Horizontal transfer of catalase-peroxidase genes between Archaea and pathogenic bacteria

David M. Faguy*
dfaguy@unm.edu

W. Ford Doolittle
ford@is.dal.ca

Canadian Institute for Advanced Research Program in Evolutionary Biology and Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, N.S., Canada.
*Department of Biology, University of New Mexico, Albuquerque, NM 87131, Canada.

The data from genome sequencing projects has had an immediate and major impact on our view of prokaryotic evolution: it has revealed the importance and prevalence of lateral gene transfers (LGTs), which have been discussed in many recent analyses and commentaries[1–7]. LGT, by implication, challenges the very existence of an evolutionary classification of prokaryotes. Other implications are less clear but, undoubtedly, we must consider lateral transfer when examining the evolutionary history of any prokaryote. We describe here evidence for transfer between Archaea and Bacteria of catalase-peroxidase genes involved in oxidation protection. This transfer might have been important in the virulence development of the major human pathogens *Escherichia coli* O157, *Yersinia pestis,* and *Legionella pneumophila*.

The catalase-peroxidase family of enzymes is involved in removing H_2O_2. They are bifunctional enzymes; capable of either reducing H_2O_2 with an external reductant (peroxidase activity) or disproportionating it to H_2O and O_2 (catalase activity). This enzyme family is found in relatively few organisms (all but one of these organisms are prokaryotic – no sequence is known from the single report there is of a eukaryotic catalase-peroxidase[8]). Catalase-peroxidase proteins are a distinct sub-class of the heme peroxidase superfamily of enzymes[9] and share no significant sequence similarity to proteins of the ubiquitous classical catalase family that is found in bacteria and eukaryotes and in one archaeon.

There are currently ~20 known catalase-peroxidases (most from complete genome sequences). Of the archaeal sequences, only three are known: from *Halobacterium salinarum, Haloarcula marismortui* and *Archaeoglobus fulgidus*. They have been identified in several diverse bacteria, including enterics, high and low GC Gram positive bacteria, and cyanobacteria. Among the Bacteria, there is no obvious phylogenetic pattern to this distribution; some organisms have the gene, whereas related organisms do not (e.g. it is present in *Bacillus stearothermophilus* but not in the complete genome of *Bacillus subtilis*). This enzyme has been implicated as a virulence factor in *Mycobacteria*[10], but has not been directly implicated in virulence in other pathogens. However, circumstantial evidence suggests that it could be important for virulence in *E. coli* O157 and *L. pneumophila*[11,12] – both of which have two copies. In *E. coli*, there is a chromosomal copy (present in all strains) and a copy on the 100 kb O157 plasmid (present only in the virulent O157 strain). This O157 catalase-peroxidase has been associated with enterohaemorrhagic hemolysin in shiga-like-toxin-producing *E. coli*[13]. This scenario is not without precedents: Fang *et al.*[14] recently reported that virulent *Salmonella typhimurium* has two dissimilar superoxide dismutases (also involved in oxidation protection).

We have performed a phylogenetic analysis of known catalase-peroxidases and find the surprising result that the gene from three human pathogens (*E. coli* O157, *L. pneumophila*, and *Y. pestis*) appear to be most closely related to those of Archaea. Figure 1 shows the results of phylogenetic analyses using three methods (distance, parsimony, and maximum likelihood). All three methods agree on the tree topology with moderate to strong statistical support (bootstraps). The most logical interpretation of this is that the gene for catalase-peroxidase was transferred between an archaeon and these pathogens. We don't believe this is an artefact of phylogenetic methods because there is statistical support, different methods agree, and the sequences are all highly conserved (there are no long branches – known to be a major source of artefacts) and this sequence conservation reduces the possibility of an alignment error resulting in an incorrect phylogenetic pattern. Also, except for the three putatively transferred genes, the tree is consistent with other data (i.e. the phylogeny of the Archaea and high GC Gram-positive bacteria are accurately recovered).

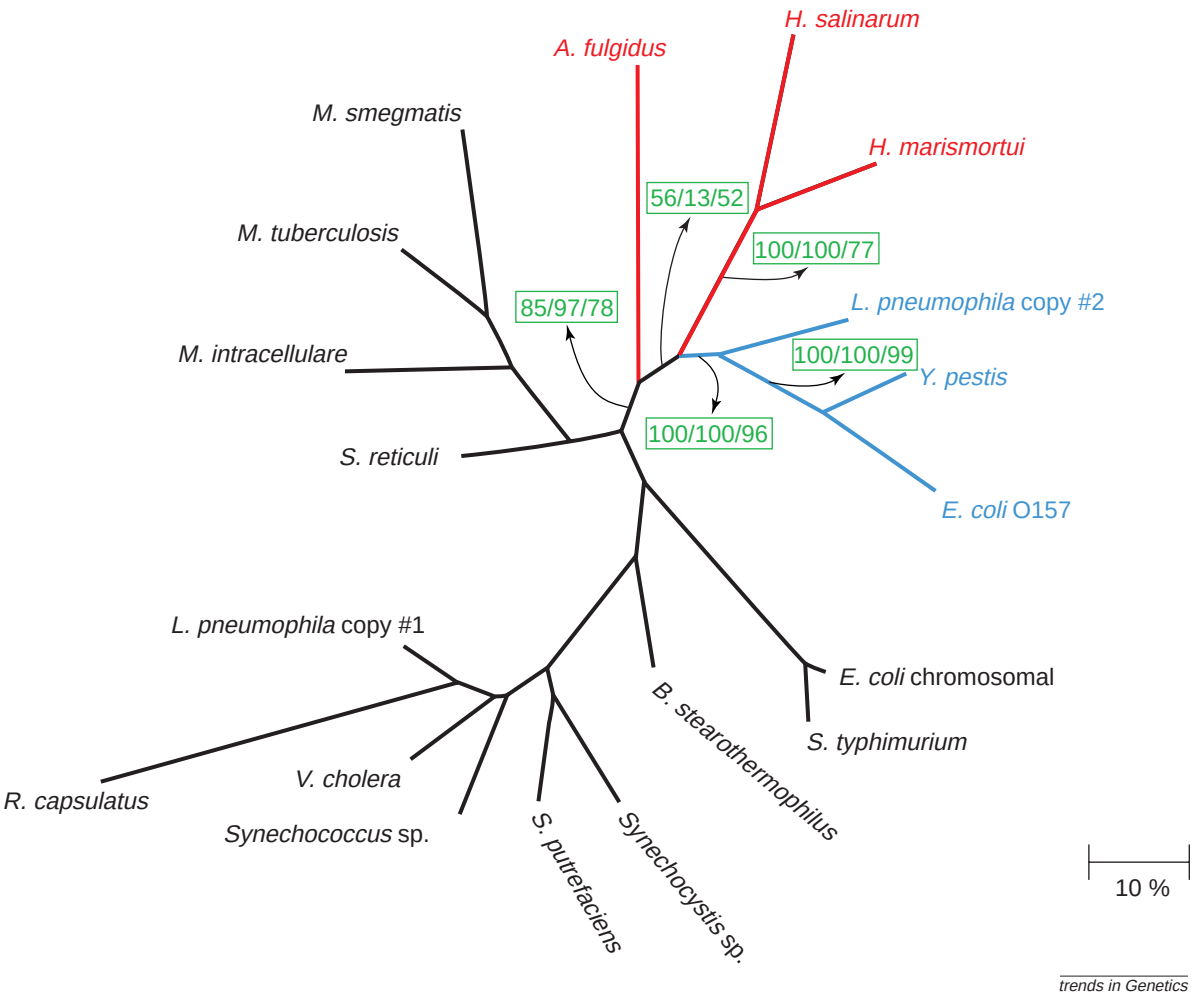
An alternative interpretation to lateral gene transfer is a series of gene duplications and/or losses. The scattered distribution in Bacteria, in the absence of lateral transfers, would require loss from a great number of distinct lineages (we know from complete genomes that more than 25 bacteria do not have a homolog of this gene). Similarly, the presence of two clearly dissimilar catalase-peroxidases in *E. coli* O157 and *L. pneumophila* in a duplication/loss scenario would require an ancient duplication and subsequent loss in all other lineages except the Archaea. Clearly, lateral gene transfer is a much simpler interpretation.

We propose that a catalase-peroxidase gene was transferred from an archaeon to a pathogenic bacterium; either directly or through an intermediate with more frequent physical contact with Archaea. The presence of two dissimilar catalase-peroxidases in *E. coli* and *L. pneumophila* strongly suggest they were on the receiving end of a lateral transfer. The catalase-peroxidase gene on the large virulence plasmid of *E. coli* O157 is flanked by transposases (which occur frequently in the plasmid) and is near a large open reading frame with high similarity to toxin B previously found only in *Clostridium difficile*[15,16].

Lateral transfer of pathogenicity islands and antibiotic resistance genes have long been known to be an important factor in the evolution of infectious disease, but few would expect non-pathogenic 'extremophiles' to provide a source of virulence genes. The lateral transfer discussed here might force us to look far afield in search of the source of disease-enhancing genes and reinforces our need to understand the biology of the Archaea.

 PII: S0168-9525(00)02007-2

FIGURE 1. Phylogenetic tree of catalase-peroxidases



Three human pathogen sequences (*Escherichia coli* O157, *Legionella pneumophila*, *Yersinia pestis* – coloured blue for emphasis) clearly group within the Archaea (red), rather than Bacteria (black). This tree was the most likely tree returned from a quick search using PROTML at the website http://BIOWEB.PASTEUR.FR/; it was also the shortest distance tree and the best tree (with two others) by parsimony. Bootstrap support (distance/parsimony/maximum likelihood; green) is shown at important nodes. Parsimony and distance phylogenetic analyses were performed with programs from the PHYLIP package, maximum likleihood analyses were performed with PUZZLE (Ref. 17) and PROTML via the above web site. Branch lengths on the tree are maximum likelihood branch lengths returned by PUZZLE. The amino acid sequences of catalase-peroxidases from GenPept plus those obtained from the unfinished complete genomes (http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html) of *Shewanella putrefaciens*, *Vibrio cholera* and *L. pneumophila* were used in the analysis. The *V. cholerae* sequence was compiled from two separate contigs, overlapping by only a few nucleotides and could represent two distinct paralogs (although both ends give similar branching patterns). This represents all the complete sequences except for duplicates, closely related organisms (*Mycobacteria* and *Bacillus stearothermophilus*), and the sequence from *Caulobacter crescentus*. The *C. crescentus* sequence was omitted because it is shorter. Including it and using a partial alignment did not affect the resulting tree topology of archaeal/pathogen sequences but did reduce overall statistical support. The alignment used was constructed initially by the CLUSTALW program, but was adjusted in three instances by deleting questionable regions. It consisted of 19 taxa with 787 sites. The alignment, sequence accession numbers, references and a table of maximum likelihood branch lengths are available at http://www.unm.edu/~dfaguy/tigsupp.htm. The scale bar represents 10% estimated sequence divergence. *H. salinarum*, *Halobacterium salinarum*; *H. marismortui*, *Haloarcula marismortui*; *A. fulgidus*, *Archeolglobus fulgidus*; *M. smegmatis*, *Mycobacterium smegmatis*; *M. tuberculosis*, *Mycobacterium tuberculosis*; *M. intracellulare*, *Mycobacterium intracellulare*; *R. capsulatis*, *Rhodobacter capsulatus*; *S. typhimurium*, *Salmonella typhimurium*; *S. reticuli*, *Streptomyces reticuli*.

References

1 Lake, J.A. *et al.* (1999) Mix and match in the tree of life. *Science* 283, 2027–2028
2 Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2129
3 Jain, R. *et al.* (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3801–3806
4 Lawrence, J.G. (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* 2, 519–523
5 Lawrence, J.G. and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9413–9417
6 Aravind, L. *et al.* (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* 14, 442–444
7 Nelson K.E. *et al.* (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329
8 Fraaije, M.W. *et al.* (1996) Purification and characterization of an intracellular catalase-peroxidase from *Penicillium simplicissimum*. *Eur. J. Biochem.* 235, 192–198
9 Welinder, K.G. (1991) Bacterial catalase-peroxidases are gene duplicated members of the plant peroxidase superfamily. *Biochim. Biophys. Acta* 1080, 215–20
10 Zhang, Y. *et al.* (1992) The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* 358, 591–593
11 Bandyopadhyay, P. and Steinman, H.M. (1998) *Legionella pneumophila* catalase-peroxidases: cloning of the katB gene and studies of KatB function. *J. Bacteriol.* 180, 5369–5374
12 Garcia, E. *et al.* (1999) Molecular characterization of KatY (antigen 5), a thermoregulated chromosomally encoded catalase-peroxidase of *Yersinia pestis*. *J. Bacteriol.* 181, 3114–3122
13 Brunder, W. *et al.* (1996) KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* 142, 3305–3315
14 Fang, F.C. *et al.* (1999) Virulent Salmonella typhimurium has two periplasmic Cu, Zn-superoxide dismutases. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7502–7507
15 Makino, K. *et al.* (1998) Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res.* 5, 1–9
16 Burland, V. *et al.* (1998) The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* 26, 4196–4204
17 Strimmer, K. and von Haeseler, A. (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969