Phylogenetic Analysis of β-Tubulin Sequences from Amitochondrial Protozoa

THOMAS D. EDLUND,* JING LI,* GOVINDA S. VISVESVARA,† MICHAEL H. VODKIN,‡ GERALD L. MCLAUGHLIN,§ AND SANTOSH K. KATYAR*†§

* Department of Microbiology and Immunology, Medical College of Pennsylvania and Hahnemann University, Philadelphia, Pennsylvania 19129; † Parasitic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia 30333; ‡ Illinois Natural History Survey, Champaign, Illinois 61820, and Department of Veterinary Pathobiology, University of Illinois, Urbana, Illinois 61801; and § Department of Pathobiology, Indiana University School of Medicine, Indianapolis, Indiana 46202

Received September 27, 1994; revised August 25, 1995

Molecular sequence data are being used increasingly to analyze phylogenetic relationships among protozoa. The most commonly used sequence is that of nucleus-encoded rRNA. Analysis of these sequences suggests that, relative to eubacteria and archaeabacteria, the eukaryotic domain is very deep (Sogin, 1989). Indeed, the number of kingdoms into which eukaryotes have been subdivided has expanded from the Plantae and Animalia to include Fungi, Protozoa, Chromista, and Archezoa, according to Cavalier-Smith (1993). The Archezoa represent the earliest branches in eukaryotic evolution, prior to the acquisition of mitochondria by endosymbiosis (Cavalier-Smith, 1987). The four phyla originally included in Archezoa, with representative pathogenic protozoa, were Archamoeba (Entamoeba histolytica), Metamonada (Giardia lamblia), Microsporidia (Encephalitozoon hellem), and Parabasalia (Trichomonas vaginalis). Subsequently, rRNA sequence analysis supported the early origins of these amitochondrial organisms, with the exception that E. histolytica diverged later, after several mitochondria-containing protozoa including Euglena gracilis (Sogin, 1989). Consistent with this, two genes encoding proteins normally localized in mitochondria have been identified in this organism (Clark and Roger, 1995). However, based on the amino acid sequence of translational elongation factor 1α (EF1α), E. histolytica appears to branch earlier than E. gracilis (Bauldauf and Palmer, 1993; Hasegawa et al., 1993). Various other aspects of its cell biology and metabolism argue for an early evolutionary origin for E. histolytica (Meza, 1992; Bakker-Grunwald and Wöstmann, 1993). Recently, Cavalier-Smith (1993) has revised his definition of Archezoa to include organisms primitively without both mitochondria and hydrogenosomes (which he suggests evolved from mitochondria), which excludes phylum Parabasalia. Furthermore, phylum Archamoeba has been more strictly defined, excluding Entamoeba species (Cavalier-Smith, 1991). These ambiguities regarding the phylogeny of amitochondrial protozoa could potentially be resolved by analyzing sequences unrelated to the translational machinery. Microtubules are a characteristic feature of...
of mitochondrial protozoa originally included in Archaea; G. lamblia (Kirk-Mason et al., 1988), T. vaginalis (Katiyar and Edlind, 1994), E. histolytica (Katiyar and Edlind, 1996), and E. hellem (Li et al., submitted for publication). By Southern blot analysis, the latter two appear to have single β-tubulin genes. In contrast, G. lamblia has three (Kirk-Mason et al., 1989) and T. vaginalis has six to seven gene copies. Representatives of all or most of the T. vaginalis gene copies were partially sequenced and shown to be >90% identical on the amino acid level (Katiyar and Edlind, 1994). Similarly, it was important to examine the remaining two G. lamblia gene copies to determine if they differed significantly from the reported copy. This was accomplished by PCR amplification of G. lamblia DNA with primers corresponding to conserved β-tubulin residues 99 to 107 and 260 to 267. Eleven independent clones of the PCR product were examined and three nearly identical sequences were detected. Each of the three sequences was identified in at least two independent PCR analyses; consequently, they are not artifacts and are presumed to represent the three β-tubulin gene copies detected by Southern analysis. The G. lamblia btub1 gene appears to encode the cDNA sequence reported by Kirk-Mason et al. (1988), except that nucleotides 367–368 are GA rather than AG as reported for the cDNA. Within the amplified region, the G. lamblia btub2 and btub3 genes have only single nucleotide differences from btub1 (not shown). On the amino acid level, btub2 β-tubulin, relative to the btub1 product (Fig. 1), has Arg in place of Lys154, and the btub3 product has Lys in place of Glu194. Thus, there appears to be minimal variation among the three G. lamblia β-tubulin gene copies.

Alignment of Tubulin Sequences

The G. lamblia btub1, T. vaginalis btub1, E. histolytica, and E. hellem β-tubulin sequences were aligned with previously reported β-tubulin sequences representing fungi, plants, animals, a chromist, and eight additional protozoan phyla, along with α- and γ-tubulin sequences from Drosophila melanogaster and Schizosaccharomyces pombe. A representative alignment is shown in Fig. 1. Up to residue 430, there were a total of only five single residue gaps or insertions in the alignment of 24 β-tubulin sequences; thus, alignment was unambiguous. Comparing the two α-tubulins to β-tubulin (e.g., from D. melanogaster), there were four gaps or insertions of 1 to 3 residues; alignment was considered ambiguous in only two regions (around β-tubulin residues 45 and 360; Fig. 1). Comparing the two γ-tubulins to β-tubulin, there were nine gaps or insertions of 1 to 3 residues, and consequently somewhat more ambiguity.

Phylogenetic Analysis of Complete β-Tubulin Sequences

The aligned tubulin sequences (residues 1 to 430) were phylogenetically analyzed by parsimony and dis-
PHYLOGENETIC ANALYSIS OF β-TUBULIN

**FIG. 1.** Representative alignment of β-tubulin sequences and an α- and a γ-tubulin sequence. β-Tubulins are from *D. melanogaster* (Dme; Michiels et al., 1987), *G. lamblia* (Gla; Kirk-Mason et al., 1988, and this study), *E. hellem* (Ehe; Li et al., submitted for publication), *T. vaginalis* (Tva; Katiyar and Edlind, 1994), and *E. histolytica* (Ehi; Katiyar and Edlind, 1996). Included are α-tubulin (DmeA; Theurkauf et al., 1986) and γ-tubulin (DmeG; Zheng et al., 1991) from *D. melanogaster*. Dots represent identity to the *D. melanogaster* β-tubulin sequence; dashes represent gaps introduced to maximize alignment.

Distance matrix methods (Felsenstein, 1993), and trees second major lineage includes β-tubulin sequences from animals (human, *D. melanogaster*, and *Caenorhabditis elegans*) and fungi (the ascomycetes or ascomycete-like *Aspergillus nidulans*, *Histoplasma capsulatum*, *Neurospora crassa*, *Pneumocystis carinii*, *Candida albicans*, and *S. pombe*; the basidiomycete *Schizophyllum commune* and the yeasts *Cryptococcus neoformans*). This animal-fungus lineage was reproduced in 70% of bootstrap resamplings.

As shown in Fig 2A, parsimony analysis generated a tree with two major lineages of β-tubulin sequences, preceded by four independently branching sequences. Five phyla from the kingdom Protozoa are represented in one major lineage: Kinetoplasta (*Trypanosoma brucei*), Percolozoa (*Naegleria gruberi*), Euglenoida (*E. gracilis*), Apicomplexa (*Toxoplasma gondii*), and Ciliophora, along with the acellular slime mold *Physarum polycephalum* (phylum Mycetozoa), branched before the animal kingdom. Organisms representing two additional kingdoms are also included: the oomycete *Achlya klebsiana*, classified in the kingdom Chromista (Cavalier-Smith, 1993), and the alga *Chlamydomonas reinhardtii* and higher plant *Zea mays* from the kingdom Plantae. This protozoa-plant lineage was reproduced in 77% of bootstrap resamplings.

The second major lineage includes β-tubulin sequences from animals (human, *D. melanogaster*, and *Caenorhabditis elegans*) and fungi (the ascomycetes or ascomycete-like *Aspergillus nidulans*, *Histoplasma capsulatum*, *Neurospora crassa*, *Pneumocystis carinii*, *Candida albicans*, and *S. pombe*; the basidiomycetes *Schizophyllum commune* and *Cryptococcus neoformans*). This animal-fungus lineage was reproduced in 70% of bootstrap resamplings.

**Fig. 1.** Representative alignment of β-tubulin sequences and an α- and a γ-tubulin sequence. β-Tubulins are from *D. melanogaster* (Dme; Michiels et al., 1987), *G. lamblia* (Gla; Kirk-Mason et al., 1988, and this study), *E. hellem* (Ehe; Li et al., submitted for publication), *T. vaginalis* (Tva; Katiyar and Edlind, 1994), and *E. histolytica* (Ehi; Katiyar and Edlind, 1996). Included are α-tubulin (DmeA; Theurkauf et al., 1986) and γ-tubulin (DmeG; Zheng et al., 1991) from *D. melanogaster*. Dots represent identity to the *D. melanogaster* β-tubulin sequence; dashes represent gaps introduced to maximize alignment.
FIG. 2. Phylogenetic trees of representative β-tubulin sequences (residues 1 to 430). Selected α- and γ-tubulin sequences were also included; D. melanogaster γ-tubulin was the designated outgroup. (A) One of two shortest trees based on parsimony analysis (PROTPARS; Felsenstein, 1993); the second shortest tree was identical except that the G. lamblia and P. polycephalum branches were switched. The consensus tree obtained following 100 bootstrap resamplings was identical to that shown; numbers indicate percentage of trees in which the group of species to the right of that branch was found (values less than 50% omitted). (B) Distance matrix-based tree with proportional branchlengths (PROTDIST and NEIGHBOR; Felsenstein, 1993). An identical consensus tree was obtained following 500 bootstrap resamplings; the bootstrap percentages are indicated. The sequence set for (B) was identical to that used in (A) except that C. albicans β-tubulin was omitted. Its addition generated a tree (lacking bootstrap support) in which the yeast (C. albicans and S. pombe) sequences branched after E. histolytica, followed by other fungi, animals, and a G. lamblia–T. vaginalis cluster. All sequences were obtained from GenBank.

While the parsimony-based tree shown in Fig. 2A, both consistently followed the T. vaginalis branch and preceded the protozoa–plant and animal–fungus division. A similar result was obtained when the P. polycephalum sequence was replaced by β-tubulin from the cellular slime mold Dictyostelium discoideum; when analyzed together, P. polycephalum and D. discoideum β-tubulins clustered (not shown). β-Tubulin from the fourth amitochondrial protozoan, the microsporidian E. hellem, unexpectedly branched within the animal–fungus lineage (Fig. 2A). Specifically, there was modest bootstrap support (68%) for its inclusion within the fungi. With most, but not all, sets of fungal sequences tested, E. hellem β-tubulin branched between ascomycete and basidiomycete β-tubulins.

While the parsimony-based tree shown in Fig. 2A is rooted with both α- and γ-tubulin outgroups, nearly identical trees, varying only in the branching of the G. lamblia and P. polycephalum sequences, were obtained using various outgroups tested individually (D. melanogaster, E. histolytica, or Plasmodium falciparum α-tubulins; D. melanogaster or P. falciparum γ-tubulins).

Using a distance matrix method with a nearly identical set of sequences to that shown in Fig. 2A, a very similar tree was generated (Fig. 2B; proportional branchlengths are shown). Specifically, β-tubulins from E. histolytica and T. vaginalis branched first. This was followed by a split leading to the fungus–animal lineage (again with E. hellem branching between the ascomycetes and basidiomycetes) and the protozoa–plant lineage (in this case, including P. polycephalum and G. lamblia at its base). Bootstrap support for the early branching of E. histolytica β-tubulin was 100%. Support for the other key features of this tree was more modest: 62% for the early branching of T. vaginalis β-tubulin and 57% for the basidiomycete–E. hellem–ascomycete cluster.

Among the β-tubulins, the E. histolytica sequence displayed the longest branchlength (Fig. 2B), reflecting its unusually low identity to other β-tubulins (54 to 58%: Katiyar and Edlind, 1996). The second longest
PHYLOGENETIC ANALYSIS OF β-TUBULIN

Partial Sequences of A. polyphaga β-Tubulins

Since the early branching of E. histolytica β-tubulin was unexpected (see Discussion), it was of interest to examine β-tubulin of another amoeba to determine if high divergence in β-tubulin was correlated with the reduced role of microtubules in these cells. For this, the free-living and aerobic A. polyphaga (phylum Rhizopoda) was selected. The PCR primers described above (corresponding to conserved β-tubulin residues 99 to 107 and 260 to 267) generated a DNA product of the expected size using a genomic DNA template from A. polyphaga. Southern blot analysis using the amplified DNA as a hybridization probe under conditions of moderate stringency suggested that A. polyphaga has four β-tubulin genes (not shown). DNA sequences of the PCR products have been deposited with GenBank; the results are summarized here. Four different sequences were identified among 10 clones examined, in agreement with the Southern blot results. Two, A. polyphaga btub1 and btub2, are closely related (98.9% identity), while btub4 has diverged considerably on the DNA level (81.6% identical to btub1). A polyphaga btub3 is very similar to btub1/btub2 except for a segment from approximately nucleotide 230 to 350, which is identical to btub4. Thus, btub3 is an apparent recombinant. On the amino acid level (not shown), the btub1 and btub2 products are identical in all but one position, while the btub4 product is highly divergent, differing from btub1 in 26 positions. Again, the btub3 product appears to be a recombinant, identical to btub1/btub2 products except for the 177 to 198 region, which has 4 residues in common with the btub4 β-tubulin.

Phylogenetic Analysis of Partial β-Tubulin Sequences

Although analyses of partial (152 residue) tubulin sequences are less likely to generate an accurate overall phylogeny, they may be useful for addressing specific questions; e.g., do all amoebal β-tubulins branch early, and do multiple β-tubulin gene copies in an organism compromise phylogenetic analysis? The A. polyphaga partial β-tubulin sequences encoded by btub1 and the divergent btub4, along with conserved and divergent β-tubulin sequences from A. nidulans (benA and tubC) and T. vaginalis (btub2 and btub3), were aligned with a representative set of sequences and phylogenetically analyzed as described above for complete β-tubulin sequences. As shown in Fig. 3 (parsimony analysis with α- and γ-tubulin outgroups), the results obtained are largely comparable to those shown in Figs. 2A and 2B. Specifically, the protozoan–plant and animal–fungus lineages were resolved, preceded by the branches leading to the T. vaginalis and E. histolytica β-tubulins. The locations of the P. polycephalum and G. lamblia branches, which were somewhat ambiguous in the analysis of complete β-tubulin sequences (see above), have shifted. P. polycephalum β-tubulin now branches within the protozoan–plant lineage, clustering with A. polyphaga β-tubulin (74% bootstrap support). The G. lamblia sequence branches at the base of the animal–fungus (Fig. 3) or protozoan–plant (not shown) lineages, depending on the sequence set used. Microsporidial β-tubulins from either E. histolytica or P. polycephalum (Fig. 3) or E. hellem (not shown) remains within the animal–fungus lineage.

This parsimony-based tree of partial β-tubulin sequences was reproduced using a distance matrix method (not shown), with one minor difference: the E. histolytica and S. commune branches were reversed. However, bootstrap support for these trees was somewhat limited, in comparison to the tree obtained with complete β-tubulin sequences. While early divergence of the E. histolytica and T. vaginalis sequences was reproduced in 85 and 71% of bootstrap resamplings, respectively, the protozoan–plant and animal–fungus lineages were variable.

In all three cases tested, the conserved and divergent copies of β-tubulin from individual species (A. polyphaga, A. nidulans, and T. vaginalis) branched together.

Comparison of α-Tubulin and β-Tubulin Phylogenies

Although the α-tubulin database is smaller and less representative, comparing α-tubulin phylogeny with that obtained for β-tubulin would provide another measure of reproducibility. A parsimony analysis of complete (residues 1 to 430), paired α- and β-tubulin sequences from 14 species, with a γ-tubulin outgroup, is shown in Fig. 4. The amitochondrial protozoa are represented by E. histolytica only (Sánchez et al., 1994). Clearly, the α- and β-tubulin trees are similar. For both, the E. histolytica sequence branches first. The fungal, animal, and protozoan–plant sequences form identical clusters, although the branching within the clusters is variable.

DISCUSSION

β-Tubulin as a Phylogenetic Marker

Several features of β-tubulin make it a potentially useful molecule for phylogenetic analysis of lower eukaryotes. In contrast to rRNA and many proteins, dele-
FIG. 3. Phylogenetic tree of representative partial $\beta$-tubulin sequences (residues 108 to 259). Selected $\alpha$-tubulin and $\gamma$-tubulin sequences were also included; S. pombe $\alpha$-tubulin was the designated outgroup. Parsimony analysis (PROTPARS; Felsenstein, 1993) was employed, and a consensus tree drawn following 100 bootstrap resamplings. Numbers indicate percentage of trees (>50%) in which the group of species to the right of that branch were found. A nearly identical tree with similar bootstrap values (500 resamplings) was obtained using a distance matrix method (PROTDIST and NEIGHBOR), with the exception that the $E. cuniculi$ and $S. commune$ branches were reversed. All sequences were obtained from GenBank.

segments or insertions which can make alignment ambigu-
ous are rare. Two distinct outgroups, $\alpha$- and $\gamma$-tubulin, which can be aligned with $\beta$-tubulin relatively unam-
biguously can be compared. While any of the three tubulins would probably be useful, the $\beta$-tubulin se-
quency database is currently the largest (over 100 full-
length sequences). Also, in contrast to rRNA and sev-
erval other proteins used for phylogenetic analysis, tubulins have been found (so far) only in eu-
akaryotes; thus, their phylogeny may more accurately parallel
that of other euakaryote-specific structures, such as the
nuclues. A potential problem in using $\beta$-tubulin for phy-
genetic analysis is presented by the multiple gene copies
found in some organisms, since individual $\beta$-tu-
bulins may have evolved specialized roles accompanied
by significant sequence divergence. In fact, in all three
cases analyzed here, involving $A. polyphaga$, $T. vagi-

nalis$, and $A. nidulans$ (Fig. 3), the conserved and diver-
gent $\beta$-tubulin copies branched together.

The results obtained here with $\beta$-tubulin sequences
are consistent in several major respects with results ob-
tained with rRNA or other protein sequences. The early
divergence of plants from a lineage leading to animals
and fungi was recently demonstrated with rRNA (Wain-
right et al., 1993) and EF1\alpha (Bauldauf and Palmer,
(1993) also present eukaryotic phylogenies based on $\beta-
tubulin sequences; however, the protozoan components
of these phylogenies cannot be directly compared to the
phylogeny presented here due to the selection of $T.$
brucei $\beta$-tubulin as outgroup in their studies.] Also, the
relatedness of Kinetoplasta to Euglenoida and of Cilio-
phora to Apicomplexa, as well as the unrelatedness of
true fungi to Mycetozoa and oomycetes, was previously
argued on both molecular and ultrastructural grounds (for review, see Cavalier-Smith, 1993), and is supported here by β-tubulin analysis.

However, the β-tubulin phylogeny is clearly different from the rRNA phylogeny with respect to the amitochondrial protozoa represented in this study. Following the initial proposal for the kingdom Archezoa (Cavalier-Smith, 1987), Archamoeba were redefined to exclude Entamoeba spp. from this kingdom, and Para-basalia (e.g., T. vaginalis) were removed based on the speculation that their hydrogenosomes evolved from mitochondria (Cavalier-Smith, 1991, 1993). These changes were meant to reflect the initial proposals for eukaryotic phylogeny based on small subunit rRNA, in which the relative branching order was G. lamblia, microsporidia, a trichomonad, and considerably higher up the tree, E. histolytica (Sogin, 1989). G. lamblia also branches before E. histolytica on an EF1α-based tree (Bauldauf and Palmer, 1993). The β-tubulin phylogeny, first of all, reverses the relative positions of G. lamblia and T. vaginalis (Fig. 2). A similar conclusion was reached when the rRNA-based analysis was expanded to include a second Metamonada species, Hexamita inflata (Leipe et al., 1993). More significantly, the β-tubulin-based phylogeny places microsporidia within the animal–fungus lineage. Finally, E. histolytica provides the earliest branch in the β-tubulin tree. (It should be noted, however, that E. histolytica β-tubulin is not unique, as a partial sequence from the reptilian parasite Entamoeba invadens reveals it to be similarly early branching; Katiyar and Edlind, 1996).

Implications for Microsporidia

A trivial explanation for the relatedness of microsporidial to fungal β-tubulins is that the presumed E. hellem sequence is actually that of a contaminant. Since the organisms were obtained from a clinical specimen and propagated on mammalian host cells, this is difficult to completely rule out; however, no fungal contaminants were detected by culture on microbiological
media. Moreover, related but distinct sequences were obtained from four different microsporidial species, three from AIDS patients (E. hellem, E. cuniculi, and Septata intestinalis) and one from an insect (Nosema locustae) (Li et al., submitted for publication). The latter was also microscopically >99% pure. This strongly argues against a fungal contaminant.

Certain anomalies in microsporidial rRNA are most readily explained by secondary deletion events (for review, see Cavalier-Smith, 1993). Such deletions would probably be accompanied by compensating changes in adjacent (and possibly distant) rRNA sequences. If so, then the rRNA-based phylogenies of these organisms become less reliable, and this could then explain the discrepancy between these and β-tubulin-based phylogenies. On a functional level, the absence of a specific rRNA structure associated with paromomycin sensitivity (Katiyar et al., 1995) is consistent with the β-tubulin but not the rRNA phylogenies. The relatedness of microsporida to fungi and animals, as suggested by the β-tubulin phylogeny, invites speculation that these obligate intracellular parasites evolved degeneratively from higher, free-living forms. It is noteworthy that, among the amitochondrial protozoa, only microsporidia display the potential for sexual recombination: formation of synaptonemal complexes associated with meiotic division (for review, see Sprague et al., 1992). Examination of additional protein-encoding sequences from the microsporidia will be of interest.

Implications for E. histolytica

It is more difficult to rationalize the dramatic difference in the placement of E. histolytica within the rRNA and tubulin (both α and β) phylogenies. The simplest explanation is that, in this case, tubulin represents the less reliable molecular clock. For example, it can be argued that tubulin sequences in this amoeba have diverged at a higher rate because they are not required for the formation of a cytoskeleton and axonemes. Highly divergent sequences may artifically branch early, as was noted with yeast (C. albicans and S. cerevisiae) β-tubulins (Fig. 2B). To examine this possibility, partial β-tubulin sequences from another amoeba, A. polyphaga, were analyzed (Fig. 3). In contrast to the E. histolytica sequence, A. polyphaga β-tubulins branched relatively high, within the protozoan–plant lineage and in a cluster with P. polypephalum β-tubulin. Thus, highly divergent tubulins are not associated with all amoeboid cells. Therefore, alternative explanations for the different rRNA versus tubulin phylogenies warrant consideration. For example, it has been speculated that microtubules, like mitochondria, were acquired from endosymbiotic bacteria (Margulis, 1981). It is possible then to envision that the ancestors of E. histolytica acquired their microtubules (and associated proteins) independently of other eukaryotes. At the very least, the different results obtained with rRNA versus tubulin support the notion that eukaryotic evolution is complex, and understanding it will require analysis of multiple molecules.

ACKNOWLEDGMENTS

Supported in part by NIH Grant AI 32433 to T. D. Edlind and Pittsburgh Supercomputing Center Grant DMBB90077P to M. H. Vodkin.

REFERENCES

Felsenstein, J. (1993). PHYLIP Version 3.5c, Distributed by the author, Department of Genetics, University of Washington, Seattle.
PHYLOGENETIC ANALYSIS OF β-TUBULIN


