (1988) *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the ''purple bacteria and their relatives.'' Int J Syst Bacteriol 38:321–325

Swofford D (1993) PAUP: phylogenetic analysis using parsimony, version 3.1 computer program. Illinois Natural History Survey, Champaign

Triggs-Raine BL, Doble BW, Mulvey MR, Sorby PA, Loewen PC (1988) Nucleotide sequence of *katG*, encoding catalase HPI of *Escherichia coli*. J Bacteriol 170:4415–4419

Von Ossowski I, Hausner G, Loewen PC (1993) Molecular evolutionary analysis based on the amino acid sequence of catalase. J Mol Evol 37:71–76

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271