GAPDH Gene Diversity in Spirochetes: A Paradigm for Genetic Promiscuity

Rainer Martin Figge and Rüdiger Cerff

Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, 38106 Braunschweig, Germany

In this study we have determined *gap* sequences from nine different spirochetes. Phylogenetic analyses of these sequences in the context of all other available eubacterial and a selection of eukaryotic Gap sequences demonstrated that the eubacterial glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene diversity encompasses at least five highly distinct gene families. Within these gene families, spirochetes show an extreme degree of sequence divergence that is probably the result of several lateral gene transfer events between spirochetes and other eubacterial phyla, and early gene duplications in the eubacterial ancestor. A Gap1 sequence from the syphilis spirochete *Treponema pallidum* has recently been shown to be closely related to GapC sequences from Euglenozoa. Here we demonstrate that several other spirochetal species are part of this cluster, supporting the conclusion that an interkingdom gene transfer from spirochetes to Euglenozoa must have occurred. Furthermore, we provide evidence that the GAPDH genes present in the protists Parabasalia may also be of spirochetal descent.

Introduction

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (phosphorylating, gene designation *gap*) is a key enzyme in glycolysis (Forthergill-Gilmore and Michels 1993) and the Calvin cycle (Martin and Schnarrenberger 1997), where it catalyzes the reversible interconversion between glyceraldehyde-3-phosphate and 1,3-diphosphoglycerate. Phosphorylating GAPDH has been shown to exist in two classes (Cerff 1995). Whereas the class I enzyme is widely distributed in eukaryotes and eubacteria (Henze et al. 1995) and is also found in one archaebacterium (Prüss, Meyer, and Holldorf 1993; Brinkmann and Martin 1996), class II enzymes seem to be restricted to archaebacteria (Hensel et al. 1989). Recently, we showed that class I GAPDH encompasses three different gene families which were given the operational designations Gap I, Gap II, and Gap III (Figge et al. 1999).

Notably, most eukaryotic GAPDH genes belong to types Gap I and Gap II. Within Gap I, the eukaryotic *GapC* gene of mainly glycolytic function is closely related to proteobacterial *gap1* and was possibly acquired via endosymbiotic gene transfer from the alpha-proteobacterial antecedents of mitochondria (Smith 1989; Martin et al. 1993; Liaud et al. 1994). Similarly, within Gap II, the nuclear genes *GapA* and *GapB*, encoding subunits A and B of Calvin cycle GAPDH (Cerff 1982), most likely originated from cyanobacterial *gap2* via gene transfer in the context of plastidal endosymbiosis (Brinkmann et al. 1989; Martin et al. 1993; Liaud et al. 1994). Apart from these prominent members of the GAPDH gene family found in eukaryotic nuclei, three other eukaryotic GAPDH lineages exist that are clearly not of mitochondrial or chloroplast descent: (1) Some Parabasalia harbor a distinct eukaryotic GAPDH gene, the origin of which has so far remained obscure (Markos, Miretsky, and Müller 1993; Viscogliosi and Müller 1998). (2) A cytosolic *GapC* gene is present in some trypanosomes that is surprisingly close to gamma-proteobacterial *gap1* (Michels et al. 1991) and has probably been acquired via a recent lateral gene transfer (LGT) from a gamma-proteobacterial donor (Henze et al. 1995; Figge et al. 1999). (3) Cytosolic *GapC* in *Euglena* (Henze et al. 1995) and its glycosomal equivalent in trypanosomes (Michels et al. 1991) have recently been shown to be closely related to a eubacterial *gap1* so far found only in the syphilis spirochete *Treponema pallidum* (Figge et al. 1999). In order to explain this curious finding, LGT between a spirochetal and a euglenozoan ancestor has been invoked. The direction of this transfer, however, could not be clarified.

LGT has just recently been recognized as a major factor in eubacterial and eukaryotic evolution (Doolittle 1999a, 1999b; Lake, Jain, and Rivera 1999). For example, it has been suggested that about 18% of the *Escherichia coli* genome is lateral acquisitions that happened during the last 100 million years (Lawrence and Ochman 1998). However, it seems that not all genes are affected by LGT to the same extent, as outlined by Rivera et al. (1998). According to these authors, operational genes (encoding functions such as energy metabolism, biosynthesis of amino acids, cofactors, etc.) are transferred easily and often, whereas informational genes (encoding genes that function in e.g., RNA, DNA synthesis and translation) are rarely passed on to another organism. Nevertheless, even informational genes such as 16S rRNA or ribosomal genes may not be safe from LGT. In fact, it was shown that 16S rRNA genes from *Es. coli* could be replaced with the corresponding genes from *Proteus vulgaris* (Asai et al. 1999) and that the ribosomal *rpS*14 gene may have undergone multiple LGT events (Brochier, Philippe, and Moreira 2000).

In order to determine the direction of the LGT between Euglenozoa and the spirochete *T. pallidum*, we have now determined *gap* genes from nine other spirochetes. Spirochetes are easily distinguishable on the basis of their unique morphology and mechanism of motility and have been shown to constitute a phylogenetically distinct eubacterial phylum (Paster et al. 1991 and

Key words: glyceraldehyde-3-phosphate dehydrogenase, gene diversity, spirochetes, lateral gene transfer, eukaryotic evolution.

Address for correspondence and reprints: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Avenue, Los Angeles, California 90095-1569. E-mail: rfigge@chem.ucla.edu.

Mol. Biol. Evol. 18(12):2240–2249. 2001

references therein). The analysis of the spirochetal sequences in the context of all other eubacterial and a selection of eukaryotic sequences revealed that the GAPDH gene diversity within spirochetes is much broader than previously assumed. LGT between spirochetes and other eubacterial phyla, in addition to early gene duplications in the eubacterial ancestor, seems to account for the phylogenetic distribution of the spirochetal Gap genes presently observed.

Material and Methods

Bacterial Strains and Plasmids

Frozen cultures of the spirochetes, *Serpulina murdochii*, *Spirochaeta aurantia*, *Spirochaeta stenostrepta*, and *T. saccharophilum*, were purchased from the Deutsche Sammlung für Mikroorganismen, Braunschweig, Germany. Cultures from the spirochetes, *Leptonema illini* and *Leptospira biflexa*, were obtained from the Royal Tropical Institute, Amsterdam, the Netherlands.

Escherichia coli strains XL1-blue, and HB101 were grown with appropriate antibiotics according to standard procedures (Sambrook, Fritsch, and Maniatis 1989). Cloning vectors used were pBluescript-SK (+) (Stratagene) and pUC18 (Vieira and Messing 1982). Plasmid transformation, selection, and testing for recombinant clones were performed as described (Sambrook, Fritsch, and Maniatis 1989).

DNA Isolation, Cloning, and Sequencing of Eubacterial *gap* Genes

DNA from spirochetes was isolated as follows: spirochete cultures were washed once in TES (10 mM Tris-HCl, pH 8.0; 0.1 M EDTA, pH 8.0; 0.15 M NaCl) and resuspended in the same buffer. Subsequently, lysozyme (*ad* [final concentration] 5 mg/ml), SDS (*ad* 1%), and proteinase K (*ad* 50 μg/ml) were added independently after 1-h incubation periods at 37°C. Then, samples were treated twice with phenol-chloroform-isoamylalcohol (25:24:1) and precipitated.

Genomic DNA from *Prochlorococcus marinus* CCMP1375 was a gift from Wolfgang Hess, Humboldt Universität, Berlin. Isabelle St. Girons, Institut Pasteur, Paris, provided DNA from *Leptospira biflexa* and *Brachyspira hyodysenteriae*. Klaus Heuner, Klinikum Charité, Berlin, supplied the DNA from *T. denticola*.

Degenerate primers (G/AINGFG: 5′-GSNATHAAYGGNTTYGG-3′; WYDNEW: 5′-CCAYTCRTTRTCRTACCA-3′) for two highly conserved regions at the N-terminal and C-terminal ends of GAPDH proteins (INGFGRI, WYDNE) were used for the amplification of *gap* genes (95% of the coding sequence) from genomic DNA samples. Polymerase chain reaction conditions were as follows—cycle 1: 93°C for 5 min; cycles 2–35: 93°C for 1 min, 50°C for 1 min, and 72°C for 2 min; and cycle 36: 72°C for 5 min for the amplification of *gap* genes from *Prochlorococcus marinus*. In the case of *Sp. stenostrepta*, *Sp. aurantia*, *S. murdochii*, and *T. saccharophilum*, the following conditions were used—cycle 1: 94°C for 3 min; cycles 2–35: 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min; and cycle 36: 72°C for 7 min. For *Leptospira biflexa*, *Leptospira interrogans*, *L. illini*, *B. hyodysenteriae*, and *T. denticola*, the following conditions were applied—cycle 1: 96°C for 3 min; cycles 2–35: 96°C for 30 s, 48°C for 30 s, and 72°C for 1 min; and cycle 36: 72°C for 7 min. All reactions were performed in a Perkin-Elmer thermocycler with a Mg^{2+} concentration of 1.5 mM and 40–200 ng of DNA per reaction (100 μl). Amplification products of appropriate size were eluted from agarose gels (Nucleospin, Macherey, and Nagel), treated with polynucleotide kinase and Klenow DNA polymerase using protocols of the supplier (New England Biolabs), and cloned into pBluescript-SK (+) or pUC18. Both strands of several independent PCR-generated clones for each gene were sequenced with appropriate oligonucleotides using the dideoxy chain termination method (Big dye, Perkin-Elmer).

The nucleotide sequences of the *gap* genes reported will appear in DDBJ/EMBL/GenBank International Nucleotide Sequence Database under the accession numbers AJ245541 (*P. marinus* CCMP1375); AJ245542 (*Sp. aurantia* DSM 1902); AJ245543 (*Sp. stenostrepta* DSM 2028); AJ245544 (*T. saccharophilum* DSM 2985); AJ245545 (*S. murdochii* ATCC 51284); AJ245546 (*Leptospira biflexa* Patoc 1); AJ245547 (*L. illini*); AJ245548 (*Leptospira interrogans* pathovar *icterohaemorragiae*); AJ245549 (*B. hyodysenteriae* ATCC 27164); AJ245550 (*T. denticola* ATCC 33521).

Phylogenetic Analysis

GAPDH sequences used for phylogenetic analysis were retrieved from Genbank with the following exceptions: *Bordetella pertussis*, *Chlorobium tepidum*, *Clostridium acetobutylicum*, *Mycobacterium avium*, *Neisseria gonorrhoeae*, *Porphyromonas gingivalis*, *Salmonella typhi*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Streptomyces coelicolor*, and *Yersinia pestis*. These sequences were all obtained from TIGR (see The Institute for Genomic Research website at http://www.tigr.org) as preliminary sequence data.

Deduced amino acid sequences were aligned with CLUSTAL W (Thompson, Higgins, and Gibson 1994) and refined by hand. The complete sequence alignment will appear at http://www.ebi.ac.uk:80/embl/Submission/alignment.html under the accession number DS45110. After the exclusion of gaps and the elimination of the N- and C-terminal regions including GINGFG and WYDNE, the resulting alignment contained 110 sequences with 250 positions. Protein phylogeny was inferred using maximum-likelihood (ML) (Molphy, version 2.3 [Adachi and Hasegawa 1996]), starting with a Neighbor-Joining (NJ) tree that was subsequently improved using the local rearrangement option of ProtML.

The robustness of the tree topology was estimated by 100 NJ-bootstrap replicates (Saitou and Nei 1987) that were calculated based on the JTT-F matrix. The same alignment was used in a maximum parsimony (MP) bootstrap (PAUP 4.0 alpha version, 2000) analysis (250 bootstrap replicates). To this end, a full heuristic search with four times random addition of the input or-

der was performed, and only bootstrap values above 50% were retained. In order to investigate the relationship between species in the Gap IB subtree, data sets containing 10, 12, or 14 sequences with 293 positions were analyzed with TREE-PUZZLE (Strimmer and von Haeseler 1996). The JTT-F amino acid substitution matrix (Jones, Taylor, and Thornton 1992) was employed with an eight-category discrete gamma model of site rate heterogeneity.

Signature sequences were defined on the basis of their presence in at least 90% of all sequences pertaining to the corresponding type of GAPDH and no more than 10% of all other GAPDH sequences included in this analysis.

Results

Phylogenetic Analysis of Currently Available Prokaryotic Gap Genes Points to the Existence of at Least Five Highly Distinct GAPDH Gene Families

Recently, we investigated the phylogenetic relationship between homologues of the three *gap* genes (*gap1*, *gap2*, and *gap3*) present in cyanobacteria. We demonstrated that these sequences belong to three distinct subtrees that we named Gap I, Gap II, and Gap III, respectively (Figge et al. 1999). Here, we have analyzed all the currently available prokaryotic and a selection of eukaryotic GAPDH sequences together with nine newly established partial spirochetal *gap* sequences (see later). Two hundred fifty positions of 110 GAPDH sequences (87 species) were used to construct phylogenetic trees using the distance matrix (NJ), MP, and protein ML (ProtML) methods. As the resulting topologies did not differ significantly, only the tree topology resulting from the ProtML analysis is shown in figure 1. In order to estimate the reliability of the tree topology, bootstrap values were calculated for both the NJ (100) and the MP (250) analyses, and are indicated above and below the nodes, respectively. The resulting tree displays five distinct subtrees, three of which were recently named types Gap I, Gap II, and Gap III. Each subtree is defined by 100% bootstrap support and by the presence of species from at least two distinct bacterial phyla or one eubacterial phylum and eukaryotic sequences (Gap II). In this study we extend this operational nomenclature by defining two additional types, Gap IV and Gap V. The species composition of subtrees Gap I and Gap II has already been described (Figge et al. 1999). Contrary to what was described in Figge et al. (1999), we wish to restrict the term Gap III to those sequences that are truly separated by 100% bootstrap support (fig. 1). Therefore, other loosely associated gamma-proteobacterial Gap sequences are now excluded from the definition of Gap III. The Gap IV subtree encompasses sequences from three different phyla, proteobacteria, gram positives with low GC-content, and spirochetes. The Gap V subtree is composed of sequences from gram positives with high GC-content and gamma-proteobacteria. Gap I itself encompasses two major subtrees that will subsequently be referred to as Gap IA and Gap IB, where Gap IB harbors spirochetal and euglenozoan sequences, and Gap IA includes all other Gap1/GapC sequences.

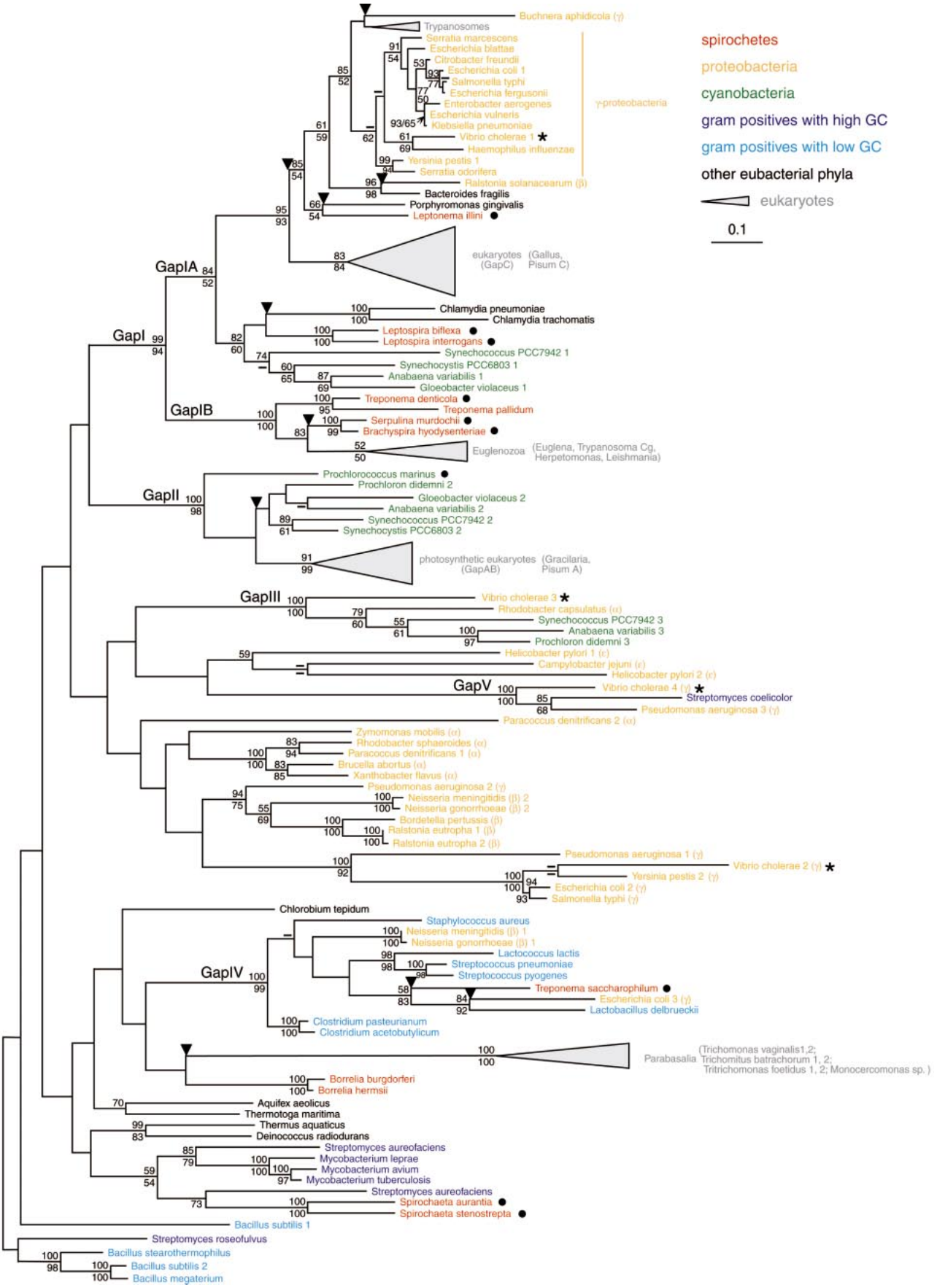
The distinctiveness of GAPDH subgroups is not only supported by high bootstrap values but also by specific sequence signatures (data not shown) and indels (fig. 2). Gap IA contains four specific signature sequences, but no indels. Within the Gap IB sequences, 11 signature sequences and the following five insertions are found: (1) the motif -GLL- (positions 22ABC), (2) a conserved aspartate (63H), (3) a nonconserved insertion (78C), (4) a conserved glycine (164A), and (5) the motif -LPG-E (301ABC-E). No specific indels exist within the Gap II and the Gap III subtree. However, Gap II has eight and Gap III has 16 common sequence signatures. Gap IV sequences possess nine signatures and have a deletion (position 266–269) and two insertions (-G- at 190B and -VG/D-G/D- at 301AB-E) in common. Within Gap V sequences, five insertions are found, but only one of these (-GI- at 87AB) is well conserved in all three species. Gap V sequences have a high number (47) of specific sequence signatures, which is probably in part caused by the low number of sequences belonging to this class.

Spirochetes Harbor at Least Five Highly Distinct GAPDH Genes

In this study we have focused on the diversity and distribution of spirochetal Gap sequences. Therefore, we determined nine novel partial *gap* sequences from the following spirochetes: *Sp. aurantia*, *Sp. stenostrepta*, *T. saccharophilum*, *S. murdochii*, *Leptospira biflexa*, *L. illini*, *Leptospira interrogans*, *B. hyodysenteriae*, and *T. denticola*. Sequencing of multiple PCR products indicated that each spirochete probably harbors only one *gap* gene. Southern analysis was used to confirm that each isolated gene is truly found on the genome of the corresponding organism (data not shown). Comparisons of 12 spirochetal Gap sequences, including three that were already present in the database, demonstrated that, unlike what we expected, Gap sequences from spiro-

→

FIG. 1.—Phylogenetic tree constructed by the ProtML (Adachi and Hasegawa 1996) method using the JTT-F model (Jones, Taylor, and Thornton 1992). Bootstrap values indicate the numbers of times that a given node was detected out of 100 NJ replicates (value above line) or 250 MP replicates (value below line, or to the right). Only bootstrap values above 50% are indicated. Dashes denote alternative topologies that are supported by bootstrap values higher than 50% in the NJ or MP analyses. The scale bar indicates 0.1 amino acid substitution per nonsynonymous site. Sequences established in this study are indicated by circles. Asterisks indicate the positions of the four Gap sequences present in *V. cholerae*. Triangles denote putative gene transfer events. Various eubacterial phyla are shown in different colors: with spirochetes (red); proteobacteria (yellow); cyanobacteria (green); gram positives with high GC-content (dark blue); and gram positives with low GC-content (light blue). Sequences from eukaryotes are shown as grey triangles that correspond in their sizes to the number of sequences currently known. Eukaryotic species included in the analysis are indicated in parentheses next to the grey triangles.


spirochetes
proteobacteria
cyanobacteria
gram positives with high GC
gram positives with low GC
other eubacterial phyla
eukaryotes
0.1
Buchnera aphidicola (γ)
Trypanosomes
Serratia marcescens
Escherichia blattae
Citrobacter freundii
Escherichia coli 1
Salmonella typhi
Escherichia fergusonii
Enterobacter aerogenes
Escherichia vulneris
Klebsiella pneumoniae
γ-proteobacteria
Vibrio cholerae 1 ★
Haemophilus influenzae
Yersinia pestis 1
Serratia odorifera
Ralstonia solanacearum (β)
Bacteroides fragilis
Porphyromonas gingivalis
Leptonema illini ●
eukaryotes (GapC) (Gallus, Pisum C)
GapIA
GapI
GapIB
Chlamydia pneumoniae
Chlamydia trachomatis
Leptospira biflexa ●
Leptospira interrogans ●
Synechococcus PCC7942 1
Synechocystis PCC6803 1
Anabaena variabilis 1
Gloeobacter violaceus 1
Treponema denticola ●
Treponema pallidum
Serpulina murdochii ●
Brachyspira hyodysenteriae ●
Euglenozoa (Euglena, Trypanosoma Cg, Herpetomonas, Leishmania)
GapII
Prochlorococcus marinus ●
Prochloron didemni 2
Gloeobacter violaceus 2
Anabaena variabilis 2
Synechococcus PCC7942 2
Synechocystis PCC6803 2
photosynthetic eukaryotes (GapAB) (Gracilaria, Pisum A)
GapIII
Vibrio cholerae 3 ★
Rhodobacter capsulatus (α)
Synechococcus PCC7942 3
Anabaena variabilis 3
Prochloron didemni 3
Helicobacter pylori 1 (ε)
Campylobacter jejuni (ε)
Helicobacter pylori 2 (ε)
GapV
Vibrio cholerae 4 (γ) ★
Streptomyces coelicolor
Pseudomonas aeruginosa 3 (γ)
Paracoccus denitrificans 2 (α)
Zymomonas mobilis (α)
Rhodobacter sphaeroides (α)
Paracoccus denitrificans 1 (α)
Brucella abortus (α)
Xanthobacter flavus (α)
Pseudomonas aeruginosa 2 (γ)
Neisseria meningitidis (β) 2
Neisseria gonorrhoeae (β) 2
Bordetella pertussis (β)
Ralstonia eutropha 1 (β)
Ralstonia eutropha 2 (β)
Pseudomonas aeruginosa 1 (γ)
Vibrio cholerae 2 (γ) ★
Yersinia pestis 2 (γ)
Escherichia coli 2 (γ)
Salmonella typhi (γ)
Chlorobium tepidum
GapIV
Staphylococcus aureus
Neisseria meningitidis (β) 1
Neisseria gonorrhoeae (β) 1
Lactococcus lactis
Streptococcus pneumoniae
Streptococcus pyogenes
Treponema saccharophilum ●
Escherichia coli 3 (γ)
Lactobacillus delbrueckii
Clostridium pasteurianum
Clostridium acetobutylicum
Parabasalia (Trichomonas vaginalis1,2; Trichomitus batrachorum 1, 2; Tritrichomonas foetidus 1, 2; Monocercomonas sp.)
Borrelia burgdorferi
Borrelia hermsii
Aquifex aeolicus
Thermotoga maritima
Thermus aquaticus
Deinococcus radiodurans
Streptomyces aureofaciens
Mycobacterium leprae
Mycobacterium avium
Mycobacterium tuberculosis
Streptomyces aureofaciens
Spirochaeta aurantia ●
Spirochaeta stenostrepta ●
Bacillus subtilis 1
Streptomyces roseofulvus
Bacillus stearothermophilus
Bacillus subtilis 2
Bacillus megaterium

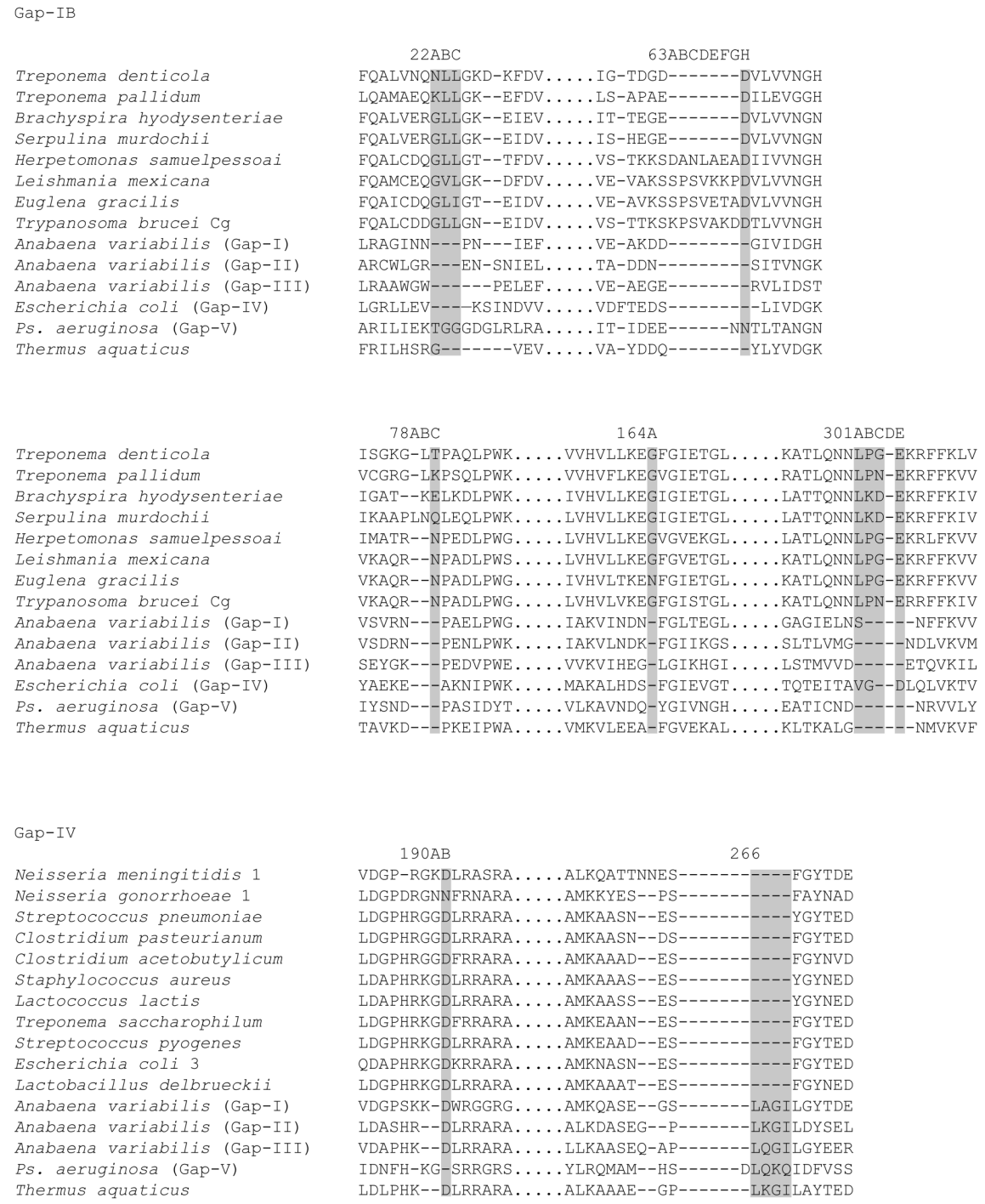


FIG. 2.—Parts of multiple-sequence alignments showing specific insertions and deletions (shaded in grey) found in the Gap types Gap IB, Gap IV, Gap V, and in the *Borrelia*/Parabasalia clade. The numbering system is in accordance with the one based on the three-dimensional structure of *Bacillus stearothermophilus* GAPDH (Clermont et al. 1993; Cerff 1995). Dots indicate parts of the complete alignment that are not shown; dashes represent gaps. Database accession numbers of duplicate species are: *Trichomitus batrachorum gap1* and *gap2*, AF022417 and AF022418; *Trichomonas vaginalis gap1* and *gap2*, L11394 and AF022414; *Tritrichomonas foetidus gap1* and *gap2*, AF022415 and AF022416.

chetes are highly divergent. An identities matrix (data not shown) revealed three surprising findings. (1) The lowest identity between two spirochetal Gap sequences (37.1%) was observed when sequences from two *Treponema* species, *T. pallidum* and *T. saccharophilum*, were compared. Even considering that *Treponema* does not form a very coherent genus (Paster et al. 1991), this finding is rather curious. (2) In stark contrast to this result is the observation that the Gap sequences from species of two different genera *Serpulina* and *Brachyspira* share 92.5% identical amino acids. However, it should be mentioned that these two organisms used to belong to the genus *Serpulina* and have just recently been reclassified in two different genera (Ochiai, Ada-

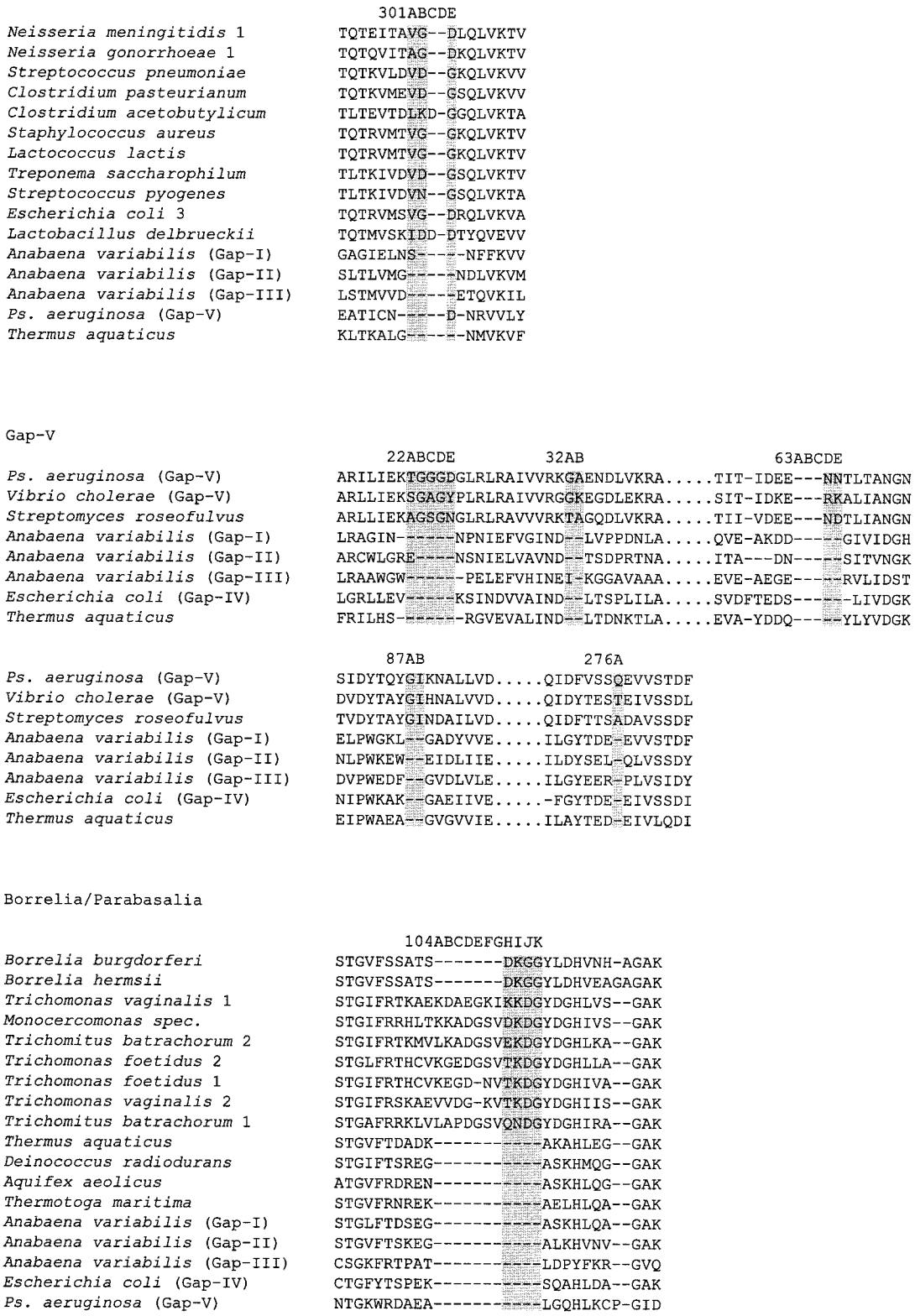


FIG. 2 *(Continued)*

chi, and Mori 1997). (3) Contrary to 16S rRNA analyses that demonstrate that *Leptonema* and *Leptospira* species belong to a distinct spirochete subgroup (Paster et al. 1991), identity values between *Leptospira* and *Leptonema* are comparatively low (63%). The phylogenetic analysis of 12 spirochete Gap sequences together with other eubacterial and eukaryotic GAPDH sequences (see earlier) confirmed the extreme heterogeneity of spirochetal *gap* genes. Indeed, spirochetal sequences were found at a total of five highly distinct positions in the

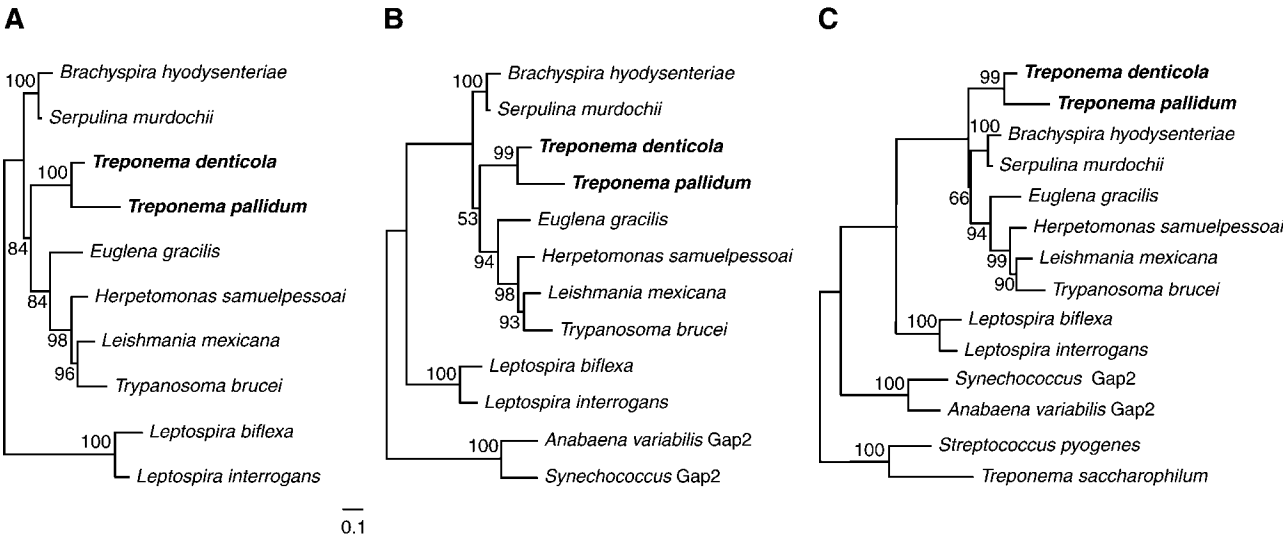


FIG. 3.—Outgroup species sampling changes the position of the *Treponema* species. In the presence of close outgroup sequences, the *Treponema* sequences cluster with the Euglenozoa with high bootstrap support (*A*, 84%). Successive addition of further, more distant outgroup sequences weakens this relationship (*B*, 53%) and, finally, separates the fast-evolving *Treponema* species completely from the Euglenozoa via an LBA artifact (*C*). Only quartet puzzling support values (Strimmer and von Haeseler 1996) above 50% are shown. The scale bar indicates 0.1 amino acid substitutions per nonsynonymous site.

GAPDH tree (fig. 1), with three of these positions located within the Gap I subtree: (1) *L. illini* Gap1 is clearly part of a cluster harboring sequences from gammaproteobacteria and two members of the phylum Bacteroides/Cytophaga/Flavobacterium. (2) The two *Leptospira* Gap1 sequences seem to be closely related to chlamydial and cyanobacterial Gap1 sequences. (3) Within the Gap IB subtree, Gap1 sequences from *T. pallidum*, *T. denticola*, *S. murdochii*, and *B. hyodysenteriae* cluster with Euglenozoan GapC sequences (see later). (4) In addition, a spirochetal Gap sequence is found in the Gap IV subtree (*T. saccharophilum*). (5) Two *Borrelia* Gap sequences form a common branch with GAPDH sequences from the eukaryotic Parabasalia. Furthermore, two *Spirochaeta* Gap sequences are loosely associated with sequences from gram positives with high GC-content.

Spirochetal Gap Sequences Are Closely Related to Eukaryotic Gap Sequences from Euglenozoa and Parabasalia

Recently, we showed that a Gap1 sequence from the syphilis-spirochete *T. pallidum* is very closely related to GAPDH genes from the eukaryotic Euglenozoa, indicating that an interkingdom gene transfer may have occurred between these organisms (Figge et al. 1999). To further our understanding of this curious finding, we have now analyzed this relationship in the presence of all available eubacterial and nine new spirochetal Gap sequences. Our analysis shows that, in addition to *T. pallidum* Gap1, three other spirochetal Gap sequences cluster with the Gap sequences from Euglenozoa. Sequences from *S. murdochii* and *B. hyodysenteriae* seem to be more closely related to Euglenozoan GapC than the two *Treponema* sequences. However, further analyses demonstrate that the outgroup species sampling influences the position of the *Treponema* species. In the presence of closely related outgroup sequences (Leptospira Gap IA), *Treponema* sequences cluster with the Euglenozoa with high bootstrap support (84%, fig. 3*A*). When two more outgroup sequences are added, bootstrap support for this relationship drops to 53%. Finally, the addition of further, more distant outgroup sequences separates the fast-evolving *Treponema* species from the Euglenozoa (fig. 3*C*, and even more prominently in fig. 1) and counterfeits a close relationship between GAPDH from *Brachyspira/Serpulina* and the Euglenozoa.

In addition to the close relationship between several spirochetal Gap1 and Euglenozoan sequences, our analysis reveals another spirochete-eukaryote connection. Figure 1*A* shows a loose association between GAPDH sequences of the two spirochetes belonging to the genus *Borrelia* and several species that are part of the protist division Parabasalia. NJ analyses found weak bootstrap support for the common branch (41%) that is also retained in the ProtML tree during local rearrangement (fig. 1). Using a smaller dataset (24 sequences), numerous alternate topologies were compared using the method of Kishino and Hasegawa (1989). However, none of the topologies was found to be significantly worse than the best ML tree in which the *Borrelia* and Parabasalia sequences branched together (data not shown). Nevertheless, a possible homologous insertion (position 104HIJK: NKD/GG, fig. 2) is consistent with the common origin of the GAPDH sequences in *Borrelia* and Parabasalia.

Discussion

Origin of the Spirochetal GAPDH Gene Diversity: Gene Duplications in the Eubacterial Ancestor or LGT

In this study, *gap* genes from nine spirochetes were amplified by PCR, cloned, and sequenced. Phylogenetic

analyses of the spirochetal Gap sequences, together with other available eubacterial and a representative set of eukaryotic Gap sequences, indicate that the eubacterial GAPDH gene diversity encompasses at least five highly distinct Gap types (Gap I to Gap V). Within the eubacterial GAPDH phylogeny, spirochetes show an exceptional degree of sequence diversity that may be explained by multiple LGT events, early gene duplications in the eubacterial ancestor, or both.

As spirochetes are part of three highly distinct gene clusters within the Gap I subtree, are found in the Gap IV subtree, and belong to two other groupings with lower bootstrap support, the phylogenetic distribution of spirochetes in the GAPDH tree can hardly be explained by organismal evolution. Therefore, it is likely that several spirochetes, which notably all seem to harbor a single *gap* gene, obtained their present *gap* gene via LGT from donors that belong to other eubacterial phyla. With the present data we cannot determine with certainty which of the spirochetes (if any) harbors the *gap* gene that was present in the spirochetal ancestor. However, the Gap IB subtree includes sequences from three different spirochete genera, and therefore it is possible that these organisms have retained the *gap* gene originally present in the spirochetal ancestor. As a consequence, all other spirochetes may have adopted *gap* genes from the following bacterial phyla: Bacteroides/Cytophaga/Flavobacterium (*L. illini*), chlamydia (two *Leptospira* species), gram positives with low GC or gamma-proteobacteria (*T. saccharophilum*), and gram positives with high GC (two *Spirochaeta* species). Thus LGT of *gap* genes seems to be especially frequent between spirochetes and other eubacterial phyla. Nevertheless, LGT probably also occurred between other eubacterial phyla, e.g., between proteobacteria (*E. coli*) and the gram positives (*Lactobacillus delbrueckii*) in the Gap IV subtree or in the Gap I subtree between proteobacteria (*Ralstonia solanacearum*) and Cytophaga/Bacteroides/Flexibacter (*Bacteroides fragilis*), as recently described by Figge et al. (1999). In conclusion, the data presented clearly underscore the importance of LGT in bacterial evolution (Ochman et al. 2000) and are in agreement with a recent analysis conducted by Rivera et al. (1998) who show that the genes encoding operational functions, such as GAPDH, are exchanged with much higher frequency than the so-called informational genes encoding the components of transcription, translation, and replication.

In contrast to the indications presented above, several other lines of evidence let us assume that part of the extant *gap* gene diversity is not the result of LGT, but was already present in the eubacterial ancestor. First of all, several eubacterial species harbor multiple, up to four, highly divergent *gap* genes as in the case of *Vibrio cholerae* (Heidelberg et al. 2000; indicated by asterisks in fig. 1). If these multiple *gap* genes have already been present in the common eubacterial ancestor, today each of these paralogous genes may be expected to form a distinct GAPDH subtree that in itself represents a eubacterial phylogeny. Indeed, the Gap I subtree has features resembling a eubacterial phylogeny (Woese 1987; Hugenholtz, Goebel, and Pace 1998) because it includes Gap sequences from five different phyla: proteobacteria (subgroups beta, gamma), cyanobacteria, chlamydiae, spirochetes, and Bacteroides/Cytophaga/Flavobacteria. Even though all other type trees (with the exception of Gap II) comprise species from only two or three eubacterial phyla, it is conceivable for two reasons that these type trees also represent rudimentary eubacterial phylogenies. First, several species may have reduced their original GAPDH gene diversity to one functional copy, a phenomenon termed reductive genome evolution, that is especially frequent within pathogens (Andersson and Kurland 1998). Second, only a minor fraction of the extant gene diversity is known, and thus the increase in sequence data may generate other type trees that resemble eubacterial phylogenies. Taken together, both LGT and gene duplications in the common eubacterial ancestor seem to account for the extremely broad GAPDH gene diversity presently observed. The fact that few gene phylogenies show a comparable diversity (Brown and Doolittle 1997) raises the question of why this particular gene has been so successful in bacterial evolution. The existence of different functions, such as erythrose-4-phosphate dehydrogenase activity that was found for the *E. coli* GapB enzyme (affiliated with the Gap III subtree; Zhao et al. 1995) or ADP-ribosylating activity on the outer membrane as determined for a GAPDH in group A streptococci (Pancholi and Fischetti 1993), may in part explain why so many divergent GAPDH genes have been retained in eubacteria. This functional pleiotropism of an ancient and highly conserved component of primary metabolism is surprising. It culminates in mammalian cells where GAPDH proteins are implicated in a number of fundamental processes not related to energy metabolism, such as tRNA export from the nucleus, DNA repair, and protein phosphorylation (for review see Sirover 1999).

Spirochetes, Donors of Eukaryotic GAPDH Genes

Recently, we showed that a Gap1 sequence from the syphilis-spirochete *T. pallidum* is very closely related to GapC sequences from Euglenozoa. The phylogenetic position of these GapC sequences made a mitochondrial and thus alpha-proteobacterial origin extremely unlikely, and therefore we invoked an additional gene transfer between *T. pallidum* and Euglenozoa (Figge et al. 1999). We have now examined this close relationship in the context of other spirochetal Gap sequences determined in this study. We show that in addition to the *T. pallidum* Gap1 sequence, several other spirochetal Gap1 sequences are closely related to Euglenozoan GapC. This raises the question of whether the gene transfer postulated did specifically occur between *T. pallidum* and the Euglenozoa, as originally assumed. In figure 1, GAPDH sequences from the spirochetes *B. hyodysenteriae* and *S. murdochii* seem to be more closely related to the Euglenozoan GAPDH sequences. However, figure 3*A* clearly shows that in the absence of distant outgroup sequences, the *Treponema* sequences cluster with the Euglenozoa with high bootstrap support. The increase

of distant outgroup sequences pulls the two *Treponema* sequences away from the Euglenozoa. Thus, in figure 1 long-branch attraction (LBA) (Felsenstein 1978) influences the position of the *Treponema* Gap sequences and makes it difficult to discern which spirochete lineage was actually involved in the transfer. The fact that the GAPDH genes of Euglenozoan nuclei are nested within eubacterial sequences makes a transfer from eukaryotes to prokaryotes very unlikely. There are two further arguments in favor of a transfer in the spirochetes-to-Euglenozoa direction. First, Gap IB is a class I GAPDH and clearly not related to class II GAPDH of Archaea, as would be expected if the gene were inherited from the archaeal/eukaryotic host cell lineage (Hensel et al. 1989; Cerff 1995). It also seems possible that the host cell which engulfed the mitochondrial ancestor did not have a glycolytic pathway at all, and hence no authentic GAPDH gene (Martin and Müller 1998; Liaud et al. 2000). Therefore, it would be difficult to explain the origin of the euglenozoan gene under the assumption that the transfer occurred in the eukaryote-to-prokaryote direction. Second, very few mechanisms are known that account for a transfer from eukaryotes to prokaryotes (Doolittle 1998), whereas many have been documented for gene transfer in the opposite direction (Martin 1999). Therefore, our phylogenetic data together with parsimony clearly favor a scenario in which a spirochete donated a gene to the Euglenozoa.

Finally, it should be noted that the GAPDH tree indicates another, though loose, relationship between spirochetes of the genus *Borrelia* and protists belonging to the division Parabasalia. In addition to the common branch that was observed in the NJ and ProtML analyses performed, these organisms share a specific insertion (see fig. 2). These findings indicate that, similar to the case described above, the GAPDH gene in Parabasalia may have been acquired from a spirochete belonging to the genus *Borrelia*.

In conclusion, several eukaryotic genes seem to be of spirochetal descent (this study; Doolittle et al. 1999b; Hannaert et al. 2000), indicating that spirochetes may have played an important role in the evolution of the eukaryotic cell.

Acknowledgments

S. Hansmann (TU-Braunschweig) and H. Brinkmann (Universität Konstanz) are acknowledged for help in the use of ProtML and TREE-PUZZLE and for helpful discussions. Further thanks are due to the following scientists for providing genomic DNA: Wolfgang Hess—Humboldt Universität Berlin (*P. marinus*), Isabelle St. Girons—Institut Pasteur Paris (*L. biflexa*, *B. hyodysenteriae*), and Klaus Heuner—Klinikum Charité Berlin (*T. denticola*). We thank TIGR for making sequence data available prior to publication and Maren Scharfe (GBF, Braunschweig) for help with the sequencing. We are very thankful to two unknown reviewers for many helpful suggestions. R.M.F. gratefully acknowledges the receipt of a Ph.D. stipend that was part of the DFG priority program, ''The molecular basis of plant evolution.''

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1996. MOLPHY. Version 2.3. Programs for molecular phylogenetics based on maximum likelihood. Pp. 1–150 *in* Computer science monographs, No. 28. Institute of Statistical Mathematics, Tokyo.

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

ASAI, T., D. ZAPOROJETS, C. SQUIRES, and C. L. SQUIRES. 1999. An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. Proc. Natl. Acad. Sci. USA **96**:1971–1976.

BRINKMANN, H., R. CERFF, M. SALOMON, and J. SOLL. 1989. Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach. Plant Mol. Biol. **13**:81–94.

BRINKMANN, H., and W. MARTIN. 1996. Higher-plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. Plant Mol. Biol. **30**:65–75.

BROCHIER, C., H. PHILIPPE, and D. MOREIRA. 2000. The evolutionary history of ribosomal protein RpS14. Trends Genet. **16**:529–533.

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

CERFF, R. 1982. Separation and purification of NAD- and NADP-linked glyceraldehyde-3-phosphate dehydrogenases from higher plants. Pp. 683–694 *in* M. EDELMAN, R. B. HALLICK, and N.-H. CHUA, eds. Methods in chloroplast molecular biology. Elsevier/North Holland, Amsterdam.

CERFF, R. 1995. The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GAPDH gene system. Pp. 205–227 *in* M. GÖ and P. SCHIMMEL, eds. Tracing biological evolution in protein and gene structures. Elsevier, Amsterdam.

CLERMONT, S., C. CORBIER, Y. MELY, D. GERARD, A. WONACOTT, and G. BRANLANT. 1993. Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase: role of the acidic residue in the fingerprint region of the nucleotide binding fold. Biochemistry **32**:10178–10184.

DOOLITTLE, R. F. 1998. The case for gene transfer between very distantly related organisms. Pp. 311–320 *in* K. SYVANEN and C. I. KATO, eds. Horizontal gene transfer. Chapman & Hall, London.

———. 1999a. Lateral genomics. Trends Cell. Biol. **9**:M5–M8.

———. 1999b. Phylogenetic classification and the universal tree. Science **284**:2124–2129.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**: 401–410.

FIGGE, R. M., M. SCHUBERT, H. BRINKMANN, and R. CERFF. 1999. Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra- and inter-kingdom gene transfer. Mol. Biol. Evol. **16**:429–440.

FORTHERGILL-GILMORE, L. A., and P. A. M. MICHELS. 1993. Evolution of glycolysis. Prog. Biophys. Mol. Biol. **59**:105–235.

HANNAERT, V., H. BRINKMANN, U. NOWITZKI, J. A. LEE, M. A. ALBERT, C. W. SENSEN, T. GAASTERLAND, M. MULLER, P. MICHELS, and W. MARTIN. 2000. Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoe-*

ba balamuthi, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. Mol. Biol. Evol. **17**:989–1000.

HEIDELBERG, J. F., J. A. EISEN, W. C. NELSON et al. (26 coauthors). 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406**:477–483.

HENSEL, R., P. ZWICKL, S. FABRY, J. LANG, and P. PALM. 1989. Sequence comparison of glyceraldehyde-3-phosphate dehydrogenases from the three urkingdoms: evolutionary implication. Can. J. Microbiol. **35**:81–85.

HENZE, K., A. BADR, M. WETTERN, R. CERFF, and W. MARTIN. 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. USA **92**:9122–9126.

HUGENHOLTZ, P., B. M. GOEBEL, and N. R. PACE. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180**:4765–4774.

JONES, D. T., W. R. TAYLOR, and J. M. THORNTON. 1992. The rapid generation of mutation data matrices from protein sequences. CABIOS **8**:275–282.

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.

LAKE, J. A., R. JAIN, and M. C. RIVERA. 1999. Mix and match in the tree of life. Science **283**:2027–2028.

LAWRENCE, J. G., and H. OCHMAN. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95**:9413–9417.

LIAUD, M. F., C. LICHTLE, K. APT, W. MARTIN, and R. CERFF. 2000. Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. Mol. Biol. Evol. **17**:213–223.

LIAUD, M. F., C. VALENTIN, W. MARTIN, F. Y. BOUGET, B. KLOAREG, and R. CERFF. 1994. The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases from *Chondrus crispus*. J. Mol. Evol. **38**:319–327.

MARKOS, A., A. MIRETSKY, and M. MÜLLER. 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote, *Trichomonas vaginalis*. J. Mol. Evol. **37**:631–643.

MARTIN, W. 1999. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. Bioessays **21**:99–104.

MARTIN, W., H. BRINKMANN, C. SAVONNA, and R. CERFF. 1993. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci USA **90**:8692–8696.

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.

MICHELS, P. A., M. MARCHAND, L. KOHL, S. ALLERT, R. K. WIERENGA, and F. R. OPPERDOES. 1991. The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distant evolutionary relationship. Eur. J. Biochem. **198**:421–428.

OCHIAI, S., Y. ADACHI, and K. MORI. 1997. Unification of the genera *Serpulina* and *Brachyspira*, and proposals of *B. hyodysenteriae* Comb. Nov., *Brachyspira innocens* Comb. Nov. and *Brachyspira pilosicoli* Comb. Nov. Microbiol. Immunol. **41**:445–452.

OCHMAN, H., and I. B. JONES. 2000. Evolutionary dynamics of full genome content in *Esherichia coli*. EMBO J. **19**: 6637–6643.

PANCHOLI, V., and V. A. FISCHETTI. 1993. Glyceraldehyde-3-phosphate dehydrogenase on the surface of group A streptococci is also an ADP-ribosylating enzyme. Proc. Natl. Acad. Sci. USA **90**:8154–8158.

PASTER, B. J., F. E. DEWHIRST, W. G. WEISBURG, L. A. TORDOFF, G. J. FRASER, R. B. HESPELL, T. B. STANTON, L. ZABLEN, L. MANDELCO, and C. R. WOESE. 1991. Phylogenetic analysis of the spirochetes. J. Bacteriol. **173**:6101–6109.

PRÜSS, B., H. E. MEYER, and A. W. HOLLDORF. 1993. Characterization of the glyceraldehyde 3-phosphate dehydrogenase from the extremely halophilic archaebacterium *Haloarcula vallismortis*. Arch. Microbiol. **160**:5–11.

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.

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

SAMBROOK, J., F. E. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SIROVER, M. A. 1999. New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. Biochim. Biophys. Acta **1432**:159–184.

SMITH, T. L. 1989. Disparate evolution of yeasts and filamentous fungi indicated by phylogenetic analysis of glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci. USA **86**:7063–7066.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

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.

VIEIRA, J., and J. MESSING. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.

VISCOGLIOSI, E., and M. MÜLLER. 1998. Phylogenetic relationships of the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase, from parabasalid flagellates. J. Mol. Evol. **47**:190–199.

WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.

ZHAO, G., A. J. PEASE, N. BHARANI, and M. E. WINKLER. 1995. Biochemical characterization of *gapB*-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and its possible role in pyridoxal 5′-phosphate biosynthesis. J. Bacteriol. **177**:2804–2812.

GEOFFREY MCFADDEN, reviewing editor

Accepted August 17, 2001