Evolutionary Analyses of the Small Subunit of Glutamate Synthase: Gene Order Conservation, Gene Fusions, and Prokaryote-to-Eukaryote Lateral Gene Transfers

Jan O. Andersson* and Andrew J. Roger

The Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Lateral gene transfer has been identified as an important mode of genome evolution within prokaryotes. Except for the special case of gene transfer from organelle genomes to the eukaryotic nucleus, only a few cases of lateral gene transfer involving eukaryotes have been described. Here we present phylogenetic and gene order analyses on the small subunit of glutamate synthase (encoded by *gltD*) and its homologues, including the large subunit of sulfide dehydrogenase (encoded by *sudA*). The scattered distribution of the *sudA* and *sudB* gene pair and the phylogenetic analysis strongly suggest that lateral gene transfer was involved in the propagation of the genes in the three domains of life. One of these transfers most likely occurred between a prokaryote and an ancestor of diplomonad protists. Furthermore, phylogenetic analyses indicate that the gene for the small subunit of glutamate synthase was transferred from a low-GC gram-positive bacterium to a common ancestor of animals, fungi, and plants. Interestingly, in both examples, the eukaryotes encode a single gene that corresponds to a conserved operon structure in prokaryotes. Our analyses, together with several recent publications, show that lateral gene transfers from prokaryotes to unicellular eukaryotes occur with appreciable frequency. In the case of the genes for sulfide dehydrogenase, the transfer affected only a limited group of eukaryotes—the diplomonads—while the transfer of the glutamate synthase gene probably happened earlier in evolution and affected a wider range of eukaryotes.

During the last 5 years it has become clear that lateral gene transfer (LGT) is a widespread and important evolutionary process among prokaryotes (10, 22). Although multicellular eukaryotes, like humans, seem to be relatively immune to LGT (2, 23, 24), it is unclear if LGT is also an important mode of evolution in unicellular eukaryotes (protists), due to the limited amount of genomic sequence data currently available for these organisms. Apart from the special case of prokaryote-to-eukaryote LGTs, which occur from the eukaryotic organelles with prokaryotic ancestry (the mitochondrion and the chloroplast) to the nucleus, there are only a handful of well-documented cases of prokaryote-to-eukaryote LGT events described in the literature (3, 8, 12, 16, 27). However, LGT appears to occur frequently between the two prokaryotic domains of life, *Bacteria* and *Archaea*, indicating that domain boundaries do not prevent successful LGT (20, 21).

We have chosen diplomonads, a group of anaerobic protists, to test the hypothesis that LGT from prokaryotes to protists is occurring at an appreciable rate. There is an ongoing whole-genome sequence project for *Giardia lamblia*, the most studied diplomonad, that releases newly generated sequences on a regular basis to GenBank (19). In our laboratory we have focused on another diplomonad, the Atlantic salmon parasite *Spironucleus barkhanus* (25), and we are, together with our collaborators, performing a sequence survey project. Genes that were laterally transferred from prokaryotes to the diplomonad lineage are expected to show higher similarity to prokaryotic genes than to any eukaryotic gene. Therefore, candidate laterally transferred genes are easily identified in these projects with similarity searches against the public sequence databases. Such genes have been further analyzed with phylogenetic methods that will give a much more reliable indication of whether the gene originated via LGT. Furthermore, if the gene is present in both the *Giardia* and *Spironucleus* projects, the phylogeny will show whether the gene has a common source in both lineages or it was a recent acquisition in one of the two lineages. Here we present such an analysis of the *sud* gene, a diplomonad gene with greater similarities to prokaryotic than to eukaryotic homologues. Sequences with the highest similarities to this gene corresponded to a conserved gene pair, *sudA* and *sudB*, present in *Thermotoga maritima*, *Treponema pallidum*, and *Pyrococcus furiosus*, suggesting an LGT event from a prokaryote to a eukaryotic ancestor of diplomonads. The gene pair encodes the two subunits of sulfide dehydrogenase (15). However, the C-terminal part of the diplomonad protein also shows high similarity to the small subunit of eubacterial glutamate synthase (*gltD*).

Glutamate synthase is a complex iron-sulfur flavoprotein that participates in ammonia assimilation processes. The eubacterial version of the enzyme is composed of two subunits. The small subunit is a flavin adenine dinucleotide-dependent NADPH oxidoreductase which shows sequence similarity to several other protein domains and enzyme subunits. Indeed, the small subunit of eubacterial glutamate synthase has been proposed to be a prototype of protein domains or enzyme subunits used in many different contexts to transfer electrons from NAD(P)H to an acceptor protein or protein domain (28).

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada. Phone: (902) 494-2881. Fax: (902) 494-1355. E-mail: joanders@is.dal.ca.

In the case of glutamate synthase, the small subunit provides electrons to the large subunit, which binds L-glutamine and 2-oxoglutarate and forms L-glutamate (28).

Sulfide dehydrogenase is an iron-sulfur flavoprotein without glutamate synthase activity (15). The enzyme consists of two subunits: the α-subunit, which is clearly homologous to the small subunit of glutamate synthase, and the β-subunit, which does not show any sequence similarity to the large subunit of glutamate synthase at all (15). Thus, the α-subunit of sulfide dehydrogenase is an example of an enzymatic subunit that is homologous (and very similar) to the small subunit of glutamate synthase but that is used in a different context. The two subunits of sulfide dehydrogenase are encoded by *sudA* and *sudB*, two genes found next to each other on the chromosome in *P. furiosus* (15). Sulfide dehydrogenase not only catalyzes the reduction of polysulfide to H_2S in *P. furiosus*, but it has also been shown to function as a reduced ferrodoxin-NADP oxidoreductase (15, 18). These dual functions of the protein complicate the interpretation of the physiological role for the enzyme. Indeed, the biological significance of the sulfide dehydrogenase activity in *P. furiosus* has recently been questioned, since the activity of the enzyme did not respond to changes in the availability of elemental sulfur (1).

To determine if the diplomonad gene shared a common ancestry with other sulfide dehydrogenases that would be suggestive of a shared function, and to elucidate its origin in the genomes of diplomonads, we retrieved all homologues of *gltD* and *gltD*-like genes from the public databases, studied the gene organization patterns around the homologues, and performed phylogenetic analyses. We identified two cases of prokaryote-to-eukaryote LGT, showing that interdomain LGT involving microbial eukaryotes may occur relatively frequently.

MATERIALS AND METHODS

Gene nomenclature. Unfortunately, the nomenclature of *gltD* and *gltD*-like genes is far from consistent in the public databases. To simplify the presentation, the gene name *gltD* is only used for genes that are linked to *gltB* (which encodes the large subunit of glutamate synthase) on the chromosome. *gltD*-like genes that are linked to *sudB* (see below) on the chromosome are annotated as *sudA*. All other genes are simply referred to as *gltD*-like.

PCR assays and nucleotide sequencing. The gene for sulfide dehydrogenase—a fusion of *sudA* and *sudB*—was amplified from *S. barkhanus* genomic DNA by degenerate PCR under standard conditions with the following amplification parameters: denaturation at 95°C for 30 s, annealing at 48°C for 1 min, and extension at 72°C for 3 min. Forty cycles were performed and a final extension step of 72°C for 10 min was used. The forward primer GLTDf1 (5′-GGTCGAGTHTGYCCHYARGA-3′) and the reverse primer GLTDr1 (5′-CAATCCATDGCDACRTTDCC-3′) were used. The PCR product was purified and sequenced directly. Uneven PCR (7) was used to obtain the 5′ end of the gene. Exact match primers from the obtained sequence were used together with arbitrary short primers in a reaction that used two radically different annealing temperatures (42 and 55°C) in alternate annealing steps, as described previously (7). This procedure yielded a PCR product that covered the 5′ end of the gene, which was cloned. Several independent clones were sequenced.

Phylogenetic analysis. All published *gltD*, *gltD*-like, and *sudA* sequences and all available unpublished *sudA* sequences were retrieved from the National Center for Biotechnology Information. The *Dictyostelium discoideum* Genome Project database (http://dicty.sdsc.edu) was also searched for homologues, with a negative result. Although the N-terminal part of eukaryotic dihydropyrimidine dehydrogenase has been shown to be homologous to the small subunit of glutamate synthase (28), we did not include it in our data set since the gene is only distantly related to *gltD*. This resulted in a data set with 399 unambiguously aligned positions. Sequences with >85% amino acid sequence identities were excluded from the data set to reduce the computational time. At the initial phylogenetic analysis, five sequences failed the χ^2 tests for deviation of amino

acid frequencies implemented in TREE-PUZZLE, version 4.02 (26) (a partial *Entamoeba histolytica gltD*-like sequence, a *Rhodobacter capsulatus gltD*-like sequence, a *Plasmodium falciparum gltS* sequence, and *gltD* sequences from *Pseudomonas aeruginosa* and *Campylobacter jejuni*). These sequences were excluded since the currently available phylogenetic methods cannot deal with strong amino acid compositional biases in the data (13). This procedure resulted in a 53-taxon data set with 399 amino acids (aa). Protein maximum likelihood (ML) phylogenies were inferred using PROML within the PHYLIP package, version 3.6 (11), with a Dayhoff (PAM 001) substitution model (the only supported model in version 3.6a of PROML) and a mixed four-category discrete-gamma model of among-site rate variation plus invariable sites (PAM plus Γ plus Inv). Ten random additions with global rearrangements were used to find the optimal tree (shown below in Fig. 1). The Γ shape parameter, α, was estimated to be 1.07, and the fraction of invariable sites, P_{inv}, was 0.01 using TREE-PUZZLE, version 4.02 (26). Protein ML distance bootstrap values for bipartitions were calculated by analysis of 500 resampled data sets using PUZZLE-BOOT (http://www.tree-puzzle.de) with a Jones-Taylor-Thornton (JTT) substitution model and a mixed eight-category discrete-gamma model (α = 1.25) of among-site rate variation plus invariable sites (P_{inv} = 0.01) (JTT plus Γ plus Inv). Due to the computational intensity, full ML bootstrapping was not performed on the 53-taxon data set.

In a separate phylogenetic analysis, the *sudA* and *sudB* gene products were combined, yielding 671 aa that could be unambiguously aligned. *Aquifex aeolicus*, *Geobacter sulfurreducens*, *Desulfovibrio vulgaris*, and *T. pallidum* failed the χ^2 amino acid composition test implemented in TREE-PUZZLE, version 4.02 (26), and were removed from further analysis. Protein ML phylogenies were inferred as described above except that an eight-category discrete-gamma model was used (α = 0.88 with no invariable sites detected). Protein ML bootstrap values for bipartitions were calculated by analysis of 500 resampled data sets using PROML within the PHYLIP package, version 3.6a (11), with one random addition with global rearrangements.

Nucleotide sequence accession numbers. The *G. lamblia sudA* sequence was assembled using the following single-pass reads deposited in GenBank by the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution (19) with the following accession numbers: AC030366, AC031039, AC031040, AC035392, AC043148, AC048816, AC051395, AC062548, AC084971, AC084972, AC086120, and AC082857. The *S. barkhanus* sequence reported here was deposited in GenBank under the accession number AF455034.

RESULTS AND DISCUSSION

A sulfide dehydrogenase gene in diplomonads. Among the sequences released from the ongoing genome project of the intestinal human parasite *G. lamblia* (19), we found a gene with greater similarities to prokaryotic genes encoding sulfide dehydrogenase than to any eukaryotic homologues, suggesting an LGT event from prokaryotes to an ancestor of *Giardia*. Furthermore, the gene appeared to represent a gene fusion, since it corresponded to two distinct genes in prokaryotes (*sudA* and *sudB*). No other sequences with significant homology to the small and/or large subunit of glutamate synthase were found among the available *G. lamblia* sequences. In order to determine if the sulfide dehydrogenase gene was a recent acquisition in the lineage leading to *Giardia* or if it was present in the last common ancestor of diplomonads, we amplified and sequenced the gene from the distantly related diplomonad *S. barkhanus*. The amplified gene did indeed show the highest similarity to the *Giardia* gene and appeared to be the result of the same gene fusion. Some diplomonads, including *S. barkhanus*, use an alternative genetic code where TAG and TAA encode glutamine rather than termination (17). Seven in-frame TAG codons and five in-frame TAA codons were found in the *S. barkhanus* sulfide dehydrogenase gene, which show that it has ameliorated to the genetic code used in the organism. To clarify the evolutionary origin of the gene in diplomonads, we retrieved homologues of the gene from the public databases and performed phylogenetic and gene order analyses.

Phylogenetic analysis of *gltD* and its homologues. The protein ML tree of the gene sequences of *gltD* and its homologues is shown in Fig. 1. The phylogenetic tree can be divided into three distinct parts: one well-supported clade (Fig. 1A) that includes mostly *gltD* sequences; one clade with *sudA* sequences (Fig. 1B); and a paraphyletic group (Fig. 1C) with six *gltD*-like sequences. However, the root of the tree is unknown and could be within one of these groupings. Within clade A, three distinct groups are found, two well-supported clades consisting of α- and γ-proteobacteria, respectively, and a large clade with a monophyletic eukaryotic clade nested within low-GC gram-positive bacteria, indicating an LGT from this eubacterial group to an early eukaryote (see discussion below) (Fig. 1A). With a few exceptions (see below), all eubacterial sequences within the *gltD* clade form a gene pair with *gltB*, which encodes the large subunit of glutamate synthase (Fig. 1A). This indicates that the glutamate synthase operon structure initially identified in *Escherichia coli* (5) is conserved within eubacteria, with the exception that the gene order between *gltB* and *gltD* is reversed in α-proteobacteria. In eukaryotes, glutamate synthase is encoded by a single gene, *gltS*, which arose by gene fusion of the eubacterial *gltB* and *gltD* genes (Fig. 1A).

Interestingly, there are also three eubacterial *gltD*-like sequences—two *aegA* genes from *E. coli* (6) and the *dsrL* gene from *Allochromatium vinosum* (GenBank accession number U84760)—nested among the *gltD* sequences (Fig. 1A). These genes probably encode functions distinct from glutamate synthase; they have N- or C-terminal extensions unrelated to the large subunit of glutamate synthase (Fig. 1) and they are not linked to *gltB*. It is likely that these enzymes have originated from gene duplications of *gltD*, followed by fusions with different domains, and they have acquired new functions in these γ-proteobacteria.

A clade consisting of sulfide dehydrogenase genes. A second large set of gene order conservation, including the *P. furiosus* sulfide dehydrogenase genes, was found for clade B (Fig. 1B). The *sudB* genes are found upstream of, or fused to, the *sudA* genes in all species in the clade except *A. aeolicus*. Gene order conservation is rare in distantly related prokaryotes and is very unlikely to indicate ancient organization that was retained solely by chance (29). Therefore, the gene order conservation between the *sudA* and *sudB* homologues (Fig. 1) strongly suggests that the gene products functionally, and probably physically, associate (interact) in vivo. Three iron-cluster-binding motifs have been proposed for the *P. furiosus* sulfide dehydrogenase (15). These show a high degree of conservation within the clade; all positions are conserved in all species containing the *sud* gene pair except in *G. lamblia*, which has substitutions in 2 of the 11 proposed iron-binding positions (Fig. 2). The monophyly of *sudA* (Fig. 1), the gene order conservation of *sudA* and *sudB* (Fig. 1), and the conservation of the iron-cluster-binding motifs strongly suggest that the gene pair (or a single gene in the diplomonad case) encodes sulfide dehydrogenases in these organisms. The most likely origin for the gene in the two diplomonads is by LGT, since the gene sequences are nested among prokaryotes and the gene is only distantly related to other eukaryotic sequences (see discussion below).

***gltD*-like sequences.** Finally, there is a paraphyletic group of six *gltD*-like sequences that are linked neither to the large subunit glutamate synthase nor to *sudB* (Fig. 1C). Four of these sequences do have N- and/or C-terminal extensions that indicate that the *gltD*-like part of the polypeptide is likely functioning in a different context in these proteins. Possibly, the two *gltD*-like sequences that lack significant N- or C-terminal extensions are functioning together with other, not-yet-identified genes. Preliminary phylogenetic analyses indicated that the partial *E. histolytica* sequence that was excluded from the large phylogenetic analysis due to a strong amino acid composition bias probably belongs to this paraphyletic group (data not shown).

Several cases of LGT within the sulfide dehydrogenase clade. Preliminary individual phylogenetic reconstructions showed congruity between the *sudA* and *sudB* gene topologies (data not shown), suggesting that they share a common history. We therefore concatenated the two proteins in a combined phylogeny to increase the resolution within the sulfide dehydrogenase clade. Five nodes in the protein ML tree of the *sud* genes show strong bootstrap support (Fig. 3). The *Pyrococcus* and the diplomonad sequences form two strongly supported monophyletic clusters, as expected (Fig. 3). The *Fibrobacter succinogenes* sequence forms a clade with the *Treponema denticola* sequence, with the *Porphyromonas gingivalis* sequence as an immediate outgroup. Finally, the low-GC gram-positive *Enterococcus faecalis* forms a strongly supported monophyletic grouping with *T. maritima* (Fig. 3), which is inconsistent with accepted views of eubacterial phylogeny (4) and which strongly suggests an LGT event between the two lineages. This is unlikely to be a rare case of LGT in the history of the *sud* genes. The scattered distribution of the *sud* genes among microbes from the three domains of life is unlikely to be solely due to differential gene loss; clearly, that would require a vast number of independent gene losses in all three domains of life. Thus, propagation of the *sud* genes via multiple LGT events combined with vertical transmission is the most parsimonious explanation of the distribution of *sud* genes among microbes. Since the gene is only found in diplomonads among the studied eukaryotes and the diplomonads are nested within a prokaryotic clade (Fig. 1 and 3), an ancestor of diplomonads almost certainly acquired the gene from a prokaryotic source via LGT. Unfortunately, the donor lineage cannot be determined due to the limited resolution in the phylogenetic reconstructions (Fig. 1B and 3).

It would be imprudent to speculate on what might have driven the propagation of the gene, since the biological activity of sulfide dehydrogenase is not very well understood (1, 15). At any rate, all organisms in the *sud* clade are anaerobes (Fig. 1B), which hints at a biological role related to anaerobiosis for the protein. However, only biochemical studies can prove or disprove if the gene products have sulfide dehydrogenase activity in the members of the clade. Such experiments in diplomonads, combined with genome data from the ongoing *G. lamblia* genome project (19), will deepen our understanding of sulfur metabolism in these organisms, which currently is very limited. It may turn out that diplomonads metabolize elemental sulfur to modulate redox potential or as an alternative energy source. If so, it would be another example where anaerobic eukaryotes metabolize inorganic sulfur compounds; some animals use sulfide as an inorganic energy source during mitochondrial sulfide oxidation (14).

A

SCHIZOSACCHAROMYCES
SACCHAROMYCES — 99
NEUROSPORA
CAENORHABDITIS — 87
DROSOPHILA — 95
ARABIDOPSIS
Bacillus subtilis — 51
Staphylococcus
Bacillus halodurans — 72
Clostridium — 60
Synechocystis
Lactococcus
Vibrio 2 — 96
Corynebacterium — 75
Mycobacterium — 97
Streptomyces
Mesorhizobium 1 — 87
Deinococcus

72
58
93

Zymomonas — 63
Azospirillum — 97
Caulobacter — 99
Rhodobacter

Xylella
Thiobacillus
Pseudomonas
Escherichia 2
Vibrio 1 — 97
Escherichia 4 — 96
Escherichia 1 — 68

FUNGI

METAZOA

VIRIDIPLANTAE

Low GC Gram Positives

Cyanobacteria
Low GC Gram Positives
γ-Proteobacteria

High GC Gram Positives

α-Proteobacteria
Deinococcales

Allochromatium γ-Proteobacteria

α-Proteobacteria

γ-Proteobacteria

γ-Proteobacteria

gltD (gltS)

gltD

gltD-like (dsrL)

gltD

gltD

gltD-like (aegA)

B

Thermotoga 1 — 89
Enterococcus
Clostridium
GIARDIA — 99
SPIRONUCLEUS
Carboxydothermus
Treponema pallidum — 86
Treponema denticola — 98
Fibrobacter
Dehalococcoides
Desulfovibrio
Geobacter — 79
Porphyromonas
Chlorobium
Pyrococcus 1 — 99
Pyrococcus 2
Aquifex

54

Thermotogales
Low GC Gram Positives

DIPLOMONADIDA

Low GC Gram Positives

Spirochetes

Fibrobacter/Acidobacteria group
Green Non-Sulfur bacteria

δ-Proteobacteria

Bacteroidaceae
Green Sulfur bacteria

Euryarchaeota

Aquificales

sudA

sudA

sudA

sudA

C

Escherichia 5 γ-Proteobacteria gltD-like
Escherichia 3 γ-Proteobacteria gltD-like
Mesorhizobium 2 α-Proteobacteria gltD-like
Thermoplasma Euryarchaeota gltD-like
Thermotoga 2 Thermotogales gltD-like
Moorella Low GC Gram Positives gltD-like (fdhA)

0.1

gltD, sudA or gltD-like gene
gltB gene
sudB gene
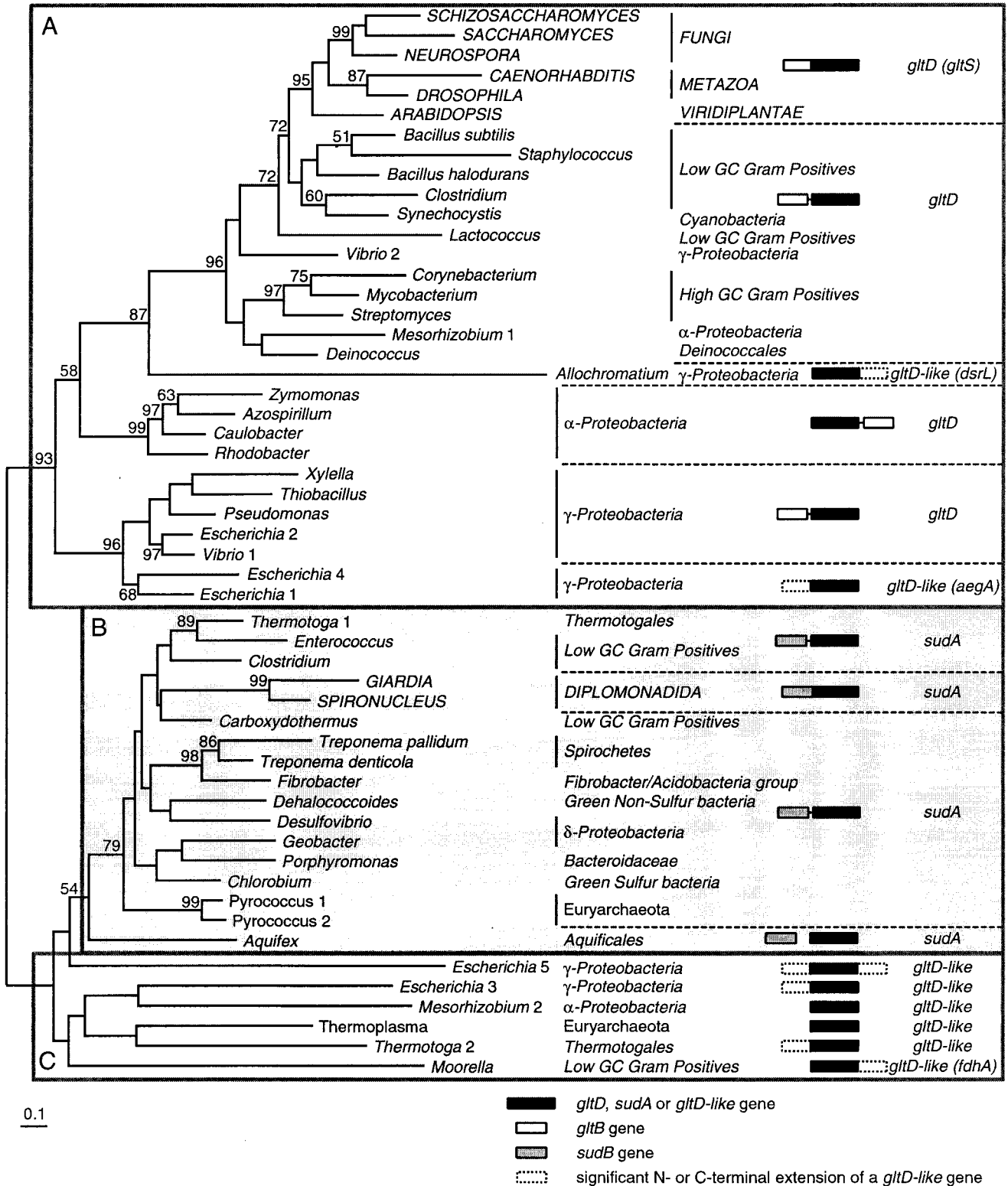significant N- or C-terminal extension of a gltD-like gene

FIG. 1. The ML tree of the inferred amino acid sequence of *gltD*, *sudA*, and *gltD*-like genes and their gene arrangements. See Materials and Methods for gene nomenclature. The different typefaces of the species names indicate whether the species belongs to the eukaryotes (all capitals), *Archaea* (lowercase, normal), or *Bacteria* (lowercase italic). Three main parts of the tree are indicated with large boxes, as follows: a clade with mostly *gltD* genes (A); a clade with only *sudA* genes (B); and a paraphyletic group with different *gltD*-like genes (C). The tree was arbitrarily rooted between the *gltD* clade and the paraphyletic group of *gltD*-like genes. Black boxes represent the homologous part of the *gltD*, *gltD*-like, and *sudA* sequences (used in analysis), open boxes represent *gltB* genes, gray boxes represent *sudB* genes, and boxes with dotted borders represent unique N- or C-terminal extensions of the *gltD*-like genes. Note that there is no detectable sequence homology between the extensions represented by dotted boxes. A thin line between two boxes indicates neighboring genes on the chromosome. Two boxes attached to each other indicate fusion of the genes. Protein ML-distance bootstrap values >50% for bipartitions are shown.

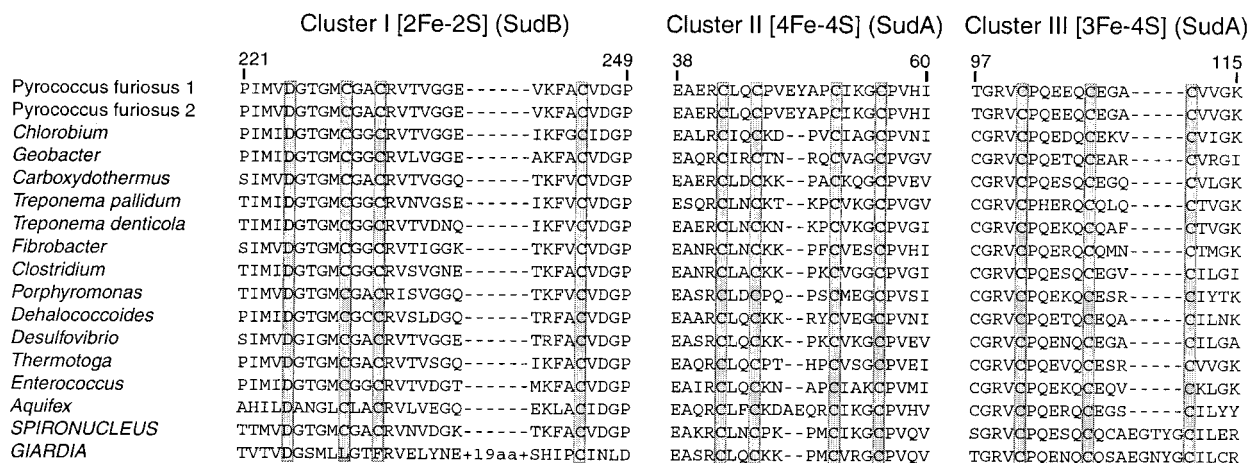|	Cluster I [2Fe-2S] (SudB)	Cluster II [4Fe-4S] (SudA)	Cluster III [3Fe-4S] (SudA)
	221　　　　　　　　　　249	38　　　　　　60	97　　　　　115
Pyrococcus furiosus 1	PIMVDGTGMCGACRVTVGGE------VKFACVDGP	EAERCLQCPVEYAPCIKGCPVHI	TGRVCPQEEQCEGA-----CVVGK
Pyrococcus furiosus 2	PIMVDGTGMCGACRVTVGGE------VKFACVDGP	EAERCLQCPVEYAPCIKGCPVHI	TGRVCPQEEQCEGA-----CVVGK
Chlorobium	PIMIDGTGMCGGCRVTVGGE------IKFGCIDGP	EALRCIQCKD--PVCIAGCPVNI	CGRVCPQEDQCEKV-----CVIGK
Geobacter	PIMIDGTGMCGGCRVLVGGE------AKFACVDGP	EAQRCIRCTN--RQCVAGCPVGV	CGRVCPQETQCEAR-----CVRGI
Carboxydothermus	SIMVDGTGMCGACRVTVGGE------TKFVCVDGP	EAERCLDCKK--PACKQGCPVEV	CGRVCPQESQCEGQ-----CVLGK
Treponema pallidum	TIMIDGTGMCGGCRVNVGSE------IKFVCVDGP	ESQRCLNCKT--KPCVKGCPVGV	CGRVCPHERQCQLQ----CTVGK
Treponema denticola	TIMIDGTGMCGGCRVTVDNQ------IKFVCVDGP	EAERCLNCKN--KPCVKGCPVGI	CGRVCPQEKQCQAF-----CTVGK
Fibrobacter	SIMVDGTGMCGGCRVTIGGK------TKFVCVDGP	EANRCLNCKK--PFCVESCPVHI	CGRVCPQERQCQMN-----CTMGK
Clostridium	TIMIDGTGMCGGCRVSVGNE------TKFACVDGP	EANRCLACKK--PKCVGGCPVGI	CGRVCPQESQCEGV-----CILGI
Porphyromonas	TIMVDGTGMCGACRISVGGQ------TKFVCVDGP	EASRCLDCPQ--PSCMEGCPVSI	CGRVCPQEKQCESR-----CIYTK
Dehalococcoides	PIMIDGTGMCGCCRVSLDGQ------TRFACVDGP	EAARCLQCKK--RYCVEGCPVNI	CGRVCPQETQCEQA-----CILNK
Desulfovibrio	SIMVDGIGMCGACRVTVGGE------TRFACVDGP	EASRCLQCKK--PKCVKGCPVEV	CGRVCPQENQCEGA-----CILGA
Thermotoga	PIMVDGTGMCGACRVTVSGQ------IKFACVDGP	EAQRCLQCPT--HPCVSGCPVEI	CGRVCPQEVQCESR-----CVVGK
Enterococcus	PIMIDGTGMCGGCRVTVDGT------MKFACVDGP	EAIRCLQCKN--APCIAKCPVMI	CGRVCPQEKQCEQV-----CKLGK
Aquifex	AHILDANGLCLACRVLVEGQ------EKLACIDGP	EAQRCLFCKDAEQRCIKGCPVHV	CGRVCPQERQCEGS-----CILYY
SPIRONUCLEUS	TTMVDGTGMCGACRVNVDGK------TKFACVDGP	BAKRCLNCPK--PMCIKGCPVQV	SGRVCPQESQCQCAEGTYGCILER
GIARDIA	TVTVDGSMLLGTFRVELYNE+19aa+SHIPCINLD	EASRCLQCKK--PMCVRGCPVQV	TGRVCPQENQCQSAEGNYGCILCR

FIG. 2. Amino acid alignment of the iron-cluster binding motifs in sulfide dehydrogenase, proposed by Hagen et al. (15). The numbers above the alignments refer to the amino acid position in the *P. furiosus* SudB (cluster I) and SudA (cluster II and III) proteins. The proposed ligands are indicated by shadowed boxes. Note that cluster I has been proposed to have a novel Asp(Cys)$_3$ binding motif (15).

Eukaryotic *gltS* may have a eubacterial origin via LGT. The LGT of the *sud* genes from a prokaryote to the diplomonad lineage may not be the only case of prokaryote-to-eukaryote transfer in the evolution of the *gltD* and *gltD*-like genes. Interestingly, eukaryotic *gltS* is clearly nested (bootstrap support value of 96) within a bacterial clade consisting of mostly gram-positive bacteria (Fig. 1A). This is unlikely the result of an endosymbiotic gene replacement from the mitochondria or chloroplast, since the eukaryotic sequences fail to show affinity to proteobacteria or cyanobacteria, the ancestors of mitochondria and chloroplasts, respectively. Therefore, an origin via LGT independent of the eukaryotic organelles, perhaps from a

gram-positive eubacterium in the common ancestor of animals, plants, and fungi, is the most likely scenario.

The *P. falciparum gltS* sequence (excluded from the large phylogenetic analysis due to its strong amino acid composition bias) showed affinity to the large bacterial clade in preliminary phylogenetic analyses but failed to branch with the other eukaryotic *gltS* sequences (data not shown). However, the *P. falciparum* sequence appears to have experienced the same fusion event as the other eukaryotic *gltS* sequences, indicating that the failure of this sequence to group with other eukaryotic *gltS* sequences is likely an artifact caused by its extreme amino acid composition bias.
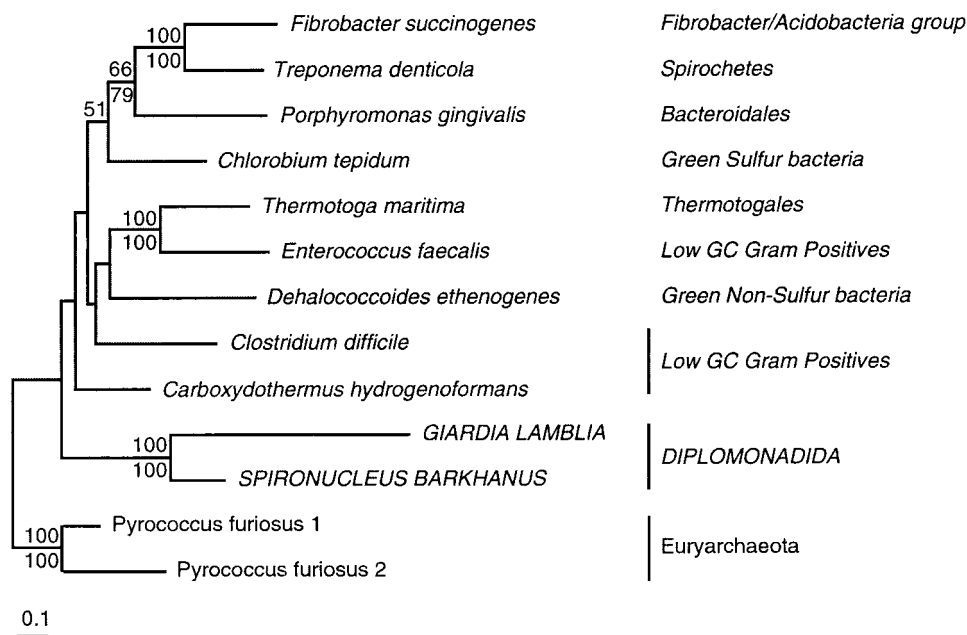

FIG. 3. Phylogeny of concatenated *sudA* and *sudB* protein sequences. Protein ML bootstrap and ML distance bootstrap support values are shown above and below the branches, respectively. Only support values >50% for bipartitions are shown.

Two examples of gene fusion events of prokaryotic operons in eukaryotes. The gene arrangements of glutamate synthase and sulfide dehydrogenase in prokaryotes versus eukaryotes are strikingly similar; in both cases a conserved gene pair occurs in prokaryotes that corresponds to a single fused gene in eukaryotes. Furthermore, in both cases, the eukaryote homologues probably originated via LGT from a prokaryotic source. The probability of successful transfers of the two subunits simultaneously was certainly increased by the conserved gene arrangements within prokaryotes. The gene fusions in eukaryotes likely reflect the differences in gene regulation between eukaryotes and prokaryotes. Fusions of genes that code for different subunits of the same enzyme are probably strongly selectively advantageous in eukaryotes, which lack operon structures, because they provide an easy solution for coregulation of the expression of the different subunits.

The gene fusions reported in this study probably resulted from two distinct fusion events: one that created the eukaryotic glutamate synthase gene and one that yielded the diplomonad sulfide dehydrogenase gene. Both appear to have occurred relatively soon after the LGT events. However, the evolutionary signal in gene fusions should not be overstated. The presence of similar fusions in distantly related eukaryotes does not necessarily mean that the genes are the result of a single transfer followed by a fusion. Two independent transfers of the same prokaryotic operon to two different eukaryotic lineages can easily lead to two independent gene fusions, since selection for coordinate regulation probably strongly favors such fusion events. Therefore, analyses of gene fusions should be interpreted in conjunction with phylogenetic analyses based on gene sequences.

LGT might play an important role in protistan genome evolution. Although it is always difficult to discern mechanisms for ancient evolutionary events, it has been suggested that a gene transfer ratchet mechanism may exist for eukaryotes whereby genes from prokaryotes taken up as food are incorporated into the eukaryotic nucleus (9). Such a mechanism would explain the transfer of sulfide dehydrogenase from a prokaryote to the common ancestor of diplomonads, which likely engulfed prokaryotes as food, as well as the putative transfer of glutamate synthase from a gram-positive eubacterium to a common ancestor of fungi, animals, and plants—an ancestor that most certainly was a single-celled organism.

Our analysis of sulfide dehydrogenase, together with several recently reported cases of prokaryote-to-eukaryote gene transfers (3, 8, 12, 16, 27), indicates that LGT may play an important role in protistan genome evolution. In addition, the putative transfer of glutamate synthase suggests that LGT also shaped the genomes of the protist ancestors of major multicellular eukaryote lineages.

ACKNOWLEDGMENTS

We thank T. Martin Embley and David S. Horner (Natural History Museum, London, United Kingdom) for a generous gift of genomic *S. barkhanus* DNA. We also thank Yuji Inagaki, Camilla Nesbø, and Åsa Sjögren for discussions and critical reading of the manuscript, and we thank Lesley Davis for help with sequencing. We utilized the preliminary data generated by The Institute for Genomic Research (www.tigr.org), the Utah Genome Center (www.genome.utah.edu), and the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution (www.mbl.edu/Giardia).

REFERENCES

1. **Adams, M. W. W., J. F. Holden, A. L. Menon, G. J. Schut, A. M. Grunden, C. Hou, A. M. Hutchins, F. E. Jenney, C. Kim, K. Ma, G. Pan, R. Roy, R. Sapra, S. V. Story, and M. F. Verhagen.** 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **183:**716–724.
2. **Andersson, J. O., W. F. Doolittle, and C. L. Nesbø.** 2001. Are there bugs in our genome? Science **292:**1848–1850.
3. **Boucher, Y., and W. F. Doolittle.** 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. Mol. Microbiol. **37:**703–716.
4. **Brown, J. R., C. J. Douady, M. J. Italia, W. E. Marshall, and M. J. Stanhope.** 2001. Universal trees based on large combined protein sequence data sets. Nat. Genet. **28:**281–285.
5. **Castano, I., F. Bastarrachea, and A. A. Covarrubias.** 1988. *gltBDF* operon of *Escherichia coli*. J. Bacteriol. **170:**821–827.
6. **Cavicchioli, R., T. Kolesnikow, R. C. Chiang, and R. P. Gunsalus.** 1996. Characterization of the *aegA* locus of *Escherichia coli*: control of gene expression in response to anaerobiosis and nitrate. J. Bacteriol. **178:**6968–6974.
7. **Chen, X., and R. Wu.** 1997. Direct amplification of unknown genes and fragments by uneven polymerase chain reaction. Gene **185:**195–199.
8. **de Koning, A. P., F. S. L. Brinkman, S. J. M. Jones, and P. J. Keeling.** 2000. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. Mol. Biol. Evol. **17:**1769–1773.
9. **Doolittle, W. F.** 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. **14:**307–311.
10. **Doolittle, W. F.** 1999. Phylogenetic classification and the universal tree. Science **284:**2124–2129.
11. **Felsenstein, J.** 1989. Phylogeny inference package (version 3.2). Clad. **5:**166.
12. **Field, J., B. Rosenthal, and J. Samuelson.** 2000. Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to Entamoeba histolytica. Mol. Microbiol. **38:**446–455.
13. **Foster, P. G., and D. A. Hickey.** 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. **48:**284–290.
14. **Grieshaber, M. K.** 1998. Animal adaptations for tolerance and exploitation of poisonous sulfide. Annu. Rev. Physiol. **60:**33–53.
15. **Hagen, W. R., P. J. Silva, M. A. Amorim, P.-L. Hagedoorn, H. Wassink, H. Haaker, and F. T. Robb.** 2000. Novel structure and redox chemistry of the prosthetic groups of the iron-sulfur flavoprotein sulfide dehydrogenase from *Pyrococcus furiosus*; evidence for a [2Fe-2S] cluster with Asp(Cys)$_3$ ligands. J. Biol. Inorg. Chem. **5:**527–534.
16. **Henze, K., D. S. Horner, S. Suguri, D. V. Moore, L. B. Sánchez, M. Müller, and T. M. Embley.** 2001. Unique phylogenetic relationship of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis*. Gene **281:**123–131.
17. **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.
18. **Ma, K., and M. W. W. Adams.** 1994. Sulfide dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*: a new multifunctional enzyme involved in the reduction of elemental sulfur. J. Bacteriol. **176:**6509–6517.
19. **McArthur, A. G., H. G. Morrison, J. E. Nixon, N. Q. Passamaneck, U. Kim, G. Hinkle, M. K. Crocker, M. E. Holder, R. Farr, C. I. Reich, G. E. Olsen, S. B. Aley, R. D. Adam, F. D. Gillin, and M. L. Sogin.** 2000. The *Giardia* genome project database. FEMS Microbiol. Lett. **189:**271–273.
20. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleishmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature **399:**323–329.
21. **Nesbø, C. L., S. L'Haridon, K. O. Stetter, and W. F. Doolittle.** 2001. Phylogenetic analyses of two "archaeal" genes in *Thermotoga maritima* reveal multiple transfers between Archaea and Bacteria. Mol. Biol. Evol. **18:**362–375.
22. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299–304.
23. **Salzberg, S. L., O. White, J. Peterson, and J. A. Eisen.** 2001. Microbial genes in the human genome: lateral transfer or gene loss? Science **292:**1903–1906.
24. **Stanhope, M. J., A. Lupas, M. J. Italia, K. K. Koretke, C. Volker, and J. R.**

Brown. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature **411:**940–944.

25. **Sterud, E., T. A. Mo, and T. T. Poppe.** 1997. Ultrastructure of *Spironucleus barkhanus* n. sp. (Diplomonadida: Hexamitidae) from grayling *Thymallus thymallus* (L.) (Salmonidae) and Atlantic salmon *Salmo salar* L. (Salmonidae). J. Eukaryot. Microbiol. **44:**399–407.

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

27. **Suguri, S., K. Henze, L. B. Sánchez, D. V. Moore, and M. Müller.** 2001. Archaebacterial relationships of the phosphoenolpyruvate carboxykinase gene reveal mosaicism of *Giardia intestinalis* core metabolism. J. Eukaryot. Microbiol. **48:**493–497.

28. **Vanoni, M. A., and B. Curti.** 1999. Glutamate synthase: a complex iron-sulfur flavoprotein. Cell. Mol. Life Sci. **55:**617–638.

29. **Wolf, Y. I., I. B. Rogozin, A. S. Kondrashov, and E. V. Koonin.** 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res. **11:**356–372.