

Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin

PATRICK J. KEELING* AND W. FORD DOOLITTLE

Department of Biochemistry, Dalhousie University, Halifax, NS Canada, B3H 4H7

Communicated by Norman R. Pace, University of California, Berkeley, CA, December 10, 1996 (received for review July 8, 1996)

ABSTRACT We have cloned and sequenced genes for triosephosphate isomerase (TPI) from the gamma-proteobacterium *Francisella tularensis*, the green non-sulfur bacterium *Chloroflexus aurantiacus*, and the alpha-proteobacterium *Rhizobium etli* and used these in phylogenetic analysis with TPI sequences from other members of the Bacteria, Archaea, and Eukarya. These analyses show that eukaryotic TPI genes are most closely related to the homologue from the alpha-proteobacterium and most distantly related to archaeobacterial homologues. This relationship suggests that the TPI genes present in modern eukaryotic genomes were derived from an alpha-proteobacterial genome (possibly that of the protomitochondrial endosymbiont) after the divergence of Archaea and Eukarya. Among these eukaryotic genes are some from deeply branching, amitochondrial eukaryotes (namely *Giardia*), which further suggests that this event took place quite early in eukaryotic evolution.

For at least two decades, we have known that the genomes of most (if not all) eukaryotes are chimeric; their nuclei and DNA-containing organelles have different evolutionary histories (1). The bulk of the nuclear genome appears to share common ancestry with modern Archaea (archaeobacteria). Rooted phylogenetic trees of translation elongation factors and aminoacyl-tRNA synthetases show that Archaea is the sister group of Eukarya (2, 3). In support of this, the sequences of many other essential components of the transcription and translation apparatus also reveal a strong archaeobacterial-eukaryotic affinity (4–6). In many instances, transcription and translation which are found in both archaeobacteria and eukaryotes are altogether absent from eubacteria (for review see ref. 7). Mitochondria, on the other hand, are the degenerate descendants of once free-living eubacteria that entered into an endosymbiotic association with a (presumably nucleated) host cell early in the evolution of eukaryotes. The genes retained in the mitochondrial genome show that this eubacterium was what we would now call an alpha-proteobacterium, a relative of modern genera such as *Rhizobium*, *Agrobacterium*, and *Rickettsia* (8). Similarly, plastid genes derive from the genome of a photosynthetic endosymbiont whose nearest modern relatives are cyanobacteria (9).

This picture of eukaryotic genome chimerism is further complicated in two ways. First, some lineages thought to have diverged soon after the origin of eukaryotes (Diplomonads and perhaps Microsporidia) have, in fact, no mitochondria. These lineages, which Cavalier-Smith has called Archezoa (10), may have never acquired mitochondria, and would therefore represent the original condition of the host in that respect. Second, many genes determining proteins that function in mitochondria or plastids actually reside in the eukaryotic

nuclear genome. These genes most often resemble eubacterial homologues and are thought to have been transferred to the nucleus from the symbiont genome, in most cases soon after the endosymbiosis was established (11), although isolated instances of organelle to nucleus transfer occurring more recently in evolution can still be documented for both mitochondria and plastids (12–15).

In nearly all widely accepted instances of such transfer, the product of the transferred gene still functions in the organelle in which it originally resided. We are aware of only one clear case in which an organelle gene seems to have replaced a nuclear homologue and assumed its cytosolic function. This is in plants, where there are two nuclear DNA-encoded phosphoglycerate kinase genes, one specific for the cytosol and one targeted to the plastid. In land plants, the cytosol-specific gene is significantly more similar to the chloroplast-specific gene (and thus to eubacterial genes) than to other eukaryotic cytosol-specific genes. This was originally attributed to a high level of intergenic recombination (16), but the data are more consistent with the nuclear-encoded chloroplast-targeted gene having duplicated at some point after the divergence of land plants from chlorophyte algae, but before dicots and monocots diverged, and replaced its nuclear encoded cytosol-specific counterpart (17).

Here we present data suggesting that eukaryotic nuclear genes encoding triosephosphate isomerase (TPI) are, like genes for mitochondrial proteins, of alpha-proteobacterial origin. TPI is central to carbohydrate metabolism and functions exclusively in the cytosol (except in photosynthetic eukaryotes where a second, plastid-specific enzyme is also present). However, an analysis of a limited number of TPI sequences by Schmidt *et al.* (18) hinted that eukaryotic TPIs branch with proteobacteria (although these authors did not comment on this result). To define the source of eukaryotic TPI more precisely by increasing the phylogenetic breadth of data, we isolated and sequenced TPI genes from three diverse eubacteria; the early-diverging green non-sulfur bacterium *Chloroflexus aurantiacus*, the gamma-proteobacterium *Francisella tularensis*, and the alpha-proteobacterium *Rhizobium etli*. Phylogenetic analyses including these new sequences support an association between the eukaryotes and proteobacteria and place *R. etli* alone as the outgroup to all known eukaryotes. Of all prokaryotic TPI genes, the archaeobacterial homologue, from *Pyrococcus woesei*, branches most distantly from that of eukaryotes.

Since the TPI genes from even the most early diverging eukaryotes represented show this affinity to alpha-proteobacterial TPI, it appears that the event leading to this relationship took place quite early in eukaryotic evolution. If all alpha-proteobacterial genes in eukaryotes derive from the alpha-proteobacterial endosymbiont that gave rise to the mitochondrion, then TPI must have been transferred in the very

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA
0027-8424/97/941270-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Abbreviation: TPI, triosephosphate isomerase.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U73962–U73966).

*Present address: Plant Cell Research Centre, School of Botany, University of Melbourne, Parkville VIC 3052, Australia.

early stages of symbiosis, perhaps even before the symbiont and host were genetically codependent. The present relationship could alternatively be the result of an isolated gene transfer from an alpha-proteobacterium, possibly arising from endosymbiotic relationships apart from the lineage that actually led to the mitochondrion. The line between these alternatives may not be clear, but in either case the proteobacterial ancestry of TPI has ramifications for current theories about early eukaryote evolution (10, 19) and for arguments based on TPI which have been used in the "introns early vs. introns late" debate (20–24).

MATERIALS AND METHODS

Strains and Culture Conditions. *Escherichia coli* DH-5aF' was used for all molecular manipulations and was grown on Luria–Bertani agar or in Luria–Bertani broth under ampicillin selection. *R. etli* CFN42 cells and DNA were provided by E. Martínez-Romero (Centro de Investigación Sobre Fijación de Nitrogeno), *F. tularensis* LVS DNA was provided by F. Nano (University of Victoria), and *C. aurantiacus* J-10-f1 DNA was provided by J. Lopez and R. E. Blankenship (both at Arizona State University).

Amplification of TPI Genes and General Molecular Techniques. Genomic DNA (50–200 ng) from *R. etli*, *F. tularensis*, and *C. aurantiacus* was used in PCR amplifications consisting of 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100, 0.2 mg/ml BSA (pH 9, 25°C), 10 mM each dNTP, 2 units of *Taq* polymerase, 0.5 unit of *Pfu* polymerase, and 1 μM each primer. Two sets of primers (provided by A. J. Roger, Dalhousie University) were used that match highly conserved blocks of all known TPI genes. One set corresponded to amino acids 6–11 and 232–237 (TF1, ACGTCTCGAGTTCGGTG-GNAAAYTGAA, and TR1, ATCTCTAGAAGTGATGC-NCCNCCNAC) and an internal set corresponded to 72–77 and 170–175 (TF4, CGAGAATTCAACGGTGCATTYAC-NGGNGA, and TR2, AGCTCTAGACCTGTNCCDAT-NGCCC), and numbering was according to the *E. coli* sequence.

Products of the expected size were isolated from agarose gels by freezing crushed gel slices in an equal volume of TE (10 mM Tris-HCl/1 mM EDTA, pH 8.0) and two volumes of phenol at -70°C, followed by centrifugation and ethanol precipitation of the aqueous phase. Purified products were cloned using TA-cloning vector, pCRII (Invitrogen), and sequenced by the dideoxy chain termination method using T7 polymerase (Pharmacia). All sequences reported are based on at least two independent clones and were completed on both strands, except the type 2 and type 3 *R. etli* clones, which are based on single clones sequenced on both strands. In no case was any discrepancy between duplicate clones observed.

Phylogenetic Analysis. New sequences were aligned with 37 existing database entries for TPI genes and the unpublished sequences from *Entamoeba histolytica* (provided by A. J. Roger) using the PILEUP program from the GCG package (25) and edited manually (Fig. 1).

Phylogenetic trees based on this alignment were inferred using distance, parsimony, and maximum likelihood methods. The type 2 and type 3 *R. etli* sequences were not included in phylogenetic analyses due to their short size, although we note that, in trees where all three *Rhizobium* sequences were included, they branched together with high affinity and had little other effect on tree topology. Corrected distance measurements were calculated for 264 sites based on the PAM 250 distance matrix, and trees were constructed by neighbor-joining using the PROTDIST and NEIGHBOR programs from the PHYLIP version 3.57 package (26). Statistical support for each node was assessed by conducting 100 bootstrap replicates. Unweighted parsimony trees were found by conducting heuristic searches with 50 rounds of stepwise addition with tree



FIG. 1. Inferred amino acid alignment of TPI. Genes reported here (*R. etli* types 1, 2, and 3, *F. tularensis*, and *C. aurantiacus*) are aligned with representatives from eukaryotes (*G. lamblia* GS/M, *Homo sapiens*, and *Plasmodium falciparum*), eubacteria (*E. coli*, *Bacillus subtilis*, and *Thermotoga maritima*), and the archaeobacterium, *P. woesei*.

bisection and reconnection using PAUP version 3.1.1 (27). Bootstrap support was also calculated by conducting 100 random replicates. Protein maximum likelihood analyses was conducted on 213 positions using constraints as described in the results, and the JTT transition probability matrix in the PROTML program from the MOLPHY version 2.2 package (28).

Additional statistical tests were performed to see if the differences observed between the tree topologies based on TPI and other markers are significant or the result of poor resolution in the TPI tree. The eukaryotic branch in both parsimony and neighbor-joining trees was moved to alternative positions in the eubacteria. These alternative topologies were then compared based on Templeton's method (29) and evaluated statistically as described by Felsenstein (30) and Kishino and Hasegawa (31). For both starting topologies, two methods were used to evaluate the difference between alternatives and the SE: maximum likelihood and parsimony. Maximum likelihood tests were carried out on a data set of 213 positions again using PROTML. Tests based on parsimony were carried out by calculating the number of steps at each of the 264 positions and using these values to calculate the SE from the best tree according to Kishino and Hasegawa (31). The Templeton test built into user tree option of PROTPARS in the PHYLIP package were also used and the results resembled those shown.

RESULTS

Identification of TPI Genes from Eubacteria. Amplification products covering over 90% of the TPI gene were isolated from *C. aurantiacus*, *F. tularensis*, and *R. etli*. With the former two species, the sequence was isolated as a single 730-bp fragment using primers TF1 and TR1. In both cases, two clones were chosen and sequenced on both strands and found to be identical. With *R. etli*, this primer combination failed, so the

same portion of the gene was amplified in two overlapping pieces, using TF1 and TR2 for the 5' end, and TF4 and TR1 for the 3' end. Six clones of the 3' end were sequenced and found to be identical, but among the six clones isolated and sequenced from the 5' end, three distinct TPI coding sequences were identified. The three variants, named type 1, 2, and 3, share between 72.5 and 65.3% identity at the amino acid level. Divergence notwithstanding, these three amino acid sequences are extremely similar to one another compared with genes from other species, and they also contain shared insertions and deletions, strongly supporting the notion that they are paralogues that diverged relatively recently. Of the three, type 1 was identical over the 282-bp overlap region with the six 3' clones, so these have been treated as two fragments of a single gene.

Phylogeny of TPI Genes. Phylogenetic trees based on TPI sequences from 22 eukaryotes, 18 eubacteria, and 1 archaeobacterium were inferred using parsimony, distance, and maximum likelihood methods. Parsimony and neighbor-joining trees are shown in Fig. 2 A and B, respectively, each with bootstrap proportions for nodes supported >30%. Parsimony analysis yielded eight trees of identical length (2745 steps) that differed slightly in the exact position of several eukaryotes (namely *Gracilaria*, *Entamoeba*, the order within the fungi, and the presence of *Arabidopsis-Coptis* and *Culex-Drosophila* clades). The neighbor-joining and parsimony trees also differ from one another in a few other branches, but none that is significantly supported in either analysis and none that is central to the questions posed here. Two constant features of all the preferred trees is the association between eukaryotic and proteobacterial TPI genes (in particular that of the alpha-proteobacterium *R. etli*) and the relatively great distance between eukaryotic and archaeobacterial sequences. Protein

maximum likelihood analysis was conducted by constraining the topology of the eukaryotes to match that shown in Fig. 2A and dividing the prokaryotes into the groups in the same figure: *Rhizobium*, gamma-proteobacteria, mycoplasmas, other low GC Gram-positive bacteria, actinomycetes, *Thermotoga*, *Chloroflexus*, and *Synechocystis*, and *Pyrococcus*. The branching order of these nine groups was then exhaustively searched, but the best 100 trees were virtually indistinguishable from one another statistically (all but a few being within 1.98 SEs from the best tree). Once again, these trees were consistent in that the eukaryotic TPI genes were always quite distant from that of the archaeobacterium and nested within the proteobacteria (data not shown).

There is now a great deal of support both from molecular phylogeny and molecular biology that the archaeobacteria are more closely related to eukaryotes than are eubacteria (2-4, 6, 7). The expected phylogenetic relationship of a eukaryotic cytosolic enzyme is therefore exactly the opposite of that observed here: if TPI is derived from the "host" genome, then eukaryotic genes ought to branch closer to the archaeobacteria than they do to eubacteria. The highly divergent nature of the *Pyrococcus* TPI gene (32) raises concerns that this gene is a specialized paralogue that could conceivably be causing an erroneous phylogeny. However, excluding the *Pyrococcus* sequence from the alignment had little effect on the outcome of the trees (data not shown) and more importantly, the recently completed genome of *Methanococcus jannaschii* contains a single TPI gene that is very similar to that of *Pyrococcus*, suggesting further that the *Pyrococcus* TPI sequence is representative of the archaeobacterial lineage to which these two species belong.

Significance of the TPI Tree Topology. The relationship between eukaryotic and proteobacterial TPI genes was tested

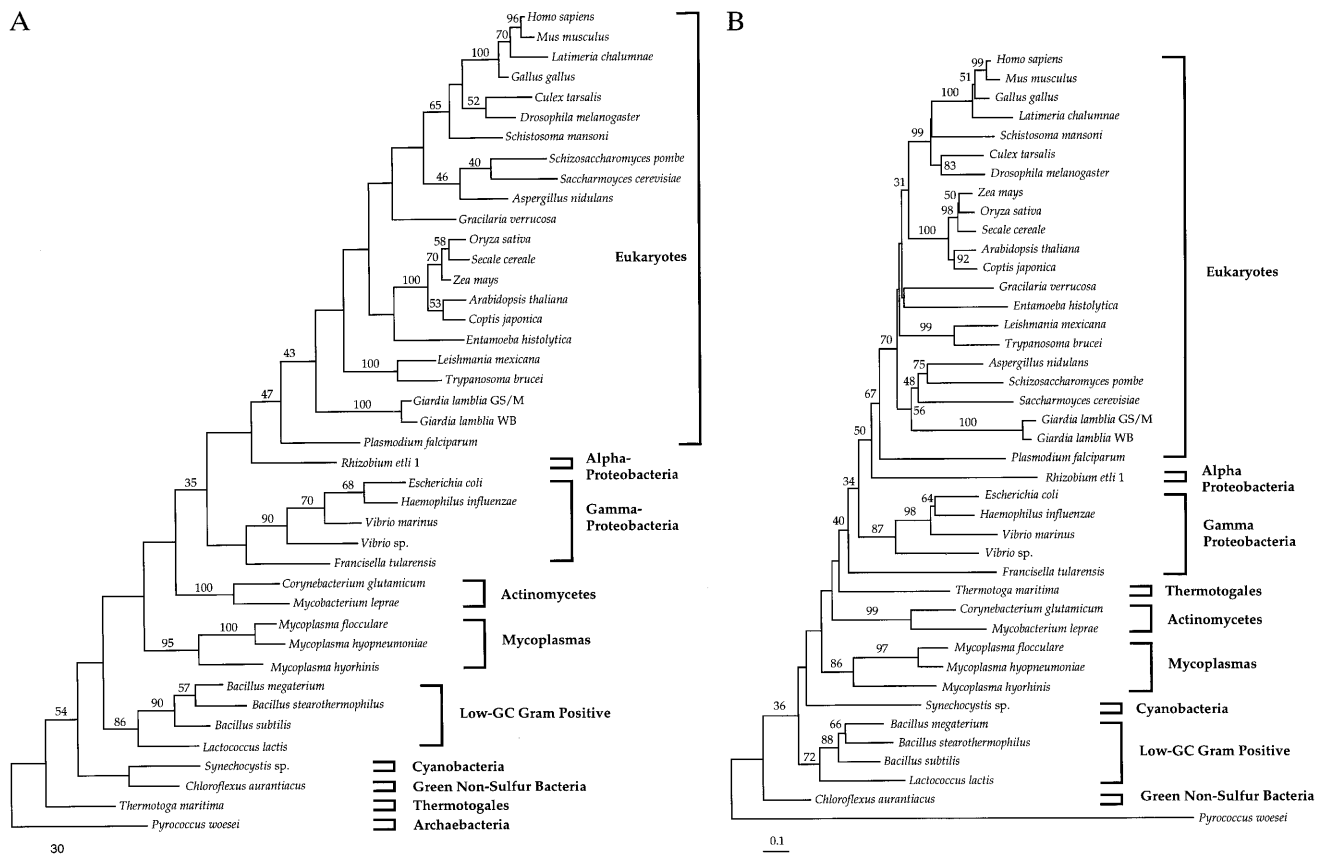


FIG. 2. Phylogenetic trees of TPI amino acid sequences from eukaryotes, eubacteria, and *P. woesei*. (A) Unweighted parsimony tree and (B) neighbor-joining distance tree showing bootstrap support for nodes where it is >30% are shown. Eukaryotes and major subdivisions of eubacteria are delineated to the right by brackets.

first by performing 100 bootstrap replicates on the distance and parsimony analyses. In both cases, the bootstrap support was almost universally low for all nodes except those between the most closely related taxa. TPI is a relatively small protein, and it is apparent that it has little power to resolve many details of organismal phylogeny (23). Nonetheless, the question posed here is a specific one—the significance of the position of the eukaryotic TPI genes within the eubacteria—and this was tested directly by comparing the “TPI tree” (using the topologies inferred by both parsimony and neighbor-joining) to alternative trees, according to the method described by Templeton (29). The alternative topologies were chosen by rooting the eukaryotes in each of the main intergroup branches within the prokaryotes. Fig. 3A shows the results of these tests, measured by parsimony and maximum likelihood, for the topology inferred by parsimony, and Fig. 3B shows the results for the neighbor-joining topology. In all tests the TPI tree is

superior to all alternatives and is significantly so in all cases except that where the outgroup of eukaryotes is all proteobacteria. Topologies G in Fig. 3A and H in Fig. 3B are of particular interest, as these trees are a very close approximation to the topology of universal trees inferred by other molecular markers (archaeobacteria as sisters to eukaryotes; refs. 2 and 3), yet these trees are between 2 and 4.4 SEs worse in both tests than the topologies actually inferred from TPI, suggesting that the phylogenetic history of TPI is different from these other molecules. Once again, these tests were repeated on the tree topologies derived by excluding *Pyrococcus*, and, as before, the archaeobacterial sequence was apparently not distorting the analysis: the ultimate result was the same and statistical significance changed very little (data not shown).

The possibility that either the branching order defined within the eukaryotes or the root of the eukaryotic subtree had some distorting effect on the likelihood of alternative topol-

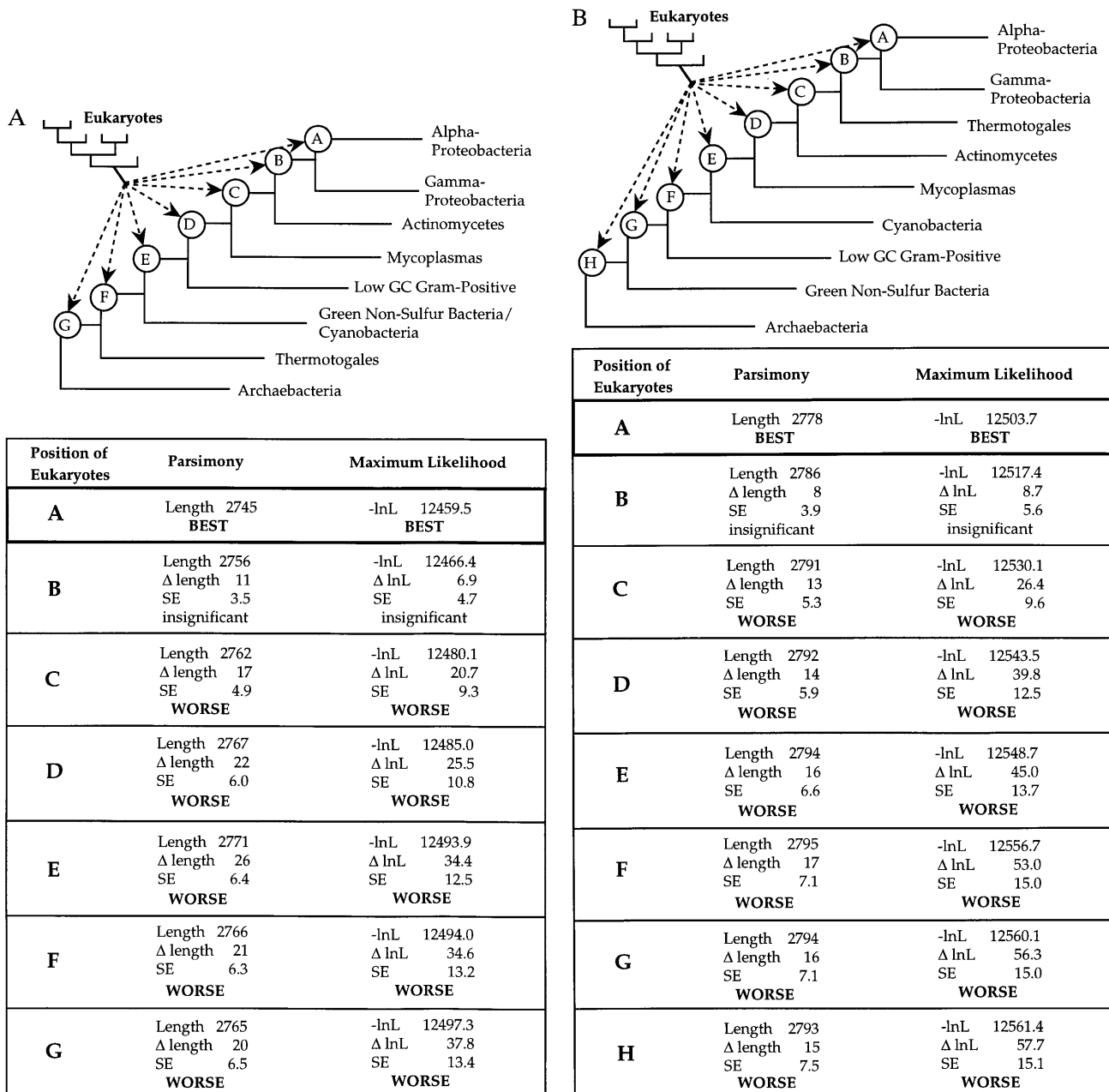


FIG. 3. Significance of TPI topology over alternatives. (A) Parsimony topology and (B) neighbor-joining topology alternatives tested by parsimony and maximum likelihood are shown. SE is calculated as described in text. Criteria for insignificant or significantly worse are >1.97 SE or 95% confidence.

ogies was also examined by conducting 100 independent maximum likelihood replicates with a random pair of eukaryotic sequences (so that there is only one topology or root within eukaryotes) and 16 eubacteria. In every one of these 100 replicates, the TPI topology was the best, although once again the difference between the *R. etli* specifically and all the proteobacteria was often insignificant.

DISCUSSION

A relationship between eukaryotic and proteobacterial TPI sequences has been supported by the addition of TPI genes from both supposedly early diverging (*Chloroflexus*) and later diverging (*Rhizobium* and *Francisella*) eubacterial lineages. Furthermore, the apparent relationship between *Rhizobium* and the eukaryotes inferred from TPI data suggests that the eukaryotic genes derive specifically from an alpha-proteobacterium.

Associations between eukaryotes and contemporary alpha-proteobacteria that might encourage some sort of lateral gene transfer are common and well known symbionts such as *Rhizobium*; intracellular parasites such as *Rickettsia* and *Ehrlichia*; and *Agrobacterium*, which has a sophisticated mechanism for transferring certain genes to the nucleus of its eukaryotic host (see ref. 33). However, since the symbiont that gave rise to the mitochondrion was also an alpha-proteobacterium, it will be difficult to distinguish independent lateral transfers from those associated with the alpha-proteobacterial endosymbiosis that ultimately produced mitochondria. This distinction is further complicated by the likelihood that the conversion of symbiont to organelle was a gradual process, with many fragments of the alpha-proteobacterial genome finding their way to the nucleus over a long period of time. Nevertheless, in general, it is imaginable that over this period some genes from symbiont genomes would become permanent additions to the nuclear gene complement, and some of these could serve cytosolic functions. Many genes of mitochondrial origin that serve mitochondrial functions are of course already known in the nucleus, and several cases of organelle proteins being replaced by cytosolic isoforms have also been documented (18, 34–36). Genes derived from organellar genomes assuming a cytosolic function are less well known, but one documented case involves replacement of a resident nuclear phosphoglycerate kinase gene by a gene derived from plastid DNA in land plants (17).

If eukaryotic TPI genes do share a common alpha-proteobacterial ancestry with genes from the mitochondrion, then we must rethink some aspects of early eukaryotic evolution. Trees in Fig. 2 include two amitochondrial eukaryotes, *Entamoeba histolytica* and *Giardia lamblia*. Both have TPI genes similar to other eukaryotes. Interestingly, although *Entamoeba* lacks cytologically defined mitochondria and is strictly anaerobic, genes whose products are normally targeted to the mitochondrion have been identified in its nuclear DNA (37). Since *Entamoeba* likely diverged from other eukaryotes after mitochondria arose (19, 38), and since the alpha-proteobacterial genes found in its nucleus (chaperonin 60 and pyridine nucleotide transhydrogenase) cluster with homologues specific to the mitochondrion, there is good support for the notion that *Entamoeba* once had a mitochondrion that it has subsequently lost. Similar suggestions for loss of mitochondria have been made for *Giardia*, but on much weaker evidence from glyceraldehyde-3-phosphate dehydrogenase (39) and immuno-crossreactivity with mitochondrial chaperonin 60 (40).

Giardia is a member of the group Cavalier-Smith called Archezoa, an assemblage of organisms that, unlike *Entamoeba*, is currently thought to have diverged before the acquisition of mitochondria (10). That *Giardia*, like other eukaryotes, possesses a TPI gene related to those of alpha-proteobacteria suggests that ancestors of *Giardia* had some close association

with alpha-proteobacteria, and raises the possibility that archezoa harbored similar endosymbionts, or perhaps even the ancestors of modern mitochondria.

The fact that TPI genes appear to have a nonnuclear, proteobacterial provenance bears on another question as well. TPI has been used extensively as a model for correlations between protein structure and gene structure in the debate over the origin of spliceosomal introns (20, 21). However, if TPI sequences presently encoded in eukaryotic genomes were actually derived from a proteobacterium, then these arguments must allow that the introns of eukaryotic TPI genes were also inherited from that proteobacterium. Since modern eubacteria have no known spliceosomal introns, this requires the independent loss of multiple introns from numerous eubacterial lineages.

We would like to thank E. Martínez-Romero, F. Nano, J. Lopez, and R. E. Blankenship for providing DNA used in this report and also all those who provided biological material who are not reported here. We would also like to thank A. J. Roger for primers and sharing unpublished data, M. Hasegawa for advice on MOLPHY, and J. R. Brown for conversations on phosphoglycerate kinase phylogeny. This work was supported by an operating Grant (MT4467) from the Medical Research Council of Canada (MRC). P.J.K. is the recipient of a MRC studentship and W.F.D. is a fellow of the Canadian Institute for Advanced Research.

- Gray, M. W. (1992) *Int. Rev. Cytol.* **141**, 233–357.
- Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355–9359.
- Brown, J. R. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2441–2445.
- Zillig, W., Palm, P., Klenk, H.-P., Langer, D., Hüdepohl, U., Hain, J., Lanzendörfer, M. & Holz, H. (1993) in *The Biochemistry Of Archaea*, eds. Kates, M., Kushner, D. J. & Matheson, A. T. (Elsevier, Amsterdam), pp. 367–391.
- Kletzin, A. (1992) *Nucleic Acids Res.* **20**, 5389–5396.
- Keeling, P. J., Baldauf, S. L., Doolittle, W. F., Zillig, W. & Klenk, H.-P. (1996) *Syst. Appl. Microbiol.* **19**, 312–321.
- Keeling, P. J. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5761–5764.
- Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4443–4447.
- Bonen, L. & Doolittle, W. F. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2310–2314.
- Cavalier-Smith, T. (1983) in *Endocytobiology*, eds. Schwemmler, W. & Schenk, H. E. A. (de Gruyter, Berlin), pp. 1027–1034.
- Weeden, N. F. (1981) *J. Mol. Evol.* **17**, 133–139.
- Baldauf, S. L. & Palmer, J. D. (1990) *Nature (London)* **344**, 262–265.
- Covello, P. S. & Gray, M. W. (1992) *EMBO J.* **11**, 3815–3820.
- Grohmann, L., Brennicke, A. & Schuster, W. (1992) *Nucleic Acids Res.* **20**, 5641–5646.
- Nugent, J. M. & Palmer, J. D. (1993) *Curr. Genet.* **23**, 148–153.
- Longstaff, M., Raines, C. A., McMorro, E. M., Bradbeer, J. W. & Dyer, T. A. (1989) *Nucleic Acids Res.* **17**, 6569–6580.
- Brinkmann, H. & Martin, W. (1996) *Plant Mol. Biol.* **30**, 65–75.
- Schmidt, M., Svendsen, I. & Feierabend, J. (1995) *Biochim. Biophys. Acta* **1261**, 257–264.
- Cavalier-Smith, T. (1993) *Microbiol. Rev.* **57**, 953–994.
- Gilbert, W., Marchionni, M. & McKnight, G. (1986) *Cell* **46**, 151–154.
- Tittiger, C., Whyard, S. & Walker, V. K. (1993) *Nature (London)* **361**, 470–472.
- Stoltzfus, A., Spencer, D. F., Zuker, M., Logsdon, J., Jr., & Doolittle, W. F. (1994) *Science* **265**, 202–207.
- Logsdon, J. M., Jr., Tyshenko, M. G., Dixon, C., D-Jafari, J., Walker, V. K. & Palmer, J. D. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 8507–8511.
- Kwiatowski, J., Krawczyk, M., Kornacki, M., Bailey, K. & Ayala, F. J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 8503–8506.
- Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Univ. of Washington, Seattle), version 3.57.

27. Swafford, D. L. (1993) PAUP, Phylogenetic Analysis Using Parsimony (Illinois Nat. Hist. Surv., Champaign, IL).
28. Adachi, J. & Hasegawa, M. (1992) MOLPHY, Programs for Molecular Phylogenetics: I. PROTML, Maximum Likelihood Inference of Protein Phylogeny (Inst. of Stat. Math., Tokyo), version 2.2, Computer Science Monographs No. 27.
29. Templeton, A. R. (1983) *Evolution (Lawrence, Kans.)* **37**, 221–244.
30. Felsenstein, J. (1985) *Syst. Zool.* **34**, 152–161.
31. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29**, 170–179.
32. Kohlhoff, M., Dahm, A. & Hensel, R. (1996) *FEBS Lett.* **383**, 245–250.
33. Zambryski, P. (1988) *Annu. Rev. Genet.* **22**, 1–30.
34. Bubunenko, M. G., Schmidt, J. & Subramanian, A. R. (1994) *J. Mol. Biol.* **240**, 28–41.
35. Brown, J. R., Masuchi, Y., Robb, F. T. & Doolittle, W. F. (1994) *J. Mol. Evol.* **38**, 566–576.
36. Henze, K., Schnarrenberger, C., Kellermann, J. & Martin, W. (1994) *Plant Mol. Biol.* **26**, 1961–1973.
37. Clark, C. G. & Roger, A. J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6518–6521.
38. Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. (1993) *Mol. Biochem. Parasitol.* **59**, 41–48.
39. Henze, K., Badr, A., Wettren, M., Cerff, R. & Martin, W. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9122–9126.
40. Soltys, B. J. & Gupta, R. S. (1994) *J. Parasitol.* **80**, 580–588.