# A Single Eubacterial Origin of Eukaryotic Pyruvate:Ferredoxin Oxidoreductase Genes: Implications for the Evolution of Anaerobic Eukaryotes

*David S. Horner, Robert P. Hirt, and T. Martin Embley*

Department of Zoology, Natural History Museum, London, England

The iron sulfur protein pyruvate:ferredoxin oxidoreductase (PFO) is central to energy metabolism in amitochondriate eukaryotes, including those with hydrogenosomes. Thus, revealing the evolutionary history of PFO is critical to understanding the origin(s) of eukaryote anaerobic energy metabolism. We determined a complete PFO sequence for *Spironucleus barkhanus,* a large fragment of a PFO sequence from *Clostridium pasteurianum,* and a fragment of a new PFO from *Giardia lamblia.* Phylogenetic analyses of eubacterial and eukaryotic PFO genes suggest a complex history for PFO, including possible gene duplications and horizontal transfers among eubacteria. Our analyses favor a common origin for eukaryotic cytosolic and hydrogenosomal PFOs from a single eubacterial source, rather than from separate horizontal transfers as previously suggested. However, with the present sampling of genes and species, we were unable to infer a specific eubacterial sister group for eukaryotic PFO. Thus, we find no direct support for the published hypothesis that the donor of eukaryote PFO was the common α-proteobacterial ancestor of mitochondria and hydrogenosomes. We also report that several fungi and protists encode proteins with PFO domains that are likely monophyletic with PFOs from anaerobic protists. In *Saccharomyces cerevisiae,* PFO domains combine with fragments of other redox proteins to form fusion proteins which participate in methionine biosynthesis. Our results are consistent with the view that PFO, an enzyme previously considered to be specific to energy metabolism in amitochondriate protists, was present in the common ancestor of contemporary eukaryotes and was retained, wholly or in part, during the evolution of oxygen-dependent and mitochondrion-bearing lineages.

## Introduction

Strict and aerotolerant anaerobic microbial eukaryotes (protists) have featured prominently in scenarios for early eukaryote evolution (Doolittle 1998; Martin and Müller 1998). As a consequence, there has been much interest in contemporary protists which occupy low-oxygen habitats, because they may have preserved relics of an ancient anaerobic eukaryotic past. A variety of phylogenetically distinct protists live under anaerobic or microaerophilic conditions (Fenchel and Finlay 1995). Parabasalids, some ciliates, and some chytrid fungi lack mitochondria but contain specialized redox organelles called hydrogenosomes (Müller 1998) which ferment pyruvate and produce hydrogen. Others, including diplomonads like *Giardia lamblia* and the amoeba *Entamoeba histolytica,* lack hydrogenosomes or mitochondria and ferment pyruvate in the cytosol (Müller 1998). Both types of anaerobe use the oxygen-sensitive iron sulfur protein pyruvate:ferredoxin oxidoreductase (PFO) to decarboxylate pyruvate to acetyl-CoA (Müller 1998). In aerobic eukaryotes, this reaction is catalyzed by the nonhomologous pyruvate dehydrogenase complex (PDH) in mitochondria (Kerscher and Oesterhelt 1982; Müller 1998).

The presence of PFO is central to anaerobic metabolism in different eukaryotes (Müller 1998), which raises the question of its evolutionary origins. Early on, it was suggested that PFO might have been supplanted by PDH from the protomitochondrion (Kerscher and Oesterhelt 1982). However, recent data suggest that all of the eukaryotes which contain PFO also contain genes of mitochondrial origin (Clark and Roger 1995; Bui, Bradley, and Johnson 1996; Germot, Philippe, and Guyader 1996; Horner et al. 1996; Roger, Clark, and Doolittle 1996; Hashimoto et al. 1998; Roger et al. 1998), suggesting that they once contained mitochondria. The ‘‘hydrogen hypothesis’’ (Martin and Müller 1998) for the contemporaneous origin of eukaryotes and the acquisition of the mitochondrion symbiont posits that PFO and PDH were both donated by the protomitochondrion (see also Embley, Horner, and Hirt 1997). Under this hypothesis, PFO genes have been preserved during the evolution of aerobic eukaryotes, so as to be recruitable in disparate lineages during a reversion to anaerobic energy metabolism. This hypothesis predicts that eukaryotic PFOs from contemporary eukaryotes will be monophyletic and root among proteobacteria (Yang et al. 1985) in a position consistent with an origin from the protomitochondrion.

Alternative hypotheses for the origin of eukaryote PFO posit one or more gene transfer(s) from prokaryotes other than the protomitochondrion, allowing repeated ‘‘reinvention’’ of an anaerobic phenotype based on PFO. For example, by analogy with the mitochondrion and chloroplast symbioses, a symbiosis between an ancestor of *Trichomonas* and an anaerobic bacterium similar to contemporary clostridia has been proposed as the source of the parabasalid hydrogenosome and its biochemistry (Whatley, John, and Whatley 1979). However, recent data are more consistent with a common symbiotic origin for mitochondria and hydrogenosome organelles (reviewed in Embley, Horner, and Hirt 1997),

Abbreviations: ML, maximum likelihood; MP, maximum parsimony; PFO, pyruvate:ferredoxin oxidoreductase.

Key words: pyruvate:ferredoxin oxidoreductase, anaerobic eukaryotes, molecular phylogeny, hydrogenosomes, mitochondria, *Saccharomyces cerevisiae.*

Address for correspondence and reprints: T. Martin Embley, Department of Zoology, Natural History Museum, London SW7 5BD, United Kingdom. E-mail: tme@nhm.ac.uk.

*Mol. Biol. Evol.* 16(9):1280–1291. 1999
© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038



Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

although this need not exclude a clostridial origin for hydrogenosomal PFO.

Phylogenetic analysis can potentially distinguish between competing hypotheses for the origin of PFO in different eukaryotes. Rosenthal et al. (1997) analyzed eight PFO sequences, including those from the anaerobic eukaryotes *Entamoeba histolytica, Giardia lamblia,* and *Trichomonas vaginalis.* They concluded that there was little support for a common ancestry of eukaryote PFO and that *Entamoeba* probably got its PFO independent of the other eukaryotes, and possibly from an enterobacterium.

Here, we present a complete PFO sequence for the diplomonad *Spironucleus barkhanus,* a large fragment of a PFO sequence from *Clostridium pasteurianum,* and a fragment of a new PFO from *Giardia lamblia.* Phylogenetic analyses of eubacterial and eukaryotic PFO genes suggest a complex history for PFO, including possible gene duplications and horizontal transfers among eubacteria. Our analyses favor a common origin for eukaryotic cytosolic and hydrogenosomal PFOs from a single eubacterial source, but we are unable to identify from which eubacterial group eukaryote PFO might have originated. We also report that several fungi and protists encode proteins with PFO domains that are likely monophyletic with the available PFO genes from anaerobic protists.

## Materials and Methods

### Strains, Culture, and Nucleic Acid Extractions

*Spironucleus barkhanus* (ATCC strain 50380) (Rozario et al. 1996) was grown following the ATCC protocol. *Clostridium pasteurianum* (National Collection of Food Bacteria: catalog number 1845) was grown anaerobically in reinforced clostridial medium at 37°C. *Giardia lamblia* biomass was a kind gift of Prof. Graham Coombs (University of Glasgow, Scotland). DNA was extracted from *G. lamblia, S. barkhanus* and *C. pasteurianum* using standard protocols. Messenger RNA from approximately $10^8$ cells of *S. barkhanus* was isolated using the Dynabead mRNA system (Dynal), and cDNA was synthesised and cloned into Lambda ZAP XR (Stratagene). *Spironucleus barkhanus* DNA partially digested with *Bam*HI was used to make a genomic library in Lambda DASHII (Stratagene).

### Cloning of a PFO Gene from *S. barkhanus* and Additional PFO Coding Sequences from *C. pasteurianum* and *G. lamblia*

Published PFO and PFO-encoding *NifJ*genes of eubacteria were used to design PCR primers 551F (5′-CGAYGGYTTYCGYACNTCNCA-3′), 1400R (5′-CTTRTTRGCACCRACRGTACC-3′), and 3175R (5′-CRATRTCRTANGCCCANCCRTC-3′). A fragment of approximately 900 bp was amplified from *S. barkhanus* DNA with primers 551F and 1400R and used to screen the cDNA library. The longest clone isolated was sequenced on both strands using primer walking. The sequence of the cDNA lacked about 15 nt from its 5′ terminus. The 5′ end terminal region of the gene was established from an overlapping genomic clone.

Primers 551F and 3175R amplified a fragment of approximately 2.5 kb from *C. pasteurianum* DNA. This product was cloned in pGEMT (Promega), and two clones were sequenced on both strands. Alignment of the conceptually translated *C. pasteurianum* sequence with reference taxa revealed the most similarity to *NifJ* genes of *Klebsiella pneumoniae* and *Enterobacter agglomerans.* Further degenerate PCR primers, C.pas 5′ PFOf (5′-AARHTBAARACWATGSAWGGNAA-3′) and C.pas 3′ PFOr (5′-CCARTAWCCWGCWTCWACW-G-3′) were designed to conserved regions in the 5′ and 3′ ends of the *Klebsiella* and *Enterobacter* coding regions and used with specific primers designed from the original PCR product to amplify almost the entire 5′ region and to within 200 bp of the expected 3′ end of the gene.

Examination of preliminary sequence data from the *G. lamblia* Genome Project at The Josephine Bay Paul Center Web site at the Marine Biological Laboratory (www.bpc.mbl.edu) suggested that *Giardia* encodes a second PFO in addition to the one already published (GenBank accession number L27221; see Rosenthal et al. 1997). Specific PCR primers Gl.PFO2:190F (5′-GATGAACTCCGAGGGGCTGCTG-3′) and Gl.PFO2:1500R (5′-CAGTCTGGAGGGCGGAGTTTATC-3′) were used to amplify a fragment of 1,539 bp, which was cloned into pGEMT (Promega). Two clones were sequenced on both strands. To distinguish between this PCR product and the published *G. lamblia* PFO gene sequence, we refer to the published sequence as *G. lamblia* PFO1 and to the gene represented by our PCR product as *G. lamblia* PFO2.

### Sequence Recovery and Alignment

We searched finished and unfinished prokaryote and eukaryote genome projects to identify PFO coding sequences (see legends to figs. 2 and 3) using prokaryote and eukaryote PFOs as probes and the BLASTn, t-BLASTn and iterative (PSI) BLAST programs of the BLAST2 package (Altschul et al. 1997).

Inferred PFO amino acid sequences were aligned using CLUSTAL W (Thompson, Higgins, and Gibson 1994) and adjusted manually using GDEv2.2 (Maidak et al. 1996). DNA coding regions were back-aligned to the protein alignment using PUTGAPS (J. O. McInerney, Natural History Museum, London). A mask was used to exclude regions which could not be unambiguously aligned, leaving 743 aligned amino acid positions for phylogenetic analyses. A reduced alignment accommodating the partial *G. lamblia* PFO2 sequence (378 aligned amino acids) was also used. Alignments are available from D.S.H. (e-mail: dsh@nhm.ac.uk).

### Phylogenetic Analyses of PFO Protein Sequences

Maximum-likelihood (ML) analyses of protein alignments were performed using PROTML in MOLPHY, version 2.3 (Adachi and Hasegawa 1996), using the heuristic quick-add OTU search and local rearrangement (R-search) methods with the JTT-F amino acid replacement model. However, the ML models in PROTML make the assumption that all sites are free to

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

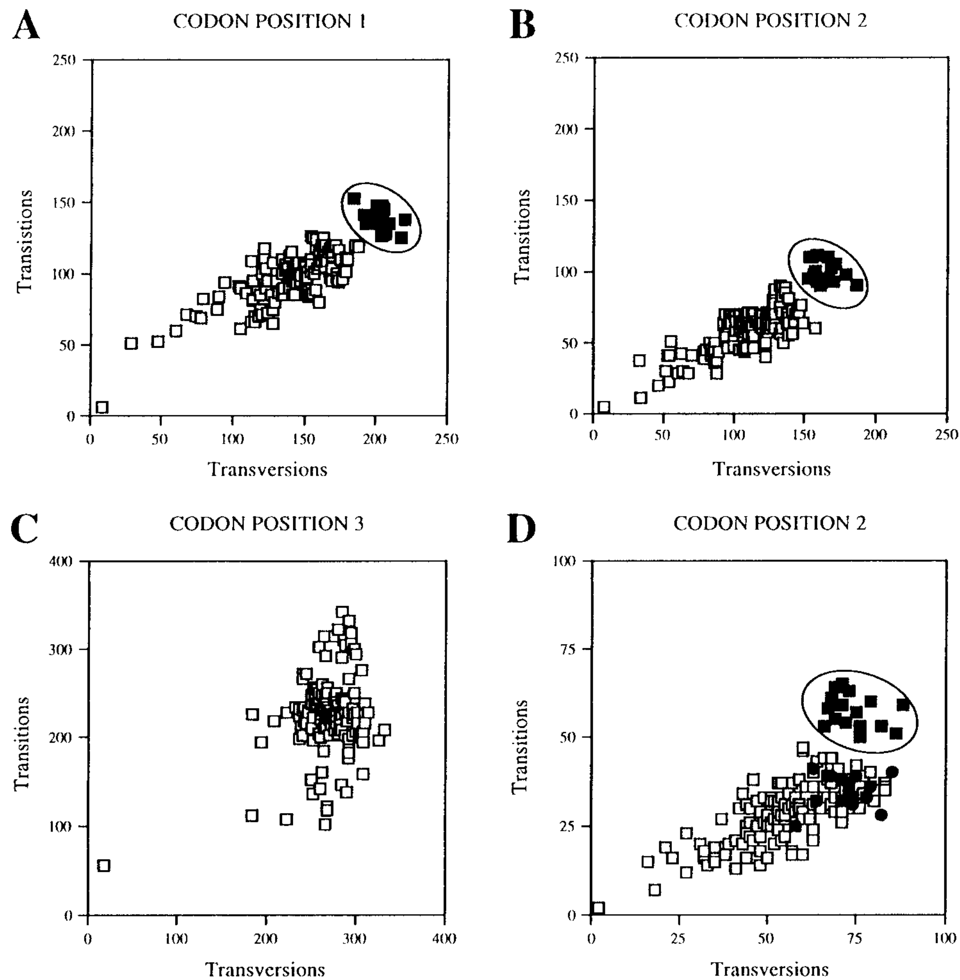


FIG. 1.—Plots of observed transitions versus transversions by codon position for all pairwise comparisons. *A–C,* The comparisons use the DNA alignment for 743 amino acids and 18 sequences. Position 3 shows extensive clustering of transitions versus transversions, indicating potential mutational saturation. Plots for codon positions 1 and 2 show a linear trend, with the exception of points involving *G. lamblia* PFO1 (filled boxes circled for codon positions 1 and 2). These points are clustered outside of the distribution of other pairwise comparisons, indicating that mutational saturation may be problematic for comparisons involving *Giardia* PFO1 (see text). *D,* Plot of transitions versus transversions for codon position 2 for a reduced alignment (378 codon positions) including the *G. lamblia* PFO2 sequence. Points involving *G. lamblia* PFO1 (filled boxes) are clustered (as in *B*), whereas comparisons involving *G. lamblia* PFO2 (filled circles) lie within the distribution of other pairwise comparisons.

vary, and when this assumption is violated, the wrong tree may be selected (Shoemaker and Fitch 1989; Lockhart et al. 1996; Hirt et al. 1999). The PFO sequences contain sites which are of unvaried amino acid composition across the alignment and are thus potentially unable to vary (i.e., invariable). We therefore removed the proportion (80%) of unvaried sites, inferred to be invariable using a codon capture/recapture method (Sidow, Nguyen, and Speed 1992). Protein ML bootstrap analyses of variable sites (629 sites for 18 taxa) were performed on 100 resampled data sets as previously described (Hirt et al. 1999) using the program PHYCON (J. O. McInerney, Natural History Museum, London).

## Phylogenetic Analyses of PFO DNA Sequences

Potential mutational saturation of PFO DNA sequences and nucleotide base composition heterogeneities between sequences were investigated using plots of transitions versus transversions and "GC trees" as previously described (Lockhart et al. 1994; Lento et al. 1995; Charleston 1998; Hirt et al. 1999). Codon position 3 was potentially saturated for all comparisons (fig. 1*C*) and manifested large base composition variation between sequences (18%–90% G+C), so we excluded it from phylogenetic analyses. Plots of transitions versus transversions for codon positions 1 and 2 gave approximately linear relationships for most pairwise comparisons, apart from those involving the published *G. lamblia* PFO1 sequence, which were clustered outside of the range of the other pairwise comparisons (fig. 1*A* and *B*). These results suggest that, with the possible exception of *Giardia* PFO1, mutational saturation should not be a severe problem for phylogenetic reconstruction using codon positions 1 and 2. Nucleotide compositional heterogeneity at variable positions was moderate at codon position 1 (50%–62% G+C) and at position 2 (41%–46% G+C). In nucleotide composition "GC trees" (Lockhart et al. 1994) for codon positions 1 and

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

FIG. 2.—LogDet bootstrap consensus tree of variable sites of DNA codon positions 1 and 2. Ninety-five percent of unvaried sites were estimated to be invariable using ML with a general time reversible model and two rate categories (variable/invariable), with all parameters estimated in PAUP* based on the MP tree. Bootstrap support is shown at each node for 1,000 replicates. The scale shown for branch length corresponds to estimated substitutions per site. PFO homologs from *Clostridium acetobutylicum, Yersinia pestis,* and *Porphyromonas gingivalis,* were recovered through the NCBI (www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html).

2, the eukaryote sequences were not monophyletic, suggesting that shared compositional bias is not responsible for bringing these sequences together in subsequent phylogenetic analyses.

Because we observed moderate base composition heterogeneity at codon positions 1 and 2, we used LogDet/Paralinear distances, a phylogenetic method in PAUP*4.1b (Swofford 1998) which performs well in simulations even when base composition varies across the tree (Lake 1994; Lockhart et al. 1994). However, the LogDet distance method does not incorporate a correction for site by site rate variation, and it assumes that all sites are free to vary. We therefore used PAUP* to estimate the fraction of sites free to vary across our alignment using a variable/invariable ML model, and we restricted LogDet analyses to only these variable positions (Waddell and Steel 1997). LogDet bootstrap analyses (1,000 replicates) and ML analysis of DNA codon positions 1 and 2 with a general time reversible model incorporating a discrete approximation to the gamma distribution (one invariable site category and eight variable rate categories) were performed in PAUP*.

### Testing Published Hypotheses for the Origins of Eukaryote PFO using ML Difference Tests

We used maximum-parsimony (MP) constrained analyses for both the DNA and the protein data sets to find the shortest trees consistent with published hypotheses for the origin(s) of eukaryote PFO from different eubacteria (Whatley, John, and Whatley 1979; Rosenthal et al. 1997; Martin and Müller 1998). These trees were supplied as user trees for Kishino-Hasegawa (1989) tests of log-likelihood differences relative to protein and DNA ML trees. DNA tree likelihood values were calculated in PAUP*. Protein tree likelihood scores were calculated in Puzzle using a discrete approximation to the gamma distribution (one invariable and eight variable rate categories) with parameters optimized on the ML tree.

## Results

### *Spironucleus barkhanus, G. lamblia,* and *C. pasteurianum* PFO Genes and Inferred Protein Sequences

The gene sequences for *C. pasteurianum* and *S. barkhanus* each encode homodimeric PFOs similar to known eukaryote enzymes and to those found in some eubacteria, but unlike the multiple subunit forms so far characterized from archaebacteria and from some other eubacteria (Müller 1998). Like all homodimeric PFOs, the new *S. barkhanus* and *Clostridium* sequences contain two typical 4Fe-4S coordinating motifs (CxxCxxCxxxCP), four additional conserved cysteine residues, and a thiamine pyrophosphate (TPP) binding domain (Kletzin and Adams 1996). Conceptual translation of the *S. barkhanus* sequence revealed several potential internal stop codons (TAA, TAG) which mainly corresponded to conserved glutamines in the alignment. This use of a noncanonical code has previously been reported for *Spironucleus* (Keeling and Doolittle 1997). Conceptual translation of the *G. lamblia* PFO2 PCR product yielded a 513-amino-acid region with 59% identity to the *Spironucleus* PFO gene but only 33% identity to the *G. lamblia* PFO1.

### Phylogenetic Analysis of Homodimeric PFO Sequences

Protein ML analysis of variable sites recovered a tree (not shown) with three clusters (labeled I, II, and III) of identical compositions to those seen in the LogDet tree (fig. 2). Bootstrap support for eukaryote monophyly (cluster III) was 70%, increasing to 75% when the *Giardia* PFO1 sequence—the longest branch

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

**Table 1**
**Significance of Differences in Log-Likelihood Values for Alternative Hypotheses of Relationships**

| Tree/Constraint | DNA ML | | | Protein ML | | |
|---|---|---|---|---|---|---|
| | −lnL | Diff lnL ± SE | Significant? | −lnL | Diff lnL ± SE | Significant? |
| ML tree | 16,471.5 | Best tree | | 14,985.7 | Best tree | |
| MP tree | 16,481.8 | 10 ± 13 | No | 14,994.4 | 9 ± 8 | No |
| (Eukaryotes + *Desulfovibrio africanus*) | 16,487.3 | 15 ± 13 | No | 15,003.5 | 18 ± 12 | No |
| (Eukaryotes + *Rhrodospirillum rubrum*) | 16,580.0 | 108 ± 22 | Yes | 15,116.1 | 130 ± 27 | Yes |
| (Eukaryotes + *Escherichia coli*/*Yersinia pestis*) | 16,529.4 | 57 ± 16 | Yes | 15,074.8 | 89 ± 25 | Yes |
| (Eukaryotes + *Klebsiella pneumoniae*/*Enterobacter agglomerans*) | 16,537.6 | 66 ± 22 | Yes | 15,093.0 | 107 ± 25 | Yes |
| (Gamma proteobacteria monophyletic) | 16,787.3 | 316 ± 36 | Yes | 15,165.1 | 180 ± 33 | Yes |
| (*Trichomonas* + *Clostridium pasteurianum*) | 16,569.5 | 98 ± 19 | Yes | 15,066.6 | 81 ± 26 | Yes |
| (Eukaryotes + *C. pasteurianum*) | 16,605.6 | 134 ± 27 | Yes | 15,116.7 | 131 ± 27 | Yes |
| (*Entamoeba* + *K. pneumoniae*/*E. agglomerans*) | 16,520.8 | 49 ± 18 | Yes | 15,092.6 | 107 ± 27 | Yes |
| (*Entamoeba* + *Trichomonas* + *K. pneumoniae*/*E. agglomerans*) | 16,539.5 | 68 ± 21 | Yes | 15,097.0 | 111 ± 26 | Yes |

Note.—Maximum parsimony (MP) was used to generate shortest trees consistent with each constraint for both the DNA and the protein data sets. Likelihood scores were calculated with gamma rate adjustment (eight variable categories and one invariable category; see *Materials and Methods*). Trees were evaluated relative to their respective (DNA or protein) maximum-likelihood (ML) and MP trees under the Kishino-Hasegawa test (see text). The negative log-likelihood values of each tree are shown (−lnL) alongside the difference from the likelihood of the ML tree and the associated standard errors. The 5% confidence level is used as a cutoff point between competing hypotheses of relationships. Eukaryote taxa are those defined in figure 2.

in the tree—was excluded from analysis. Evaluation of the log-likelihoods of the 179 trees found by the ProtML quick-add OTU search option using a rate variable model in PUZZLE identified 62 trees in which eukaryote sequences were not monophyletic and which could not be distinguished from the ML tree by the criterion of 1.96 SE of log-likelihood differences ($\Delta$lnL = 1.96, $P$ = 0.95). All of these trees were worse than the ML tree by at least the criterion of 1 SE ($P$ = 0.62) of log-likelihood differences, with the best being 1.1 SE ($P$ = 0.73), where eukaryotes were monophyletic with the exception of the *Giardia* PFO1 sequence. In fact, in about half of the 62 trees, eukaryotes were monophyletic with the exception of the *Giardia* PFO1.

The eukaryote sequences were recovered in cluster III in LogDet distance analyses of DNA sequences with bootstrap support of 74% (fig. 2), increasing to 78% when the *Giardia* PFO1 sequence was excluded from analysis. Support rose to 85% when the δ-proteobacterium *Desulfovibrio africanus,* which sometimes (10% of bootstrap partitions) branched with different eukaryotes rather than with other eubacteria, was excluded. The best tree we found in which *D. africanus* PFO grouped with eukaryotes rather than at the base of eubacterial cluster II was 1.2 SE ($P$ = 0.77) of log-likelihood differences worse than the DNA ML tree, or 1.5 SE ($P$ = 0.87) of log-likelihood differences worse than the protein ML tree (table 1).

LogDet analyses of codon positions 1 and 2 (587 variable sites) and unconstrained protein ML analyses of a reduced data set incorporating the partial *G. lamblia* PFO2 sequence (323 variable sites) always recovered a sister relationship between *Spironucleus* and *G. lamblia* PFO2 (100% bootstrap support in all analyses; data not shown). This is consistent with relationships inferred from elongation factor 1-alpha (Keeling and Doolittle 1997) and GAPDH (Rozario et al. 1996), whereby these two diplomonads form a monophyletic group.

## Testing Published Hypotheses for the Origins of Eukaryotic PFO using ML Difference Tests

It has been suggested (Embley, Horner, and Hirt 1997; Martin and Müller 1998) that eukaryote PFO may have originated with the protomitochondrion, which is thought to share common ancestry with contemporary α-proteobacteria (Andersson et al. 1998; Gray et al. 1998). Examination of the 179 trees found by the quick-add OTU search option in PROTML revealed that the eukaryotic sequences never branched in a position which provided direct support for this hypothesis. To further investigate if the available PFO data are consistent with an α-proteobacterial origin, we used two constraints, one whereby we constrained the sequence from the (single) α-proteobacterium *Rhodospirillum* (cluster I) with eukaryotes, and one whereby we constrained the γ-proteobacteria from cluster II with eukaryotes. The latter constraint is based on multiple molecular data sets which recover α- and γ-proteobacteria as part of a monophyletic group with β-proteobacteria (Olsen, Woese, and Overbeek 1994; Viale et al. 1994; Eisen 1995) and was implemented because there are no α-proteobacterial sequences in cluster II. Both of these trees could be rejected at the 5% level (table 1). A sister group relationship between the available *Clostridium* sequences and the *Trichomonas* sequences (Whatley, John, and Whatley 1979) could also be rejected at the 5% confidence level. However, it was not possible to exclude a sister group relationship between all of the eukaryotic sequences and the sequence from the δ-proteobacterium *D. africanus* at the 5% level (table 1).

Rosenthal et al. (1997) hypothesized that the PFO genes of *Entamoeba,* and possibly those of *Trichomo-*

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

*nas,* may share a common ancestry with enterobacteria to the exclusion of other eukaryote PFOs. However, constrained trees enforcing relationships between PFOs from *Entamoeba* and/or *Trichomonas* and PFOs from *K. pneumoniae* and/or *E. agglomerans* could all be rejected at the 5% confidence level (table 1).

### Gene Duplications and Horizontal Transfers May Have Contributed to Generation of Contemporary Diversity of Eubacterial Homodimeric PFO Sequences

A notable feature of the eubacterial PFO tree (fig. 2) is that it splits proteobacterial PFOs into two clusters, with a strongly supported central branch. Cluster I contains the α-proteobacterium *Rhodospirillum rubrum* and the γ-proteobacteria *Escherichia coli* and *Yersinia pestis,* whereas cluster II contains the δ-proteobacterium *D. africanus* and the γ-proteobacteria *K. pneumoniae* and *E. agglomerans.* While published trees differ in the amount of support they provide for a monophyletic group containing α-, γ-, and δ-proteobacteria, there is strong support from multiple data sets that γ-proteobacteria are a monophyletic group (Ludwig et al. 1993; Olsen, Woese, and Overbeek 1994; Van de Peer et al. 1994; Viale et al. 1994; Eisen 1995). However, the signal for nonmonophyly of γ-proteobacterial PFO sequences is strong, because trees in which γ-proteobacteria were constrained as a monophyletic group could all be rejected at the 5% confidence level (table 1).

The nonmonophyly of γ-proteobacterial PFO sequences can be reconciled with these other gene trees only if processes are invoked which could have produced multiple copies of PFO genes in ancestral genomes, such as gene duplication(s) or horizontal transfer(s). Thus, one way of reconciling the PFO tree with the consensus that γ-proteobacteria are a monophyletic group is to hypothesize that a single horizontal transfer of a PFO gene occurred from a Gram-positive bacterium to the common ancestor of *Klebsiella* and *Enterobacter.* Thus, *Klebsiella* and *Enterobacter* branch with clostridia in figure 2 because the ''Gram-positive'' copy, rather than the ortholog of the *E. coli* and *Yersinia* PFOs, has been sampled. Support for this hypothesis could come from demonstrating that *Klebsiella* and *Enterobacter* PFOs branched robustly with a specific sister group inside of the Gram-positive clade. Additionally, if ''Gram-positive'' PFOs among γ-proteobacteria were subsequently shown to be restricted to a monophyletic group centering on *Klebsiella* and *Enterobacter,* then this would also argue for a single local transfer event subsequent to the diversification of other γ-proteobacteria. Additional sampling of PFOs from Gram-positive eubacteria and proteobacteria are needed to further explore the ''horizontal transfer'' hypothesis. Interestingly, it is already apparent that the complete genomes of *Synechocystis, E. coli,* and *Treponema pallidum* contain only a single PFO coding sequence of the type found in cluster I; under the ''horizontal transfer'' hypothesis, they never had a cluster II type PFO.

An alternative explanation of the incongruent γ-proteobacterial relationships depicted in the PFO tree is that clusters I and II represent subtrees, each of which is tracking organismal relationships. This situation could arise if there were two copies of PFO in the common ancestor of the sampled eubacteria. It is apparent that different copies of PFO exist in the genomes of *Giardia, Trichomonas, Clostridium acetobutylicum,* and *C. pasteurianum,* so inferring that multiple copies might also have existed in some ancestral eubacterial genomes is not unreasonable. Clusters I and II could then represent incomplete sampling of PFO diversity from individual genomes (undiscovered genes) or loss of individual copies of PFO from some genomes (sorting). Support for this ''ancestral gene diversity hypothesis'' could come from finding both cluster I and cluster II copies of PFO in the genomes of phylogenetically distinct eubacteria. Under this hypothesis, the absence of a cluster II PFO copy in the genomes of *Synechocystis, E. coli,* and *T. pallidum* is the result of gene loss.

The position of the *Desulfovibrio* (a δ-proteobacterium) PFO in figure 2 is also worthy of comment, because 16S rRNA (Olsen, Woese, and Overbeek 1994; Van de Peer et al. 1994), elongation factor EF-Tu (Ludwig et al. 1993), GroEL (Viale et al. 1994), and RecA (Eisen 1995) all favor a relationship between δ-proteobacteria and the other proteobacteria. However, a sister group relationships between *Desulfovibrio* and either the cluster II γ-proteobacteria or the cluster I α- and γ-proteobacteria are both rejected by the PFO data set at the 5% confidence level of log-likelihood differences. The apparently anomalous position of the *Desulfovibrio* PFO sequence is not considered under the two hypotheses discussed above to reconcile γ-proteobacteria relationships. However, if both hypotheses are combined, i.e., to invoke skewed sampling of ancestral PFO gene diversity and a single horizontal transfer, then the PFO tree can be reconciled to traditional notions of proteobacterial relationships. Whatever hypothesis is favored, an additional event is required to explain the occurrence of at least two copies of PFO in *C. acetobutylicum* (A and B in fig. 2), one of which (A) is more closely related to a sequence from *C. pasteurianum* than it is to *C. acetobutylicum* B. A partial (369 amino acids) *C. pasteurianum* PFO sequence recovered from GenBank (AF064550) forms the sister group to the *C. acetobutylicum* B gene (not shown). This topology suggests retention of two copies of PFO which were both present in the genome of the common ancestor of these two *Clostridium* species.

Other relationships within each subtree (cluster I and cluster II) are not generally significantly incongruent, as judged by low bootstrap values or by the Kishino-Hasegawa test of log-likelihoods of competing trees, with relationships supported by other genes. The apparently incongruent position of *Rhodospirillum* as the weakly supported sister group of cyanobacteria (fig. 2), rather than as the sister group of γ-proteobacteria as commonly reported, is not significant at the 5% confidence level. Moreover, in the protein ML tree, *Rhodospirillum* does group with the γ-proteobacteria in cluster I.

### PFO Domains in Other Eukaryotes

The *MET10* gene of *Saccharomyces cerevisiae* (fig. 3) encodes a protein (Met10p) which contains a large

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

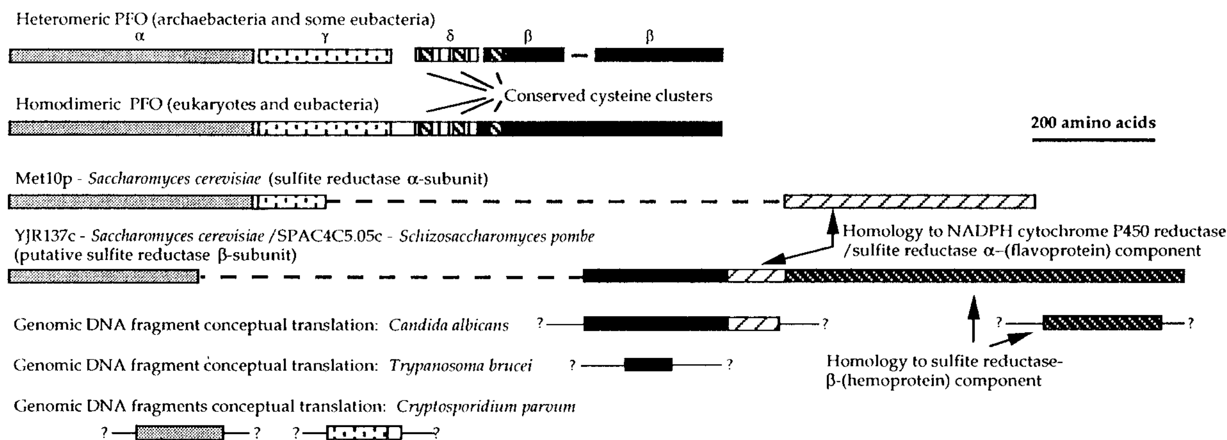


FIG. 3.—Schematic representation of sequences containing PFO fragments detected in eukaryote genomes. Shading patterns indicate inferred sequence homology. Enclosed white regions are "linker regions" present in homodimeric PFOs but not in the heteromeric PFOs of archaebacteria and some eubacteria. The fragments flanked by thin continuous lines were recovered through similarity searches of random genome survey projects (see below). Question marks indicate that the sequences flanking these regions are presently unknown. Dashed lines indicate artificial gaps which have been inserted to "align" regions of inferred homology. PFO fusion proteins YJR137c (*Saccharomyces cerevisiae,* Z49637) and SPAC4C5.05c (*Schizosaccharomyces pombe,* Z98560) were recovered through BLAST searches of the nonredundant nucleotide databases. Fragments of *Candida albicans* genes similar to YJR137c were identified using BLASTn at the *Candida* genome project (alces.med.umn.edu). *Trypanosoma brucei* (B13566) and *Cryptosporidium parvum* (AQ023784 and AQ023783) fragments with inferred homology to PFO were identified using BLAST searches at the EBI (www.ebi.ac.uk/parasites/parasite_blast_server.html).

fragment of PFO fused to an NADPH cytochrome P450 reductase/sulfite reductase α-(flavoprotein) component (Hansen, Cherest, and Kielland-Brandt 1994). We identified further ORFs (fig. 3) encoding PFO fusion proteins in the genomes of *S. cerevisiae* (YJR137c) and *Schizosaccharomyces pombe* (SPAC4C5.05c). YJR137c and SPAC4C5.05c are 44.5% identical in amino acid sequence and contain the same fragments of PFO fused to the same domains from other redox proteins and thus are likely orthologous. Met10p and YJR137c/SPAC4C5.05c lack the central region of PFO which contains conserved cysteine residues involved in Fe-S coordination. YJR137c/SPAC4C5.05c contain a 200-amino-acid fragment of the β domain of PFO which is absent from Met10p (fig. 3). The putative TPP binding motif of PFO (Kletzin and Adams 1996) may not be functional in YJR137c/SPAC4C5.05 because the functionally important "GDG" signature is absent.

YJR137c/SPAC4C5.05c also contain a region of strong similarity to the flavodoxin-like domain of the NADPH-cytochrome *c* (P450) reductase/NADPH-ferrihemoprotein reductase gene product of animals and plants, but this is much shorter than, and does not overlap with, the fragment present in Met10p (fig. 3). Instead, the extreme C-terminal regions of YJR137c/SPAC4C5.05c contain an almost complete β-(hemoprotein) component of NADPH-sulfite reductase. A search of the yeast expressed sequence tag database revealed that YJR137c is transcribed. Partial sequence data obtained from the *Candida* genome project also suggest the presence of a gene similar to YJR137c in this organism (fig. 3).

The PFO regions of YJR137c and SPAC4C5.05c were subjected to phylogenetic analyses (not shown) that weakly supported a specific relationship to other eukaryote sequences. However, other relationships recovered with strong support in analyses of the full-length data set were also poorly supported by the reduced data set. Inspection of bootstrap partitions revealed that the fungal sequences had no strong tendency to cluster with particular eubacterial genes—rather, decay in bootstrap support for eukaryote monophyly was caused equally by instability of all of the eukaryotic sequences.

The genomes of the protist parasites *Trypanosoma brucei* and *Cryptosporidium parvum* also contain fragments of PFO (fig. 3). The hits were short (200–450 nt) sequence fragments from random genomic survey projects. Thus, we cannot tell if these fragments derive from complete PFO coding sequences or from fusion proteins like those found in fungi. However, the conceptual translation product of one of the *C. parvum* sequences corresponds to a region of PFO which is not present in the fungal fusion proteins (fig. 3). Phylogenetic analysis of the *C. parvum* sequence suggested a weakly supported but specific relationship with the other eukaryotic sequences.

## Discussion

The nuclear genome of eukaryotes is a mosaic, containing some genes which appear closer to archaebacteria and others which appear closer to eubacteria. Genes involved in replication, transcription, and translation are almost exclusively of the "archaebacterial type," while the "eubacterial type" are generally those involved in, for example, energy metabolism, amino acid biosynthesis, and intermediary metabolism (Brown and Doolittle 1997; Ribeiro and Golding 1998; Rivera et al. 1998). Analyses of eukaryotic PFOs are also consistent with a eubacterial, rather than an archaebacterial, origin. Characterized eukaryotic PFOs are all homodimeric, like many eubacterial enzymes (Hrdy and Müller 1995; Rosenthal et al. 1997). In contrast, all archaebac-

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

teria so far investigated, and some eubacteria, contain a multimeric form of PFO which, while considered homologous to homodimeric PFO, is only distantly related in phylogenetic trees (Kletzin and Adams 1996; Müller 1998).

It has recently been suggested (Rosenthal et al. 1997) that the PFO genes of *Entamoeba* and *Giardia* probably originated through separate horizontal transfers from different eubacteria. We find no support for this hypothesis from our analyses, which consistently favor a single common origin for eukaryotic PFO. The *Giardia* PFO1 sequence contributes most to the instability of the eukaryotic cluster, but it is also the longest branch in phylogenetic trees (fig. 2). In addition, plots of transitions versus transversions suggest potential mutational saturation for pairwise comparisons involving PFO1 (fig. 1*A, B,* and *D*), whose amino acid composition deviates significantly (5% $\chi^2$ test) from those of the other PFOs. It also contains amino acid substitutions at positions which are otherwise conserved among PFO sequences. Together, these observations suggest that the phylogenetic position of the *Giardia* PFO1 sequence must be interpreted with care. Moreover, we show that *Giardia* contains at least two distinct PFO sequences, one of which (PFO2) clusters strongly with the *Spironucleus* PFO. Such a relationship between *Spironucleus* and *Giardia* is supported by ultrastructure and other molecular data sets (Rozario et al. 1996; Keeling and Doolittle 1997). The amino acid composition of the new *Giardia* PFO2 sequence is similar to those of other PFOs and shows fewer anomalies than PFO1 in terms of amino acid composition at conserved positions. In addition, plots of transitions versus transversions for pairwise comparisons involving PFO2 codon position 2 fall within the distribution of points for all comparisons, except for those involving PFO1 (fig. 1*D*). We therefore infer that the new *Giardia* PFO2 sequence is orthologous to the *Spironucleus* PFO, rather than PFO1, which may be a paralog. Evidence suggesting gene duplication of eukaryote PFO is already apparent in the genome of *Trichomonas,* which has two very similar PFOs, and we have preliminary data which suggest that *Spironucleus* has another copy similar to the PFO published here. Interestingly, a second 2-oxoacid oxidoreductase activity, which preferentially utilizes α-ketobutyrate as a substrate, has been detected in *Giardia* (Townson, Upcroft, and Upcroft 1996). It is possible that the two PFO-like genes demonstrated in this study correspond to the two activities which have been described. This scenario would resemble the situation in some bacteria and archaea for which related genes encoding different 2-oxoacid oxidoreductase activities have been described (Kletzin and Adams 1996).

The tree for eubacterial homodimeric PFOs (fig. 2) is incongruent with species relationships based on multiple molecular data sets (Ludwig et al. 1993; Olsen, Woese, and Overbeek 1994; Van de Peer et al. 1994; Viale et al. 1994; Eisen 1995). To reconcile the PFO tree with these other gene trees, we hypothesize that the eubacterial PFO tree contains genes which are the products of gene duplication or horizontal transfer. Both are recognized phenomena affecting genome content, but distinguishing between them as the sources of particular genes can be difficult, especially when (as here) sampling of genes and species is sparse (Smith, Feng, and Doolittle 1992; Doolittle 1997; Martin and Schnarrenberger 1997; Nowitzki et al. 1998; Page and Charleston 1998).

Our attempts to root the tree of homodimeric PFOs using the multimeric PFOs (Kletzin and Adams 1996) yielded trees with very long central branches, for which all rootings relevant to eukaryote PFO origins were permitted at the 5% confidence level of log-likelihood differences (data not shown). The inability to locate a root, plus the complex tree for homodimeric PFO, confounds inferences about the identity of a potential eubacterial donor of eukaryote PFO. However, we found no support for a sister group relationship between the eukaryotic PFO sequences and *Rhodospirillum* (α-proteobacterium, cluster I) or cluster II γ-proteobacterial PFO sequences. Thus, based on current gene and species sampling, we find no direct support for the hypothesis that eukaryote PFO originated with the protomitochondrion (Embley, Horner, and Hirt 1997; Martin and Müller 1998).

We could not exclude a relationship between the PFO from the δ-proteobacterium *Desulfovibrio* and eukaryotic PFO, although it should be emphasized that this relationship received no strong support from any of our analyses. A relationship between *Desulfovibrio* PFO and eukaryote PFO could be interpreted to provide support for the hypothesis that eukaryotes obtained their PFO through horizontal transfer from a δ-proteobacterium. Moreover, the recently published ''syntrophic hypothesis'' (Moreira and López-García 1998) indeed suggests that a symbiosis between an archaeal methanogen and a hydrogen-producing δ-proteobacterium gave rise to an ancestral amitochondriate eukaryote. Thus, although PFO is not actually mentioned in this paper, the ''syntrophic hypothesis'' could be interpreted to predict a δ-proteobacterial origin for eukaryote PFO. However, it is possible from the PFO tree that the *Desulfovibrio* PFO is a different copy of PFO from that sampled from the α-proteobacteria in figure 2. If this is the case, then orthologs of the *Desulfovibrio* PFO may exist or may have existed in the genomes of α-proteobacteria including that of the protomitochondrion. The absence of PFO homologs in the genome of the α-proteobacterium *Rickettsia prowazekii* (Andersson et al. 1998) or in the genomes of contemporary mitochondria (Gray et al. 1998) clearly does not exclude the common ancestor of mitochondria and *Rickettsia* from possessing PFO genes in its genome (Embley and Martin 1998). Both the *Rickettsia* and the mitochondria genomes have undergone reduction in content during their separate evolutionary histories as intracellular residents (Andersson and Kurland 1998; Gray et al. 1998). Moreover, while multiple data sets indicate an α-proteobacterial (mitochondrial) contribution to the eukaryote genome (Andersson et al. 1998; Gray et al. 1998), we are unaware of published molecular phylogenies which robustly support a δ-proteobacterial component.

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

The finding that PFO sequences from diverse anaerobic protists are monophyletic prompted us to search for retention of PFO sequences in the genomes of other eukaryotes. Our finding of PFO fragments in all of the genomes we investigated supports a long and complex history for PFO in the evolution of eukaryotes generally. We infer that these PFO fragments are monophyletic with PFOs from anaerobic protists, because we find no consistent support for alternative origins from the limited amount of data currently available for analysis. The *C. parvum* sequence contains a region of PFO which is not present in fungal fusion proteins (see below) and includes part of a linker region common to all eukaryote homodimeric PFOs. *Cryptosporidium parvum* is reported to lack mitochondria (Tetley et al. 1998), and *Cryptosporidium* infections are typically initiated in the host gut. Thus, it seems possible that there could be a role for PFO during at least some life cycle stages. *Trypanosoma* is related to *Euglena gracilis,* which has been reported to contain a PFO fusion protein in its mitochondria (Inui et al. 1991; Hrdy and Müller 1995; Müller 1998).

All of the fungal genomes we investigated contain fusion proteins that contain fragments of PFO sequence joined to fragments of other redox proteins (fig. 3). At least one of these fusion proteins (YJR137c/SPAC4C5.05c) is common to *Saccharomyces, Schizosaccharomyces,* and, tentatively, *Candida,* suggesting that it was present in the most recent common ancestor of these fungi. Several lines of evidence allow us to speculate on the role of YJR137c/SPAC4C5.05c. Each contains an almost intact β-(hemoprotein) component of a sulfite reductase gene, the product of which normally interacts with an α-(flavoprotein) sulfite reductase component, such as the one encoded by *MET10,* which is involved in methionine biosynthesis in *Saccharomyces* (Hansen, Cherest, and Kielland-Brandt 1994). The functional partner of Met10p is a peptide of ~165 kDa, which is close to the predicted molecular weight of 161 kDa for YJR137c (Kobayashi and Yoshimoto 1982). Biochemical analysis of the *Saccharomyces* sulfite reductase holoenzyme included an analysis of amino acid composition (Kobayashi and Yoshimoto 1982). Most interestingly, when the amino acid composition of the *MET10* fusion protein (including PFO fragments) is subtracted, the remaining composition data match closely with the prediction for YJR137c (fig. 4). In addition, YJR137c and SPAC4C5.05c contain flavodoxin-like domains, as was predicted for the functional partner of Met10p (Hansen, Cherest, and Kielland-Brandt 1994). Lastly, upstream of YJR137c, there are the consensus regulatory elements $UAS_{met}$ and $URS_{met}$, which are associated with methionine biosynthesis genes in yeast (Thomas, Cherest, and Surdin-Kerjan 1989). These data suggest that YJR137c and Met10p together constitute the sulfite reductase in *Saccharomyces* and that this enzyme contains a substantial contribution from PFO. At present, there are too few data concerning the functions of different parts of the PFO protein to say what the possible role(s) of the PFO fragments might be. For example, we cannot say whether the PFO fragments detected on the fungal fusion proteins are sufficient to utilize pyruvate as an electron donor for reduction of sulfite (Kobayashi and Yoshimoto 1982; Hansen, Cherest, and Kielland-Brandt 1994).

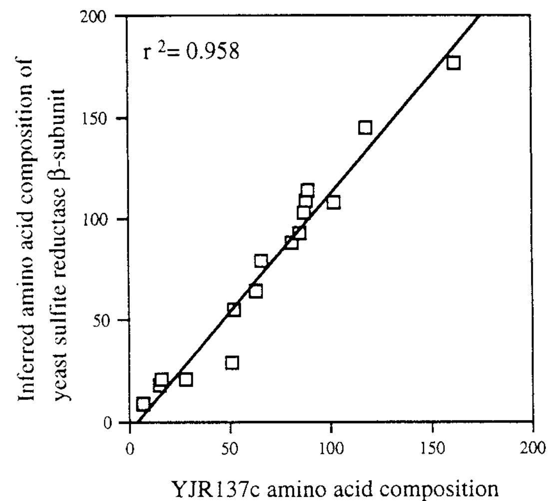


FIG. 4.—Evidence that *Saccharomyces cerevisiae* YJR137c is the β subunit of sulfite reductase. The empirically determined amino acid composition of *S. cerevisiae* sulfite reductase β subunit (after Kobayashi and Yoshimoto 1982; Hansen, Cherest, and Kielland-Brandt 1994) was plotted against the inferred amino acid composition of *S. cerevisiae* YJR137c, which contains PFO-derived domains. In addition to the experimental error inherent in chemical determination of amino acid content, the slight numerical discrepancies between the inferred amino acid content of YJR137c and the empirically determined amino acid content of the *S. cerevisiae* sulfite reductase β subunit may be attributable to errors in the estimate of holoenzyme molecular weight made by Koboyashi and Yoshimoto (1982). The coefficient of determination is shown. Interestingly, the YJR137c protein is also identified as the closest match to the experimentally determined amino acid frequencies and molecular weight by the AACompIdent identification tool (available at the Swiss Institute of Bioinformatics, ExPASy WWW server; Appel, Bairoch, and Hochstrasser 1994).

## Conclusions

Because PFO is essential for energy metabolism (ATP synthesis) in all amitochondriate eukaryotes studied to date, the evolutionary history of the enzyme is critical to understanding the origin of the anaerobic energy metabolism common to these groups (Müller 1998). Here, we have shown that full-length PFOs from *Trichomonas, Giardia, Spironucleus,* and *Entamoeba* are monophyletic, suggesting a single common origin for PFO in these organisms. In contrast, in the small-subunit ribosomal RNA tree, *Trichomonas* and *Giardia*/*Spironucleus* are separated from *Entamoeba* by aerobic mitochondrion-containing lineages (Sogin et al. 1996). Recent analyses of other genes have indicated that *Trichomonas* and *Giardia* may form a monophyletic group (reviewed in Embley and Hirt 1998), which would be consistent with a single origin for their PFO genes. However, no previous reports have suggested that *Entamoeba* might also be monophyletic with *Trichomonas* and *Giardia.* We also found PFO fragments in the genomes of *S. cerevisiae, S. pombe, Candida albicans, T. brucei,* and *C. parvum,* which we infer to be

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

monophyletic with comparable PFO sequences from anaerobic protists. Thus, despite currently conflicting views of protist evolution, the topology of the PFO tree is consistent with the view that PFO was present in the common ancestor of contemporary eukaryotes and was retained, wholly or in part, during the evolution of oxygen-dependent and mitochondrion-bearing lineages (Martin and Müller 1998).

A central question for understanding the evolution of anaerobic metabolism in contemporary protists is that of identification of the source(s) of key genes such as the one for PFO. Most modern theories on the origins of the eukaryotic cell postulate, at some stage of the cell's evolution, significant genetic contributions from an archaebacterium and a eubacterium (reviewed in Doolittle 1998; Martin and Müller 1998). Genomics has revealed data compatible with this general picture, with informational genes of eukaryotes resembling those of archaebacteria, while most operational (e.g., energy metabolism) genes are eubacteria-like (Brown and Doolittle 1997; Ribeiro and Golding 1998; Rivera et al. 1998). Eukaryote PFOs follow this pattern in that they more closely resemble eubacterial homodimeric PFO, being colinear with these enzymes along their entire length, than the multimeric PFOs so far described for archaebacteria (Kletzin and Adams 1996). Some of the operational genes in eukaryotic genomes can already be traced to likely gene transfers from the mitochondrial or chloroplast endosymbionts, and many others may also have originated with these symbionts (Ribeiro and Golding 1998; Rivera et al. 1998). In this context, it is worth emphasizing that all eukaryotes known to contain PFO also contain genes of mitochondrial origin. Our analyses favor a common origin for eukaryotic cytosolic and hydrogenosomal PFOs from a single eubacterial source, rather than separate horizontal transfers as previously suggested (Rosenthal et al. 1997). However, given the current sampling of bacterial PFO genes and the absence of any strong support for a eubacterial sister group to eukaryote PFOs, we conclude that the current data are insufficient to resolve a donor lineage for eukaryote PFO. In particular, we find no direct support for the hypothesis that the donor of eukaryote PFO was the common α-proteobacterial ancestor of mitochondria and hydrogenosomes.

## Acknowledgments

We would like to thank Andrew Roger, Bengt Sennblad, Mark van der Giezen, Miklós Müller, and Mark Wilkinson for useful discussions and comments on the manuscript. We are indebted to Mitchell Abrahamsen and Larry Ruben for permission to use their unpublished sequence data on *Cryptosporidum* and *Trypanosoma.* Preliminary sequence data for a partial *Giardia* PFO, which we used to design PCR primers, were obtained with permission from the *Giardia lamblia* Genome Project at The Josephine Bay Paul Center Web site at the Marine Biological Laboratory (www.bpc.mbl.edu). Sequencing at this center is supported by the National Institute of Allergy and Infectious Diseases. We also utilized data generated by the Sanger Centre (www.sanger.ac.uk), the Institute for Genomic Research (www.tigr.org), and Genome Therapeutic Corporation (www.cric.com/genomesequences). D.S.H. is supported by the Natural History Museum, R.P.H. is the recipient of Wellcome Trust Fellowship number 045702/Z/95/Z. The sequences reported have been deposited in GenBank (accession numbers AF132673, AF132546, AF132547, and AF146519).

## LITERATURE CITED

Adachi, J., and M. Hasegawa. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Comput. Sci. Monogr. **28**:1–150.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

Andersson, S. G. E., and C. G. Kurland. 1998. Reductive evolution of resident genomes. Trends Microbiol. **6**:263–268.

Andersson, S. G. E., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Pontén, U. C. M. Alsmark, R. M. Podowski, A. K. Näslund, A.-S. Eriksson, H. H. Winkler, and C. G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature **396**:133–140.

Appel, R. D., A. Bairoch, and D. F. Hochstrasser. 1994. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem. Sci. **19**:258–260.

Brown, J. R., and W. F. Doolittle. 1997. Archaea and the prokaryote to eukaryote transition. Microbiol. Mol. Biol. Rev. **61**:456–502.

Bui, E. T. N., P. J. Bradley, and P. J. Johnson. 1996. A common evolutionary origin for mitochondria and hydrogenosomes. Proc. Natl. Acad. Sci. USA **93**:9651–9656.

Charleston, M. 1998. Spectrum—a tool for spectral analysis of phylogenetic data. Bioinformatics **14**:98–99.

Clark, C. G., and A. J. Roger. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica.* Proc. Natl. Acad. Sci. USA **92**:6518–6521.

Doolittle, R. F. 1997. Microbial genomes opened up. Nature **392**:339–343.

Doolittle, W. F. 1998. A paradigm gets shifty. Nature **392**: 15–16.

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.

Embley, T. M., and R. P. Hirt. 1998. Early branching eukaryotes? Curr. Opin. Genet. Dev. **8**:624–629.

Embley, T. M., D. S. Horner, and R. P. Hirt. 1997. Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria? Trends Ecol. Evol. **12**:437–441.

Embley, T. M., and W. Martin. 1998. A hydrogen-producing mitochondrion. Nature **396**:517–519.

Fenchel, T., and B. J. Finlay. 1995. Ecology and evolution in anoxic worlds. Oxford University Press, Oxford, England.

Germot, A., H. Philippe, and H. L. Guyader. 1996. Presence of a mitochondrial type 70kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. Proc. Natl. Acad. Sci. USA **93**: 14614–14617.

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

GRAY, M. W., B. LANG, R. CEDERGREN et al. (15 co-authors). 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. **26**:865–878.

HANSEN, J., H. CHEREST, and C. KIELLAND-BRANDT. 1994. Two divergent MET10 genes, one from *Saccharomyces cerevisiae* and one from *Saccharomyces carlsbergiensis,* encode the a subunit of sulfite reductase and specify potential binding sites for FAD and NADPH. J. Bacteriol. **176**:6050–6058.

HASHIMOTO, T., L. B. SANCHEZ, T. SHIRAKURA, M. MÜLLER, and M. HASEGAWA. 1998. Secondary absence of mitochondria in *Giardia* and *Trichomonas* revealed by valyl-tRNA synthetase phylogeny. Proc. Natl. Acad. Sci. USA **95**:6860–6865.

HIRT, R. P., J. M. LOGSDON, B. HEALY, M. W. DOREY, W. F. DOOLITTLE, and T. M. EMBLEY. 1999. Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc. Natl. Acad. Sci. USA **96**:580–585.

HORNER, D. S., R. P. HIRT, S. KILVINGTON, D. LLOYD, and T. M. EMBLEY. 1996. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. Proc. R. Soc. Lond. B Biol. Sci. **263**:1053–1059.

HRDY, I., and M. MÜLLER. 1995. Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis.* J. Mol. Evol. **41**:388–396.

INUI, H., R. YAMAJI, H. SAIDOH, K. MIYATAKE, Y. NAKANO, and S. KITAOKA. 1991. Pyruvate : $NADP^+$ oxidoreductase from *Euglena gracilis*: limited proteolysis of the enzyme with trypsin. Arch. Biochem. Biophys. **286**:270–276.

KEELING, P. J., and W. F. DOOLITTLE. 1997. Widespread and ancient distribution of a noncanonical genetic code in diplomonads. Mol. Biol. Evol. **14**:895–901.

KERSCHER, L., and D. OESTERHELT. 1982. Pyruvate : ferredoxin oxidoreductase—new findings on an ancient enzyme. Trends Biochem. Sci. **7**:371–374.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

KLETZIN, A., and M. W. W. ADAMS. 1996. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima.* J. Bacteriol. **178**:248–257.

KOBAYASHI, K., and A. YOSHIMOTO. 1982. Studies on yeast sulfite reductase IV. Structure and steady state kinetics. Biochim. Biophys. Acta **705**:348–356.

LAKE, J. A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl. Acad. Sci. USA **91**:1455–1459.

LENTO, G. M., R. E. HICKSON, G. K. CHAMBERS, and D. PENNY. 1995. Use of spectral analysis to test hypotheses on the origins of pinnipeds. Mol. Biol. Evol. **12**:28–52.

LOCKHART, P. J., A. W. D. LARKUM, M. A. STEEL, P. J. WADDEL, and D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. USA **93**:1930–1934.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

LUDWIG, W., J. NEUMAIER, N. KLUGBAUER et al. (12 co-authors). 1993. Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes. Antonie Van Lueewenhoek **64**:285–305.

MAIDAK, B. L., G. J. OLSEN, N. LARSEN, R. OVERBEEK, M. J. MCCAUGHEY, and C. R. WOESE. 1996. The Ribosomal Database Project (RDP). Nucleic Acids Res. **24**:82–85.

MARTIN, W., and M. MÜLLER. 1998. The hydrogen hypothesis for the first eukaryote. Nature **392**:37–41.

MARTIN, W., and C. SCHNARRENBERGER. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. **32**:1–18.

MOREIRA, D., and P. LÓPEZ-GARCÍA. 1998. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. **47**:517–530.

MÜLLER, M. 1998. Enzymes and compartmentation of core energy metabolism of anaerobic protists—a special case in eukaryotic evolution. Pp. 109–127 *in* G. H. COOMBS, K. VICKERMAN, M. A. SLEIGH, and A. WARREN, eds. Evolutionary relationships among protozoa. Kluwer Academic Publishers, London.

NOWITZKI, U., A. FLECHNER, J. KELLERMAN, M. HASEGAWA, C. SCHNARRENBERGER, and W. MARTIN. 1998. Eubacterial origin of nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase from spinach: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. Gene **214**:205–213.

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.

PAGE, R. D. M., and M. A. CHARLESTON. 1998. Trees within trees: phylogeny and historical associations. Trends Ecol. Evol. **13**:356–359.

RIBEIRO, S., and G. B. GOLDING. 1998. The mosaic nature of the eukaryotic nucleus. Mol. Biol. Evol. **15**:779–788.

RIVERA, M. C., R. JAIN, J. E. MOORE, and J. A. LAKE. 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA **95**:6239–6244.

ROGER, A. J., C. G. CLARK, and W. F. DOOLITTLE. 1996. A possible mitochondrial gene in the early branching amitochondriate protist *Trichomonas vaginalis.* Proc. Natl. Acad. Sci USA **93**:14618–14622.

ROGER, A. J., S. G. SVARD, J. TOVAR, C. G. CLARK, M. W. SMITH, F. D. GILLIN, and M. L. SOGIN. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc. Natl. Acad. Sci. USA **95**:229–234.

ROSENTHAL, B., M. ZHIMING, D. CAPLIVSKI, S. GHOSH, H. DE LA VEGA, T. GRAF, and J. SAMUELSON. 1997. Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica.* J. Bacteriol. **179**:3736–3745.

ROZARIO, C., L. MORIN, A. J. ROGER, M. W. SMITH, and M. MÜLLER. 1996. Primary structure and phylogenetic relationships of glyceraldehyde-3-phosphate dehydrogenase gene of free-living and parasitic diplomonad flagellates. J. Eukaryot. Microbiol. **43**:330–340.

SHOEMAKER, J. S., and W. M. FITCH. 1989 Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. Mol. Biol. Evol. **6**:270–289.

SIDOW, A., T. NGUYEN, and T. P. SPEED. 1992. Estimating the number of invariable codons with a capture-recapture method. J. Mol. Evol. **35**:253–260.

SMITH, M. W., D.-F. FENG, and R. F. DOOLITTLE. 1992. Evolution by acquisition: the case for horizontal gene transfers. Trends Ecol. Evol. **17**:489–493.

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022

SOGIN, M. L., H. G. MORRISON, G. HINKLE, and J. D. SILBERMAN. 1996. Ancestral relationships of the major eukaryotic lineages. Microbiologia SEM **12**:17–28.

SWOFFORD, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, Mass.

TETLEY, L., S. M. BROWN, V. MCDONALD, and G. H. COOMBS. 1998. Ultrastructural analysis of the sporozoite of *Cryptosporidium parvum.* Microbiology **144**:3249–3255.

THOMAS, D., H. CHEREST, and Y. SURDIN-KERJAN. 1989. Elements involved in S-adenosylmethionine-mediated regulation of the *Saccharomyces cerevisiae* MET25 gene. Mol. Cell. Biol. **9**:3292–3298.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 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.

TOWNSON, S. M., J. A. UPCROFT, and P. UPCROFT. 1996. Characterisation and purification of pyruvate:ferredoxin oxidoreductase from *Giardia duodenalis.* Mol. Biochem. Parasitol. **79**:183–193.

VAN DE PEER, Y., J.-M. NEEFS, P. DE RIJK, P. DE VOS, and R. DE WACHTER. 1994. About the order of divergence of the major bacterial taxa during evolution. Syst. Appl. Bacteriol. **17**:32–38.

VIALE, A. M., A. K. ARAKAKI, F. C. SONCINI, and R. G. FERREYRA. 1994. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. Int. J. Syst. Bacteriol. **44**:527–533.

WADDELL, P. J., and M. A. STEEL. 1997. General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites. Mol. Phylogenet. Evol. **8**:398–414.

WHATLEY, J. M., P. JOHN, and F. R. WHATLEY. 1979. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. Proc. R. Soc. Lond. B Biol. Sci. **204**: 165–187.

YANG, D., Y. OYAIZU, H. OYAIZU, G. J. OLSEN, and C. R. WOESE. 1985. Mitochondrial origins. Proc. Natl. Acad. Sci. USA **82**:4443–4447.

WILLIAM MARTIN, reviewing editor

Accepted June 1, 1999

Downloaded from https://academic.oup.com/mbe/article/16/9/1280/2925528 by U.S. Department of Justice user on 16 August 2022