J Mol Evol (1995) 41:841–849

JOURNAL OF MOLECULAR EVOLUTION

The *Giardia lamblia* Actin Gene and the Phylogeny of Eukaryotes

Guy Drouin,[1] Mário Moniz de Sá,[1] Michael Zuker[2]

[1] Department of Biology, University of Ottawa, 30 Marie Curie, Ottawa, ON, K1N 6N5, Canada
[2] Institute for Biomedical Computing, Box 8036, Washington University, 700 S Euclid Avenue, St. Louis, MO 63110, USA

Received: 18 February 1994 / Accepted: 17 August 1995

Abstract. The single-copy actin gene of *Giardia lamblia* lacks introns; it has an average of 58% amino acid identity with the actin of other species; and 49 of its amino acids can be aligned with the amino acids of a consensus sequence of heat shock protein 70. Analysis of the potential RNA secondary structure in the transcribed region of the *G. lamblia* actin gene and of the single-copy actin gene of nine other species did not reveal any conserved structures. The *G. lamblia* actin sequence was used to root the phylogenetic trees based on 65 actin protein sequences from 43 species. This tree is congruent with small-subunit rRNA trees in that it shows that oomycetes are not related to higher fungi; that kinetoplatid protozoans, green plants, fungi and animals are monophyletic groups; and that the animal and fungal lineages share a more recent common ancestor than either does with the plant lineage. In contrast to small-subunit rRNA trees, this tree shows that slime molds diverged after the plant lineage. The slower rate of evolution of actin genes of slime molds relative to those of plants, fungi, and animals species might be responsible for this incongruent branching.

Key words: Actin genes — *Giardia lamblia* — GC-content — Phylogeny — Slime molds — Relative rate test

Correspondence to: G. Drouin

Introduction

Ribosomal RNA gene sequences are the molecules most extensively used to infer species phylogenies. They have allowed testing of different hypotheses concerning the evolutionary relationships of taxonomic groups for which there were only a few morphological characters with which to uncover these relationships (Sogin 1991). On the other hand, molecular phylogenies based on protein coding genes have on occasion been found to be inconsistent with rRNA-based phylogenies. For example, rRNA phylogenies suggest that the absence of mitochondria in the protozoan species *Entamoeba histolytica* is likely due to a secondary loss, whereas phylogenies based of elongation factor 1α and RNA polymerase protein sequences suggest that this species diverged before the lineage leading to eukaryotes with mitochondria (Sogin 1991; Sogin et al. 1993; Cavalier-Smith 1993; Hasegawa and Hashimoto 1993; Hasegawa et al. 1993). Obviously, all phylogenies based on a given gene are subject not only to statistical error, but also to the particular alignment used, to the rate of evolution and GC-content of this gene in different taxonomic groups, as well as to other factors (Schlegel 1994). It is therefore of interest to corroborate molecular phylogenies using different gene sequences.

Actin genes are well suited for the purpose of corroborating rRNA-based phylogenies because of the slow evolutionary rate of the proteins they encode and the fact that they can be unambiguously aligned (Doolittle 1992). Furthermore, amino acid sequences are usually thought to be less affected by GC-content biases than rRNA genes and are therefore often considered a reliable indi-

cator of phylogenetic relationships when genes from distantly related species having widely different GC-content are being compared (Loomis and Smith 1990; Sidow and Wilson 1990; Martin et al. 1992). The fact that actin genes have been cloned and sequenced in more than 100 species also provides a wealth of sequence information for phylogenetic inference (Sheterline and Sparrow 1994).

Considerable interest was generated when 18S rRNA sequences were used to show that the absence of mitochondria in Archezoan species was likely a relic of the earliest phase of eukaryote cell evolution, i.e., before the endosymbiotic event (or events) that gave rise to mitochondria, and not due to the secondary loss of this organelle (Vossbrinck et al. 1987). *Giardia lamblia* is a relatively well known Archezoan species and has been shown to represent one of these early lineage in the eukaryotic line of descent (Sogin 1989; Sogin et al. 1989; Kabnick and Peattie 1991; Cavalier-Smith 1993; Hashimoto et al. 1994). We therefore used its actin gene sequence as an outgroup to root our phylogenetic trees based on 65 actin sequences from 43 species.

Materials and Methods

Isolation, Cloning, Sequencing, and Copy Number. The lambda EMBL3 genomic DNA library of *Giardia lamblia* was a generous gift from Debra A. Peattie. It was screened using the GC-rich ardC actin cDNA clone of *Physarum polycephalum,* a generous gift from Dominick Pallotta (Hamelin et al. 1988). The 2.5-kb *Bam*HI fragments of two independent clones and the 1.5-kb *Pst*I fragments of two other clones were subcloned into pBluescript SK+ plasmids. All clones were found to contain likely allelic variants of the same gene, and only one 1.5-kb *Pst*I fragment was selected for further study. Both strands of the selected clone were sequenced repeatedly via the dideoxynucleotide chain-termination method (Sanger et al. 1977) using Sequenase (USB), synthesized internal primers, and dGTP, 7-deaza-dGTP, and dITP nucleotide mixtures.

This 1.5-kb *Pst*I fragment was gel purified, labeled with ^{32}P-dCTP using an oligonucleotide labeling kit (Pharmacia), and hybridized to a Southern blot of *G. lamblia* genomic DNA (a generous gift from Debra A. Peattie) digested with *Bam*HI and *Pst*I in a solution of 50% formamide, 5× SSC, 5× Denhardt's, 0.5% SDS, and 100 mg/ml salmon sperm DNA at 42°C. The membrane was washed twice in 2× SSC, 0.1% SDS at room temperature for 5 min, once in 2× SSC, 0.1% SDS at room temperature for 10 min, and twice in 2× SSC, 0.1% SDS at 42°C for 15 min.

DNA Sequences and Alignment. The actin gene sequences used for folding analysis are listed in Table 1. They were selected because they had been reported to exist as single-copy genes in their respective species. All these genes code for actin proteins of 375 or 376 amino acids in length, except for the *Volvox carteri* protein, which is 377 amino acids long. An alignment of heat shock-70 proteins was performed using the PILEUP program of the GCG package. The sequences used are listed in Table 2.

The accession numbers of the actin sequences used for phylogenetic inference, as well as their alignments, are available by anonymous FTP at bio01.bio.uottawa.ca in the /pub/giardia directory.

Phylogenetic Analyses. Phylogenetic trees were constructed using programs running on a SUN SPARC station. The programs PROTDIST and NEIGHBOR from the PHYLIP 3.52 package were used to infer phylogenies from protein sequences based on the Dayhoff PAM model of amino acid substitutions (Felsenstein 1989). The programs DNADIST and NEIGHBOR were used to infer phylogenies from DNA coding sequences based on the Kimura two-parameter model of nucleotide substitutions. Bootstrap analyses were performed using the SEQBOOT and CONSENSE programs from the same package. The tree was drawn using the DRAWGRAM program of the PHYLIP package and the public domain XFIG program.

Table 1. Actin gene sequences

GenBank accession number	Organism	Alignment[a]
M22869	*Aspergillus nidulans*	2
X16377	*Candida albicans*	1
L29032	*Giardia lamblia*	2
M25826	*Kluyveromyces lactis*	2
X15900	*Phytophthora megasperma*	2
Y00447	*Schizosaccharomyces pombe*	2
M13939	*Tetrahymena thermophila*	1
X07463	*Thermomyces lanuginosus*	2
M33963	*Volvox carteri*	0
L00026	*Saccharomyces cerevisiae*	2

[a] Number of amino acid gap(s) inserted after the fourth amino acid of the different sequences to align them

Table 2. Heat shock 70 protein sequences

SwissProt name	Organism
Hs70_Chick	*Gallus gallus*
Hs70_Human	*Homo sapiens*
Hs70_Plafa	*Plasmodium falciparum*
Hs70_Schma	*Schistosoma mansoni*
Hs70_Xenia	*Xenopus laevis*
Hs74_Trybb	*Trypanosoma brucei*

RNA Secondary Structures. To search for "well-determined" secondary structures in the coding region of the *G. lamblia* actin mRNA, we employed both thermodynamic methods and comparative sequence analysis. For the latter, we assembled the coding regions of the nine other single-copy actin genes listed in Table 1. These sequences are very similar in length and are easy to align. This makes the search for common base-paired regions particularly simple and unambiguous. To generate foldings for individual sequences with the thermodynamic approach, we used the "Irna" program from the "mfold" package of Zuker and colleagues (Jaeger et al. 1989, 1990; Zuker 1989, 1994).

Results

The Actin of Giardia lamblia *Is Encoded by a Single Gene*

The Southern blot analysis of restriction digests of *G. lamblia* genomic DNA shows that its actin is encoded by a single gene (Fig. 1). Even though the membrane was washed under relatively low stringency, which would have allowed sequences related to our probe to cross-hybridize, a single hybridizing fragment is seen in the lane of both restriction digests. Surprisingly, the sizes of

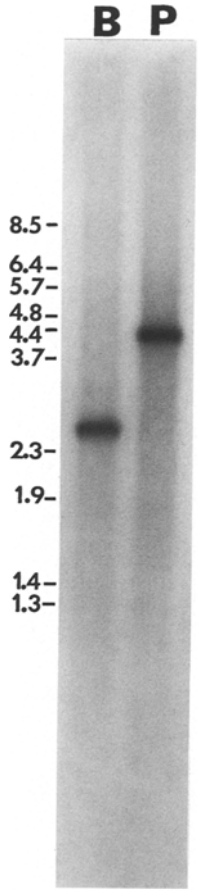

Fig. 1. Southern blot of *Giardia lamblia* genomic DNA digested with *Bam*HI (*B*) and PstI (*P*) and probed with the sequence shown in Fig. 2. The positions of the molecular weight standards (in kb) are indicated to the left of the autoradiogram.

the *Bam*HI and *Pst*I restriction fragments observed by Southern blot analysis, 2.6 kb and 4.3 kb, respectively, do not correspond to the size of the restriction fragments isolated from the lambda EMBL3 genomic DNA library (see above). This discrepancy is likely due to the fact that the genomic DNA used for the lambda EMBL3 genomic library and the genomic DNA use for Southern blot analysis came from different isolates. Recent studies have shown that different isolates of *G. lamblia* exhibit extensive karyotypic heterogeneity. For example, after only 8 division cycles, 20% of the clones derived from a single trophozoite can be shown, using pulsed field gel electrophoresis, to have mutant karyotypes (Le Blancq 1994).

Sequence of the Giardia lamblia *Actin Gene*

The *G. lamblia* actin gene sequence is shown in Fig. 2. This gene has no introns and codes for a protein of 375 amino acids. When compared to the amino acid sequence of the actin genes listed in Table 1, the *G. lamblia* actin has an average of 58.3% similarity, being most similar to *V. carteri* (59.7%) and least similar to the *Aspergillus nidulans* (56.8%) sequences. The amino acid sequence of the *G. lamblia* actin gene was also compared to both a consensus actin sequence derived from the nine sequences listed in Table 1 and a consensus heat shock protein 70 (hsp70) sequence derived from the six sequences listed in Table 2 (Fig. 3). Of the 377 amino acids of the actin consensus, 290 are conserved according to our above criteria. Of these, 204 are also found in the *G. lamblia* sequence (represented in Fig. 3 by * above of the actin consensus sequence), whereas 259 are found in the *Tetrahymena thermophila* sequence (Fig. 3, and results not shown). Comparison of the *G. lamblia* actin sequence with the aligned hsp70 consensus shows that 49 amino acids are conserved between these two sequences (represented by * below the hsp70 consensus sequence, Fig. 3).

Search for Common RNA Secondary Structures in the Coding Regions of Single-Copy Actin Genes

We folded the *G. lamblia* and *Saccharomyces cerevisiae* sequences to produce both energy dot plots and lists of optimal and suboptimal structures. The *S. cerevisiae* sequence was chosen arbitrarily for the purpose of comparison. The *G. lamblia* dot plot reveals an enormous number of base pairing possibilities in foldings within 10 kcal of the minimum energy of −274.6 kcal/mol. No clearly defined structural domains or "well-determined" features can be identified in the plot (results not shown). The automatic trackback feature found 19 different foldings within 10 kcal from the energy minimum, despite the fact that the distance parameter (Zuker et al. 1991) was set to 20, which is fairly large. No repeated structural elements were found in these foldings. The *S. cerevisiae* dot plot was less cluttered with potential base pairs, and the minimum energy folding had an energy of only −217.8 kcal/mol (results not shown). Although two short-range well-determined hairpin regions were detected (a helix of length 10 starting with the G:293-C:315 base pair, and a series of helices in the region from nucleotides 760 to 810) these features are not found in the *G. lamblia* folding. We computed 13 different foldings using the automatic feature of "lrna."

The failure of the thermodynamic approach to find either well-determined or conserved structural features in the *G. lamblia* and *S. cerevisiae* sequences led us to try the phylogenetic approach (Woese et al. 1983; Winker et al. 1990). Using the unambiguous alignment detailed above, and the "compbc" program of Chan et al. (1991), we looked for common helices with at least two compensating base pair changes per helix. We found no helices of size 4 or greater that were conserved in all species. Relaxing the conditions to 1 compensating base pair change yielded 15 conserved helical regions, but many of these overlap, and we feel that 1 compensating base change is not sufficient to reliably determine a helix. The results for conserved helices with at least two compen-

```
ctgcagcaatgtgtgctttatattatctaattctgtacgatagtacatgtgctatattta   60

aagatcgcagaaatgtgcactgagaactccttgccgatgagcctgtgtagaagcgtgcat  120

cttgaagtacccgttagtatcctataagtttatggctcttaggcatggtttctggaattt  180

                                                MetThrAspAspAsnProAlaI
gattttctcaatctttttaaatttatttgcaaagtaaaatgacagacgacaaccctgcca  240

leValIleAspAsnGlySerGlyMetCysLysAlaGlyPheAlaGlyAspAspAlaProA
tagtcattgataacggctccggaatgtgcaaggcaggctttgctggggacgatgcgccac  300

rgAlaValPheProThrValValGlyArgProLysArgGluThrValLeuValGlySerT
gtgctgtgttccccacagtagtcggtcgcccgaaacgtgagactgtgcttgtgggctcca  360

hrHisLysGluGluTyrIleGlyAspGluAlaLeuAlaLysArgGlyValLeuLysLeuS
cccacaaggaggagtacataggagacgaggccctagcaaagcgtggggtgctgaagctct  420

erTyrProIleGluHisGlyGlnIleLysAspTrpAspMetMetGluLysValTrpHisH
cctacccgatagagcatggacagataaaggactgggacatgatggagaaggtgtggcacc  480

isCysTyrPheAsnGluLeuArgAlaGlnProSerAspHisAlaValLeuLeuThrGluA
attgctacttcaatgagctgcgagcacagccttcggaccatgccgtactccttaccgagg  540

laProLysAsnProLysAlaAsnArgGluLysIleCysGlnIleMetPheGluThrPheA
ctccaaagaacccaaaggctaaccgagaaaagatttgccagatcatgttcgagacatttg  600

laValProAlaPheTyrValGlnValGlnAlaValLeuAlaLeuTyrSerSerGlyArgT
ccgtacctgccttctatgtccaggtacaggcagtcctggctctctacagttccgggcgca  660

hrThrGlyIleValIleAspThrGlyAspGlyValThrHisThrValProValTyrGluG
ccaccgggattgttatcgatacaggcgacggggtgacgcatactgttcctgtgtacgaag  720

lyTyrSerLeuProHisAlaValLeuArgSerGluIleAlaGlyLysGluLeuThrAspP
gatattctctaccacatgcagttctccgttctgagattgcgggcaaggagcttacagact  780

heCysGlnIleAsnLeuGlnGluAsnGlyAlaSerPheThrThrSerAlaGluPheGluI
tctgccaaatcaacctacaggagaacggcgcatcgtttaccacgagcgcagagtttgaga  840

leValArgAspIleLysGluLysLeuCysPheValAlaLeuAspTyrGluSerValLeuA
tagtgcgtgatataaaagagaagctctgcttcgttgcccttgattatgagtccgttttgg  900

laAlaSerMetGluSerAlaAsnTyrThrLysThrTyrGluLeuProAspGlyValValI
ctgcctccatggagagtgccaactacacaaaaacatatgagctgccagatggggtcgtca  960

leThrValAsnGlnAlaArgPheLysThrProGluLeuLeuPheArgProGluLeuAsnA
ttaccgtgaaccaggcccgcttcaagacccctgagctcctcttcaggccggagctcaaca 1020

snSerAspMetAspGlyIleHisGlnLeuCysTyrLysThrIleGlnLysCysAspIleA
atagcgacatggacggcattcaccagctatgttacaagaccattcagaagtgtgacattg 1080

spIleArgSerGluLeuTyrSerAsnValValLeuSerGlyGlySerSerMetPheAlaG
atattaggtcagagctttatagtaacgtggtcctgagcggcgggagcagcatgtttgcag 1140

lyLeuProGluArgLeuGluLysGluLeuLeuAspLeuIleProAlaGlyLysArgValA
gcctccccgagagactggagaaggagcttcttgacctcattccagccgggaagcgtgtgc 1200

rgIleSerSerProGluAspArgLysTyrSerAlaTrpValGlyGlySerValLeuGlyS
gcataagtagtcccgaggacagaaagtactctgcctgggttggtggaagcgtattgggga 1260

erLeuAlaThrPheGluSerMetTrpValSerSerGlnGluTyrGlnGluAsnGlyAlaS
gcctggcaaccttcgagtctatgtgggtaagctctcaagaatatcaagagaatggcgctt 1320

erIleAlaAsnArgLysCysMetEnd
ccattgcaaaccgtaagtgtatgtgaacacctatcaacgcttttctttccctttgcattg 1380

ttgtgctgtccacaatcttcctagacgacctacaagtggggtctcgcctatagatttgta 1440

ataaaattgtttaataaaaaatcaagtcctgtctatgcttacctccagatccgttgacct 1500

gcag  1504
```

Fig. 2. Sequence of the *G. lamblia* actin gene and its putative amino acid sequence. Potential promoter elements are *underlined.* This sequence has been deposited in GenBank (accession number L29032).

sating base pair changes are given in Table 3. Many of these helices conflicted with one another, and none of the ones we examined could be seen in the energy dot plots, indicating that they are not energetically favorable.

Effect of the GC-Content of the Coding Regions on the GC-Content of Codon Positions

In order for nucleotide sequences to be phylogenetically informative, they have to evolve independently from diverse constraints. We calculated the correlation coefficient between the percent GC in the first, second, and third bases of codons against the GC-content of the coding regions of the 65 genes shown in Fig. 4 in order to ascertain whether the nucleotides found at these different codon positions were independent of the overall GC-content of these genes. The second bases of codons are not influenced by the overall GC-content of their coding region ($R^2 = 0.07$). In contrast, the base composition of the first base of codons is affected by the GC-content of

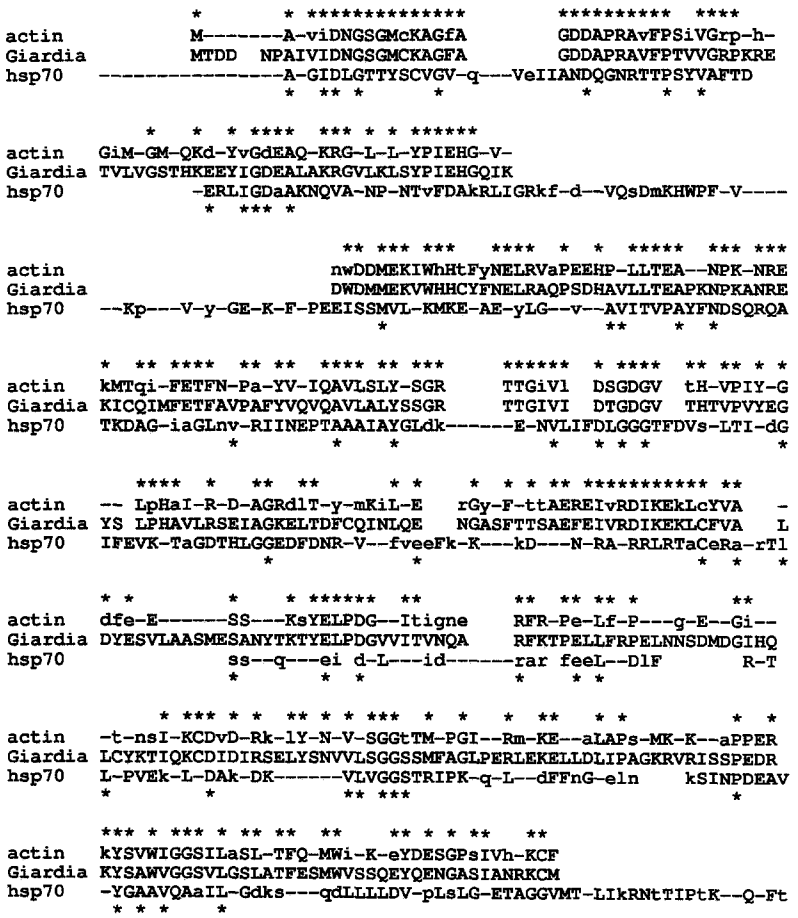

Fig. 3. Comparison of the *G. lamblia* actin amino acid sequence with a consensus actin sequence and a consensus heat shock protein 70 sequence. In the actin consensus sequence, *capital letters* represent perfectly conserved amino acids in all nine genes whereas *lowercase letters* represent amino acids conserved in eight of the nine genes. In the hsp70 consensus, a *capital letter* represents an amino acid sequence perfectly conserved in all six genes, whereas a *lowercase letter* represents an amino acid conserved in five of the six hsp70 genes. In both consensus sequences, *hyphens* represent less well conserved amino acids, whereas spaces represent gaps introduced for alignment. Amino acids that are conserved between the *G. lamblia* actin sequence and the actin consensus sequence or the consensus heat shock protein 70 sequence are indicated by * on top and bottom of this alignment, respectively.

Table 3. Numbers of conserved helical regions of length 4 or greater[a]

x	Number of helices
7	32
8	12
9	4
10	0

[a] Conserved in *x* out of 10 sequences

the coding region ($R^2 = 0.51$), and the base composition of the third base of codons is almost entirely determined by the GC-content of the coding region ($R^2 = 0.98$).

The GC-content bias in the first position of codons is likely entirely due to synonymous substitutions in the first base of arginine and leucine codons. In fact, removing all the positions corresponding to these codons from our alignment eliminated the correlation between the GC-content of this codon position and that of the coding region ($R^2 = 0.06$).

Phylogenetic Analyses

Figure 4 shows the tree obtained from neighbor-joining analysis of the protein sequences. It shows that the *G. lamblia* sequence is highly divergent from all other species; that kinetoplatid protozoans, green plants, fungi, and animals are monophyletic groups; and that the animal and fungi lineages share a more recent common ancestor than either does with the plant lineage. Within fungi, yeasts (the Endomycetales species *Candida albicans, Kluyveromyces lactis,* and *Saccharomyces cerevisiae*) and filamentous fungi (the Plectomycetes species *Aspergillus nidulans* and *Thermomyces lanuginosus* and the Pyrenomycetes species *Trichoderma reesei*) form two distinct groups which originated after the divergence of the early ascomycetes species *Schizosaccharomyces pombe.* This tree also gives strong support to the grouping of the brown algae *Fucus disticus* with the Oomycete species *Achyla bisexualis* and *Phytophthora megasperma* and shows that this lineage is not related to higher fungi. The lineage leading to *Entamoeba histolytica* is shown to have diverged after the lineage leading to Oomycetes and before the lineage leading to plants, animals, and fungi. Surprisingly, it indicates that slime molds (*Dictyostelium discoideum* and *Physarum polycephalum*) and *Acanthamoeba castellanii* form a lineage which diverged after the plant lineage, that the ciliated protozoan species (*Euplotes crassus, Oxytricha nova,* and *Tetrahymena thermophila*) are polyphyletic, and that the two *Plasmodium falciparum* genes are paraphyletic.

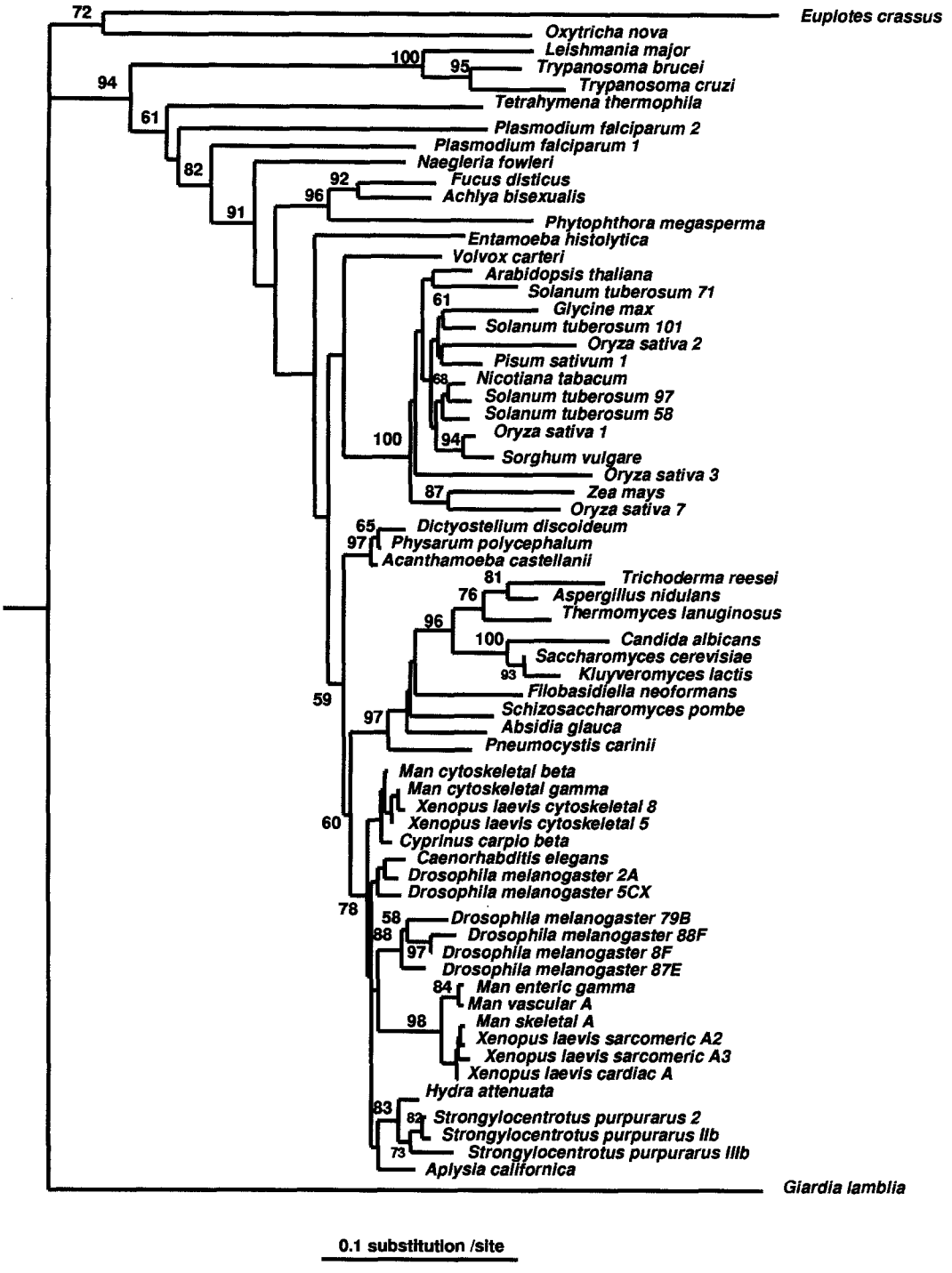

Fig. 4. Phylogenetic tree of actin protein coding sequences. *Numbers* at the nodes represent the bootstrap values out of 100 replicates. Only bootstrap values greater than 50 are indicated.

Phylogenetic analysis using the second base of codons produced a tree very similar to that shown in Fig. 4, except that the two *Plasmodium falciparum* genes were shown to be monophyletic, that the lineage leading to *Naegleria fowleri* was shown to have diverged after the lineage leading to Oomycetes, that the lineage leading to *Entamoeba histolytica* was shown to have diverged after the lineage leading to slime molds, and that *Hydra attenuata* was shown to be an outgroup to all other animal species (results not shown).

Discussion

The amino acid sequence of the *G. lamblia* actin gene shares less than 60% similarity with the actin proteins of the species listed in Table 1. Despite this relatively high degree of divergence, it still shares 204 identical amino acids with the actin consensus shown in Fig. 3. It thus shares 54% amino acid similarity with this consensus, whereas it does not share more than 60% similarity with any single actin from other species. This emphasizes the

highly conserved nature of actin and suggests that most of these conserved amino acids are of functional significance. Furthermore, the fact that the *Tetrahymena thermophila* actin sequence shares 259 amino acids with this consensus, whereas *G. lamblia* shares only 204 such positions, is also consistent with the ancestral nature of the *G. lamblia* sequence.

The determination of the three-dimensional structure of actin has revealed that this protein is related to hsp70 heat shock proteins (Kabsch et al. 1990; Flaherty et al. 1991; Bork et al. 1992). The similarity in three-dimensional structure between actin and hsp70 can also be detected at the amino acid sequence level when gaps corresponding to deletions based on the X-ray structures are forced into an alignment of these two proteins (Doolittle 1992). Doolittle's alignment showed 28 conserved amino acids between four actin sequences and four hsp70 proteins. Our alignment (Fig. 3) shows that the *G. lamblia* actin shares 49 conserved amino acids with a consensus hsp70 sequence when aligned according to Doolittle (1992).

It has been suggested that there is selective pressure on the secondary structure of some mRNAs, such as histone mRNAs (Huynen et al. 1992). Although the *G. lamblia* actin gene mRNA contains a large number of base pairing possibilities with foldings within 10 kcal from the minimum energy of –274.6 kcal/mol, there does not seem to be any significant or conserved RNA secondary structure in either the *G. lamblia* or other single-copy actin gene mRNA coding sequences. Similarly, no stable secondary structures could be detected in either the 5′ or 3′ regions of the *G. lamblia* actin gene. In fact, the 3′ region seems unusual in that it contains no secondary structures (results not shown).

Our phylogenetic tree shows that the *G. lamblia* actin sequence is highly divergent relative to the actin sequences of other eukaryote species. This is consistent with *G. lamblia* having diverged early in evolution, i.e., before the endosymbiotic event which gave rise to mitochondria (Sogin 1989; Sogin et al. 1989; Cavalier-Smith 1993; Hashimoto et al. 1994). Our tree is consistent with previously published trees based on 18S rRNA, tubulin, elongation factor-1α, and actin gene sequences: oomycetes are not true fungi; kinetoplastid protozoans, plants, fungi, and animals are monophyletic groups; and fungi and animals are each others' closest relatives (Förster et al. 1990; Cavalier-Smith 1993; Hasegawa et al. 1993; Taylor et al. 1993; Wainright et al. 1993; Baldauf and Palmer 1993; Nikoh et al. 1994; Bhattacharya and Ehlting 1995).

Within the fungi, only the evolutionary relationships of the ascomycetes are consistent between our tree and the trees based on 18S ribosomal RNA gene sequences (Fig. 4; Taylor et al. 1993; Berbee and Taylor 1993). *Schizosaccharomyces pombe* is shown to diverge before the lineages leading to filamentous ascomycetes and to yeasts, but another likely early ascomycete, *Pneumocystis carinii,* is instead found a the base of the fungi lineage (Fig. 4; Bruns et al. 1992; Berbee and Taylor 1993). Also puzzling is the fact that the basidiomycete species *Filobasidiella neoformans* is part of the ascomycete lineage (Fig. 4).

As mentioned above (see Introduction), the phylogenetic position of *Entanmoeba histolytica* is still controversial. Our tree is consistent with the rRNA phylogenies insofar as they show *E. histolytica* to have diverged after kinetoplastid and amoeboflagellate species (Fig. 4; Sogin 1991; Sogin et al. 1993). However, given the low bootstrap value for the position of this lineage, it is not possible to make a definitive conclusion regarding its relationship to other mitochondria containing eukaryotic lineages.

The phylogenetic position of the cellular slime mold *Dictyostelium discoideum* has also been controversial. Phylogenies based on rRNA suggest that this species diverged before the lineage leading to ciliates, plants, fungi, and animals (Gunderson et al. 1987; Hendricks et al. 1991; Sogin 1989, 1991; Vossbrinck et al. 1987; Wolters 1991: Cavalier-Smith 1993). In fact, the sequence of the small-subunit rRNA molecule of *D. discoideum* is often used as an outgroup to root rRNA phylogenies of ciliates, plant, fungi, and animal phylogenies (e.g., Field et al. 1988; Wainright et al. 1993; Bhattacharya et al. 1995). Furthermore, the recent report by Cole et al. (1995) in which they show that *D. discoideum* mitochondrial DNA encodes a NADH:ubiquinone oxidoreductase subunit which is nuclear encoded in plants, fungi, and animals strongly suggests that the lineage leading to this species diverged before the lineage leading to plants, animals, and fungi. In contrast, phylogenies based on protein sequences suggest that *D. discoideum* is unlikely to be an outgroup to the lineage leading to plants, fungi, and animals. The study of eight protein sequences by Loomis and Smith (1990) led them to conclude that *D. discoideum* is more closely related to animals and plants than to fungi and that this species diverged from the line leading to animals at about the same time as the plant/animal divergence. They also suggested that the relatively high A + T content of the small-subunit rRNA gene of this species was responsible for its erroneous placing in rRNA trees, whereas the protein sequences they used were likely less influenced by the relatively high A + T content of the *D. discoideum* genome and thus likely provided a more reliable phylogenetic position of this species. The study of elongation factor 1 α by Hasegawa et al. (1993) supports the suggestion of Loomis and Smith (1990) insofar as it concluded that *D. discoideum* is unlikely to have diverged from the lineage leading to plants, fungi, and animals before these three kingdoms separated from each other. Our phylogenetic tree agrees with this suggestion and shows that *D. discoideum* diverged after the lineage leading to plants but before the lineage leading to fungi and animals (Fig. 4). Similar results using the first and sec-

Table 4. Relative-rate test between the actin genes of different taxonomic groups and slime molds[a]

Taxon 1	Taxon 2	K_{13}–K_{23}
Animals	Slime molds	3.62 ± 2.65
Angiosperms	Slime molds	5.82 ± 2.39*
Fungi	Slime molds	3.46 ± 2.20

[a] Mean number of estimated nucleotide substitutions at the second position of codons/100 sites and their standard error
* Significantly different at the 5% level

ond codon positions of actin genes or actin amino acid sequences were previously reported (Bhattacharya et al. 1991; Baldauf and Palmer 1993; Bhattacharya and Ehlting 1995). To explain these conflicting results between rRNA and actin phylogenies, Bhattacharya et al. (1991) measured the relative evolutionary rate of actin genes vs the small-subunit rRNA of diverse species and observed that the actin of *D. discoideum* had an evolutionary rate ten times smaller than its small-subunit rRNA. They suggested that the position of this species in the actin tree might reflect the sequence divergence rate slowdown of actin in this species rather than its "true" phylogenetic position (see also Bhattacharya and Ehlting 1995).

Evolutionary Rates

We performed relative-rate tests to determine whether slime mold actin genes are evolving at a rate similar to those of angiosperms, animals, and fungi (Sarich and Wilson 1973; Li and Tanimura 1987). The mean number of nucleotide substitutions at the second position of codons of the sequences shown in Fig. 4, relative to the *Fucus disticus* actin gene was estimated using the Kimura two-parameter model. Only the second position of codons were used to perform these relative-rate tests because it is the only codon site which did not show a correlation with the GC-content of the coding regions (see above), and it was not possible to use the number of nonsynonymous substitutions to perform these tests because such sites were saturated in several pairwise sequence comparisons (results not shown).

The results obtained are consistent with the suggestion of Bhattacharya et al. (1991) that the position of *D. dictyostelium* in the actin tree does not reflect its true phylogenetic position due to a sequence divergence rate slowdown of actin in this species (Table 4). The actin genes of slime molds are evolving significantly slower than those of angiosperms, but their rate of evolution is not significantly different from those of fungi and animal species. The absence of statistically significant slowdown in slime molds as compared to fungi and animals might be due to the fact that we used *Fucus disticus* as an outgroup to the four taxonomic groups on which we performed the relative-rate test. This averages the nucleotide substitutions which occurred in these genes over a time span close to 1 billion years. It might be that the different evolutionary rates reflected by the lengths of the terminal branches within the different taxonomic groups shown in Fig. 4 would become significant if a closer outgroup was available. In our relative-rate estimations, the use of an outgroup as distantly related as *F. disticus* was unavoidable since the relative-rate test requires the use of a sequence that is an outgroup to all the sequences on which the test is to be performed, and a measure of the actual evolutionary rate of actin in slime molds would require one to know the yet-undetermined divergence time between *D. discoideum* and *P. polycephalum.*

Acknowledgments. We thank Walter Gilbert, in whose laboratory the cloning and sequencing of the *Giardia lamblia* actin gene described here was performed. This research was supported in part by a research grant from the National Sciences and Engineering Research Council of Canada to G.D.

References

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci USA 90:11558–11562

Berbee ML, Taylor JW (1993) Dating the evolutionary radiations of the true fungi. Can J Bot 71:1114–1127

Bhattacharya D, Stickel SK, Sogin ML (1991) Molecular phylogenetic analysis of actin genic regions from *Achlya bisexualis* (Oomycota) and *Costaria costata* (Chromophyta). J Mol Evol 33:525–536

Bhattacharya D, Helmchen T, Bibeau C, Melkonian M (1995) Comparisons of nuclear encoded small-subunit ribosomal RNAs reveals the evolutionary position of the Glaucocystophyta. Mol Biol Evol 12:415–420

Bhattacharya D, Ehlting J (1995) Actin coding regions: gene family evolution and use as a phylogenetic marker. Arch Protistenkd 145:155–164

Bork P, Sander C, Valencia A (1992) An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. Proc Natl Acad Sci USA 89:7290–7294

Bruns TD, Vilgalys R, Barns SM, Gonzales D, Hibbett DS, Lane DJ, Simon L, Stickel S, Szaro TM, Weisburg WG, Sogin ML (1992) Evolutionary relationships within the fungi: analyses of nuclear small subunit rRNA sequences. Mol Phylogenet Evol 1:231–241

Cavaler-Smith T (1993) Kingdom protozoa and its 18 phyla. Microbiol Rev 57:953–994

Chan L, Zuker M, Jacobson AB (1991) A computer method for finding common base paired helices in aligned sequences: application to the analysis of random sequences. Nucleic Acids Res 19:353–358

Cole RA, Slade MB, Williams KL (1995) *Dictyostelium discoideum* mitochondrial DNA encodes a NADH:ubiquinone oxidoreductase subunit which is nuclear encoded in other eukaryotes. J Mol Evol 40:616–621

Doolittle RF (1992) Reconstructing history with amino acid sequences. Protein Sci 1:191–200

Felsenstein J (1989) PHYLIP—phylogeny inference package. Cladistics 5:164–166

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselin MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239:748–753

Flaherty KM, McKay DB, Kabsch W, Holmes KC (1991) Similarity of the three-dimensional structures of actin and the ATPase fragment

of a 70-kDa heat shock cognate protein. Proc Natl Acad Sci USA 88:5041–5045

Förster H, Coffey MD, Elwood H, Sogin ML (1990) Sequence analysis of the small subunit ribosomal RNAs of three zoosporic fungi and implications for fungal evolution. Mycologia 82:306–312

Gunderson JH, Elwood H, Ingold A, Kindle K, Sogin ML (1987) Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. Proc Natl Acad Sci USA 84:5823–5827

Hamelin M, Adam L, Lemieux G, Pallotta D (1988) Expression of the three unlinked isocoding actin genes of *Physarum polycephalum*. DNA 7:317–328

Hasegawa M, Hashimoto T (1993) Ribosomal RNA trees misleading? Nature 361:23

Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993) Early branchings in the evolution of eukaryotes: ancient divergence of entamoeba that lacks mitochondria revealed by protein sequence data. J Mol Evol 36:380–388

Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, Goto N, Okamodo K, Hasegawa M (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. Mol Biol Evol 11:65–71

Hendricks L, De Baere R, Van de Peer Y, Neffs J, Goris A, De Wachter R (1991) The evolutionary position of the rhodophyte *Porphyra umbilicalis* and the basiodiomycete *Leucosporidium scottii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. J Mol Evol 32:167–177

Huynen MA, Konings DMA, Hogeweg P (1992) Equal G and C content in histone genes indicate selection pressures on mRNA secondary structure. J Mol Evol 34:280–291

Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures for RNA. Proc Natl Acad Sci USA 86:7706–7710

Jaeger JA, Turner DH, Zuker M (1990) Predicting optimal and suboptimal secondary structure for RNA. Methods Enzymol 183:281–306

Kabnick KS, Peattie DA (1991) Giardia: a missing link between prokaryotes and eukaryotes. Am Sci 79:34–43

Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC (1990) Atomic structure of the actin:DNase I complex. Nature 347:37–44

Le Blancq SM (1994) Chromosome rearrangements in *Giardia lamblia*. Parasitology Today 10:177–179.

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Loomis WF, Smith DW (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. Proc Natl Acad Sci USA 78:9093–9097

Martin W, Somerville CC, Loiseaux-de Goër S (1992) Molecular phylogenies of plastid origins and algal evolution. J Mol Evol 35:385–404

Nikoh N, Hayase N, Iwabe N, Kuma K-i, Miyata T (1994) Phylogenetic relationship of the kingdoms animalia, plantae, and fungi inferred from 23 different protein species. Mol Biol Evol 11:762–768

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sarich VM, Wilson AC (1973) Generation time and genomic evolution in primates. Science 179:1144–1147

Schlegel M (1994) Molecular phylogeny of eukaryotes. Trends Ecol Evol 9:330–335

Sheterline P, Sparrow JC (1994) Actin. Protein Profile 1:1–121

Sidow A, Wilson AC (1990) Compositional statistics: an improvement of evolutionary parsimony and its application to deep branches in the tree of life. J Mol Evol 31:51–68

Sogin ML (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. Am Zool 29:487–499

Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science 243:75–77

Sogin ML (1991) Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1:457–463

Sogin ML, Hinkle G, Leipe DD (1993) Universal tree of life. Nature 362:795

Taylor JW, Bowman BH, Berbee ML, White TJ (1993) Fungal model organisms: phylogenetics of *Saccharomyces, Aspergillus,* and *Neurospora*. Syst Biol 42:440–457

Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR (1987) Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature 326:411–414

Wainright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260:340–342

Winker S, Overbeek R, Woese CR, Olsen GJ, Pfluger N (1990) Structure detection through automated covariance search. Comput Appl Biosci 6:365–371

Woese CR, Gutell RR, Gupta R, Noller HF (1983) Detailed analysis of the higher order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev 47:621–669

Wolters J (1991) The troublesome parasites—molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-cilliate clade. Biosystems 25:75–83

Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. Science 244:48–52

Zuker M (1994) Prediction of RNA secondary structure by energy minimization. Methods Mol Biol 25:267–294

Zuker M, Jaeger JA, Turner DH (1991) A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res 19:2707–2714