Giardia lamblia Expresses a Proteobacterial-like DnaK Homolog

Hilary G. Morrison,* Andrew J. Roger,† Todd G. Nystul,‡ Frances D. Gillin,‡ and Mitchell L. Sogin*†

*The Josephine Bay Paul Center of Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts; †Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada; and ‡Department of Pathology, Division of Infectious Diseases and Center for Molecular Genetics, University of California at San Diego School of Medicine

We identified a novel gene encoding molecular chaperone HSP70 in the amitochondriate parasite Giardia lamblia. The predicted protein is similar to bacterial DnaK and mitochondrial HSP70s. The gene is transcribed and translated at a constant level during trophozoite growth and encystation. Alignment of the sequence with a data set of cytosolic, endoplasmic reticulum (ER), mitochondrial, and DnaK HSP70 homologs indicated that the sequence was extremely divergent and contained insertions unique to giardial HSP70s. Phylogenetic analyses demonstrated that this sequence was distinct from the cytosolic and ER forms and was most similar to proteobacterial and mitochondrial DnaKs. However, a specific relationship with the alpha proteobacterial and mitochondrial sequences was not strongly supported by phylogenetic analyses of this data set, in contrast to similar analyses of cpn60. These data neither confirm nor reject the possibility that this gene is a relic of secondary mitochondrial loss; they leave open the possibility that it was acquired in a separate endosymbiotic event.

Introduction

In many phylogenetic studies, Giardia lamblia (Diplomonadida) consistently represents one of the most basal eukaryotic branches (Leipe et al. 1993; Cavalier-Smith and Chao 1996; Hashimoto and Hasegawa 1996), although its deep-branching position has been questioned (Stiller and Hall 1999; Morin 2000). Giardia lacks mitochondria and peroxisomes, and its energy metabolism appears to be cytosolic. There is no compelling evidence yet for the presence of introns or functional splicesomal machinery. Giardia quite successfully makes its living as a parasite of the mammalian gut, often using metabolic pathways that are most closely related to prokaryotes. Giardia’s 12-Mb genome (Adam 2000) is small compared with many eukaryotic genomes, e.g., those of plants, animals, and many protists. What is the common denominator that creates a eukaryotic cell? We undertook sequencing of the Giardia genome to answer this and other fundamental questions about this presumptively early diverging eukaryote.

A key event in the evolution of the eukaryotic cell was the acquisition of the mitochondrion via the endosymbiosis of a prokaryote. For many years, this endosymbiotic event was assumed to have occurred subsequent to the divergence of the most basal lineages. However, evidence has been accumulating that mitochondria or other ancestral symbionts occurred in these early-branching lineages. Genes have recently been discovered in the nuclear genomes of these amitochondriates that appear to have been introduced by an ancestral endosymbiont. Coding regions for α-proteobacterial or mitochondrion-targeted molecular chaperones have been discovered in the nuclear genomes of microsporidia (Germot, Philippe, and Le Guyader 1997; Hirt et al. 1997; Peyretaillade et al. 1998), trichomonads (Bui, Bradley, and Johnson 1996; Germot, Philippe, and Le Guyader 1996; Horner et al. 1996; Roger, Clark, and Doolittle 1996; Hashimoto et al. 1998), and Giardia (Roger et al. 1998). The most parsimonious explanation of this phylectic distribution pattern is that mitochondria were present in the earliest eukaryotes (the mitochondria-early hypothesis). However, phylogenetic analyses of cpn60 and HSP70 do not consistently recover topologies congruent with nuclear gene trees as represented by rRNA analyses. Either rRNA trees are unreliable, or the chaperonins lack phylogenetic resolving power, or the chaperonin genes have been acquired independently by different lineages.

To address these issues, we characterized a new HSP70 gene from G. lamblia. Heat shock protein 70s (HSP70s) are the most prevalent of the molecular chaperones and are found in all domains of life. The bacterial HSP70 homolog is known as DnaK. Most eukaryotes have at least three HSP70 genes, and the gene products are compartmentalized to the cytosol, the endoplasmic reticulum (ER), or the mitochondrion. The cytosolic and ER forms result from an ancient gene duplication in the eukaryotic lineage, whereas the mitochondrial form was acquired from the endosymbiosis of the DnaK-containing proteobacterium that became the mitochondrion. Gupta et al. (1994) reported the cloning of the ER and cytosolic homologs of HSP70 in Giardia, but no mitochondrial homolog was discovered. Here, we report on the cloning, sequencing, expression studies, and phylogenetic analysis of a DnaK-like form of HSP70 from Giardia that is clearly of bacterial origin.

Materials and Methods

Parasites and Culture

Giardia lamblia strain WB was isolated from a patient with chronic symptomatic giardiasis (Smith et al. 1982) and belongs to the most common group of isolates from humans (Ey et al. 1996). WB clone C6 (ATCC...
number 50803) trophozoites were routinely subcultured twice weekly in filter-sterilized Diamond’s TYI-S-33 medium with 10% adult bovine serum (Irvine Scientific, Santa Ana, Calif.) (Diamond, Harlow, and Cunnick 1978) and 0.5 mg/ml of bovine bile (Keister 1983), without added vitamins, antibiotics, or iron, at pH 7.0–7.1.

Cloning and Sequencing of *G. lamblia* HSP70

The construction of a *Giardia* genomic DNA library in the plasmid pBluescript has previously been described (Henze et al. 1998). A basic local alignment tool (BLASTX; Altschul et al. 1990) search of sequence data from one clone returned a highly significant hit to hydrogenosomal HSP70 from *Trichomonas vaginalis* and to various bacterial DnaK proteins. A 558-bp fragment was excised from the clone by EcoRI digestion and used to probe a *Giardia* λ-ZapII genomic library using a digoxigenin detection protocol (Boehringer Mannheim, Indianapolis, Ind.). A clone containing an insert of approximately 9 kb was identified, and the pBluescript phagemid was excised following the manufacturer’s protocols (Stratagene, Valencia, Calif.). A 2,130-bp region containing the HSP70 coding sequence was sequenced on both strands. Sequencing was carried out using the Excel II cycle sequencing protocol (Epicentre Technologies, Madison, Wis.). Sequencing reaction products were run on LI-COR 4200 automated sequencers (Middendorf et al. 1992; Roger et al. 1998).

HSP70 Expression

Expression of total HSP70 proteins during differentiation and heat and oxygen stress was monitored in preliminary experiments by immunoblotting using polyclonal antibodies to yeast ssc-1 (mitochondrial HSP70) as a probe. (The polyclonal antibodies were a kind gift from Dr. Elizabeth Craig, University of Chicago.) Since this reagent was made against a region that was similar between giardial ER, cytoplasmic, and mitochondrial HSP70 genes, this gave an estimate of overall changes in HSP70 levels. Since no significant changes were noted in the initial studies, Northern analyses were carried out using conditions of the greatest stress tolerated by the parasites. Protein and RNA isolation were as described below.

Immunoblot Analyses

Parasites were harvested at the indicated stages of differentiation or after the stress conditions specified by extraction with RNAzol B according to the manufacturer’s instructions (Tel-Test Inc., Friendswood, Tex.). Samples of total RNA (10 μg per lane) were fractionated in 1.5% formaldehyde-agarose gels, downward capillary blotted in 20 × SSC, and immobilized onto nylon membranes (Zeta-Probe, Bio-Rad, Hercules, Calif.) by UV cross-linking. For Northern hybridization, a probe corresponding to the open reading frame of the giardial DnaK-like HSP70 was purified and radiolabeled by random priming (Prime It II kit, Stratagene). Blots were prehybridized in 6 × SSC, 5 × Denhardt’s solution, 0.5% (w/v) SDS, and 20 μg/ml salmon sperm DNA for 1 h at 65°C. Hybridization at 65°C was continued overnight in the presence of the HSP70 probe (Knodler et al. 1999). The membrane was washed twice in 2 × SSC/0.1% (w/v) SDS at room temperature for 15 min, and then once at 60°C for 15 min in 0.2 × SSC/0.1% (w/v) SDS. The washed membrane was autoradiographed overnight.

Induction of Encystation

Pre-encysting cultures were grown to late log phase (66 h) in TYI-S-33 medium (pH 7.0) without bile but containing the antibiotics piperacillin (500 μg/ml, Led-erle Laboratories, Carolina, Puerto Rico) and amikacin (125 μg/ml; Bristol Laboratories, Syracuse, N.Y.), which do not affect *G. lamblia* growth or differentiation (Meng, Hetsko, and Gillin 1996). Encystation was initiated by removing the spent medium and nonadherent cells and refeeding the adherent trophozoite monolayer with encystation medium: TYI-S-33 with antibiotics but without bovine bile, adjusted to pH 7.8 with NaOH and supplemented with 0.25 mg/ml porcine bile and 5 mM lactic acid, (hemicalcium) (Meng, Hetsko, and Gillin 1996), which increases biological activity of cysts (Boucher and Gillin 1990).

Heat Shock

For protein isolation, 250 ml of growth medium was inoculated at 700 cells/ml and grown to confluence with NaOH, boiled for 6 min, and stored at −70°C. Twenty microliters of parasite extract was separated by SDS-PAGE on 4%–20% gradient gels (Novex, San Diego, Calif.) (Laemmli 1970). The separated antigens were transferred to nitrocellulose membranes (Towbin, Staehelin, and Gordon 1979) for 18 h at 30 V and then for 1 h at 70 V. Membranes were blocked in PBS with 7.5% (w/v) BSA for 2 h and washed thoroughly. The blots were probed for 2 h with polyclonal antibodies to yeast ssc1 diluted 1:500 in PBS, washed in Tris-buffered saline, reacted with Protein A-HRP (Zymed, San Francisco, Calif.) diluted 1:2,000 in Tris-buffered saline for 1 h, washed in Tris-buffered saline, and developed with ECL reagents according to the manufacturer’s instructions (Amersham Life Sciences, Arlington Heights, Ill.).

RNA Isolation

Total RNA was isolated from *G. lamblia* at the indicated stages of differentiation or after the stress conditions specified by extraction with RNAzol B according to the manufacturer’s instructions (Tel-Test Inc., Friendswood, Tex.). Samples of total RNA (10 μg per lane) were fractionated in 1.5% formaldehyde-agarose gels, downward capillary blotted in 20 × SSC, and immobilized onto nylon membranes (Zeta-Probe, Bio-Rad, Hercules, Calif.) by UV cross-linking. For Northern hybridization, a probe corresponding to the open reading frame of the giardial DnaK-like HSP70 was purified and radiolabeled by random priming (Prime It II kit, Stratagene). Blots were prehybridized in 6 × SSC, 5 × Denhardt’s solution, 0.5% (w/v) SDS, and 20 μg/ml salmon sperm DNA for 1 h at 65°C. Hybridization at 65°C was continued overnight in the presence of the HSP70 probe (Knodler et al. 1999). The membrane was washed twice in 2 × SSC/0.1% (w/v) SDS at room temperature for 15 min, and then once at 60°C for 15 min in 0.2 × SSC/0.1% (w/v) SDS. The washed membrane was autoradiographed overnight.
over 3 days at 37°C. Attached cells were isolated by pouring off medium containing free-swimming cells, refed with fresh media, harvested by scraping, and transferred to 8-ml glass tubes to facilitate heat transfer. After a 2-h preincubation period at 37°C, cells were incubated in a 42°C water bath for 15, 30, or 45 min. The tubes were periodically inverted during the incubation period to promote even heat transfer. Following the 42°C incubation, cells were allowed to recover at 37°C for 0, 30, or 60 min and harvested by chilling on ice for 20 min. For RNA isolation, cells were grown and heat shocked as described above except that the cells were preincubated at 37°C for 1 h and the 42°C incubation was carried out in 16-ml tubes for 45 min. Following the 42°C incubation, cells were allowed to recover at 37°C for 30 or 60 min and then harvested with cell scrapers.

Oxygen Stress

For protein isolation, 65-ml culture flasks were filled to 40%, 50%, 60%, 70%, 80%, 90%, and 100% capacity with growth medium and inoculated with 2 x 10^7 cells total. Cells were grown overnight at 37°C, and protein was isolated as above. For RNA isolation, 65-ml culture flasks were filled to 50% capacity with growth medium (the maximum oxygen exposure tolerated well by the cells in the initial experiments) and inoculated with 1.5 x 10^7 cells total. Cells were grown overnight at 37°C, and total RNA was isolated as above.

Rapid Amplification of cDNA Ends (RACE) Analysis

Rapid amplification of cDNA ends (5′ RACE) was employed to identify the start of transcription of the DnaK-like HSP70. 5′ RACE was performed using the 5′ RACE System for Rapid Amplification of cDNA ends, version 2.0 (Life Technologies, Gaithersburg, Md.), according to the manufacturer’s instructions. Oligo HSP70.3 (TGG-TTC-CTT-GCT-GTG-TCT-GGC-C) was used as the first-strand primer, and oligo HSP70.4 (TTC-ATG-GAC-TTG-GTG-TCT-GG) was used as the nested primer (Knodler et al. 1999).

Sequence Alignment

A database containing 56 HSP70 homologs from the Archaea, Bacteria, and Eukarya was assembled from GenBank. Protein sequences were aligned using CLUSTAL W (Thompson, Higgins, and Gibson 1994). The alignment was refined manually within the SEQLAB module of the GCG package (Genetics Computing Group, Madison, Wis.), taking into account structural elements identified in bovine HSP70 (hsp70) and Escherichia coli DnaK (Flaherty et al. 1994; Zhu et al. 1996). Regions of questionable alignment were removed from the data sets used in phylogenetic analyses.

Phylogenetic Analyses

Phylogenies were inferred using parsimony, distance, and maximum-likelihood (ML) methods on the aligned amino acid sequences. Protein ML trees were inferred using the quick-add OTU (-q option) tree-searching procedure implemented in the PROTML program, version 2.2, using the JTT-F amino acid substitution model (Adachi and Hasegawa 1996). To avoid inconsistency problems introduced by the presence of invariable sites in the alignment, the ML estimate of the proportion of invariable sites (P_inv) in the HSP70 data set was obtained using the PUZZLE program, version 4.0 (Strimmer and von Haeseler 1996), and invariable positions were selectively removed from the alignment before PROTML analysis. To study the impact of among-sites rate variation on the phylogenetic trees, the quartet puzzling ML method was used to infer trees with a mixed eight-category discrete gamma and invariable-sites model of rate heterogeneity and the JTT-F substitution model (JTT-F+Γ+Inv). P_inv and the gamma shape parameter α were estimated by ML optimization on a neighbor-joining topology. ML protein distances were inferred with PUZZLE using the same model of evolution, and distance trees were estimated using the Fitch-Margoliash method with global rearrangements (using the FITCH program of PHYLIP, version 3.57c; Felsenstein 1996). Unweighted maximum-parsimony analysis was carried out by 100 rounds of random stepwise addition heuristic searches with tree bisection-reconnection (TBR) branch swapping using the PAUP* program, version 4b1 (Swofford 1997). Bootstrap analyses for protein distance trees were carried out using the PUZZLEBOOT program (http://www.tree-puzzle.de/puzzleboot.sh). Deviations of amino acid frequencies in a given sequence from the overall frequencies in the data set were detected by a chi-square test implemented in PUZZLE.

Results

Characterization of the G. lamblia HSP70 Gene

The intronless putative coding region is 1,923 bp, encoding 640 amino acids (GenBank AF274582). Transcriptional start sites were located 13 and 18 bases upstream of the start codon (AAATGAAATGCGGCCATG). Both transcript start sites are A/T-rich and are located downstream from another A/T-rich sequence, as is typical for Giardia (Holberton and Marshall 1995; Hilario and Gogarten 1998). However, no giardial CAT box sequence, shown to be important for expression of the glutamate dehydrogenase gene, was present (Yee et al. 2000). The overall G+C content of the coding sequence was 50.13%. The first of two potential polyadenylation signals (CGTAA) (Adam 1991; Que et al. 1996) overlaps the stop codon. The second (TGTAAT) is very similar to the polyadenylation signal of the Giardia protein disulfide isomerase-3 gene (Knodler et al. 1999).

Expression

Total levels of HSP70 protein were constant during trophozoite growth and encystation (fig. 1A). Giardia is sensitive to oxygen; however, we found that HSP70 protein was not significantly increased by sublethal oxygen
Giardia lamblia DnaK Homolog

**A. Encystation Western**

![Western blot image](image)

**B. Oxygen Stress Western**

Percent air: 0, 10, 20, 30, 40, 50, 60

![Western blot image](image)

**C. Heat Shock Western**

Exposure (min): C 15, 30, 45

Recovery (min): 0, 30, 60

![Western blot image](image)

**D. Heat Shock Northern**

Exposure (min): C 45

Recovery (min): 30, 60

![Northern blot image](image)

**Fig. 1.**—HSP70 is expressed constitutively and not upregulated in response to heat shock or oxygen stress. **A,** Encystation, Western analysis. Protein was harvested from trophozoites (T) and 24-h encysting (24) cells. **B,** Oxygen stress, Western analysis. Cells were grown overnight in flasks filled to 90%, 80%, 70%, 60%, 50%, or 40% capacity with normal growth medium. The control flask was filled to 100% with normal growth medium. **C** and **D,** Heat shock. Cells were heat-shocked at 42°C for 15, 30, or 45 min and then allowed to recover for 0, 30, or 60 min before protein (C) or RNA (D) was harvested. The control cells were incubated at 37°C, while the rest were heat shocked and recovered. In D, the ratio of HSP70 message to that of the GDH control did not change significantly with the most severe heat shock treatment.

stress (fig. 1B) or heat shock (fig. 1C). The constant expression of the DnaK-like HSP70, in response to oxygen stress or heat shock, was confirmed by Northern analyses with a specific probe relative to a glutamate dehydrogenase control (fig. 1D and data not shown).

Alignment of the *Giardia lamblia* DnaK Homolog

An alignment of the translated protein sequence with 56 cytosolic, ER, mitochondrial-like, and bacterial HSP70s from a variety of taxa (fig. 2) revealed that this *Giardia lamblia* DnaK homolog is distinct from the Giardia cytosolic and ER HSP70s. The amino terminus of the Giardia DnaK-like HSP70 is hydrophobic, consistent with potential translocation across an organelle membrane but not with import into the ER (Nielsen et al. 1997). No C-terminal membrane spanning region or signature sequences related to endoplasmic reticulum targeting or retrieval are obvious. The protein contains many of the residues conserved among mitochondrial and α-proteobacterial HSP70s (Flaherty et al. 1994; Wilbanks, DeLuca-Flaherty, and McKay 1994; Kamath-Loeb et al. 1995; O’Brien, Flaherty, and McKay 1996; Zhu et al. 1996; Suh et al. 1998; Wang et al. 1998), as indicated in figure 2. However, other signature sequences (Germot, Philippe, and Le Guyader 1996, 1997; Hirt et al. 1997; Peyretaillade et al. 1998) are not well conserved. For example, the signature sequence GDAW, present in the mitochondrial-like HSP70 of *Trichomonas vaginalis*, is replaced by GEAM in Giardia. Similarly, the signature sequence YSPAQIG is not well conserved in Giardia, appearing as VSPIEVG.

The Giardia sequence contained several small insertions not found in any other sequence (fig. 2). The first occurred at residues 49–51, a site that falls between two β-strand sheets in bovine hsc70, and the second occurred at residues 147–149, between a hydrogen-bonded turn and a β-strand sheet in bovine hsc70 (fig. 2). A third insertion occurred at residues 199–202; interestingly, the genes for the cytosolic and ER HSP70s in Giardia also each contain a distinct insertion at this location. Finally, a larger insertion occurred at residues 263–276. This site falls between a helical region and a hydrogen-bonded turn in bovine hsc70. These regions were excluded from the data set used to infer trees, and their significance is unknown.

**Phylogenetic Analysis of the HSP70 Data Set**

Protein ML and maximum parsimony trees were inferred for 56 taxa using 465 reliably aligned sites (fig. 3 and data not shown). Both methods placed Giardia within the proteobacteria, but not specifically within the α-proteobacterial or mitochondrial groups. In figure 3, Giardia DnaK, the mitochondrial HSP70s, and all of the proteobacterial sequences branched together with a bootstrap value of 66%. Within this group, the mitochondrial HSP70s plus the α-proteobacterial DnaKs, excluding the Ehrlichia and microsporidial sequences, grouped with a bootstrap value of 56%. Giardia did not branch within this group, nor was there significant bootstrap support for its placement as a sister group to the Ehrlichias. The cytosolic and ER proteins grouped together, with the Giardia cytosolic and ER proteins branching most deeply. The Giardia protein branched with the three microsporidial proteins in both analyses; however, we suspected that this was a problem of long-branch attraction, and the group did not hold up on further examination. Since it was apparent that the new Giardia sequence was specifically related to the bacterial DnaK/mitochondrial HSP70 group, the cytosolic and ER sequences were not included in the later analyses in order to reduce the number of taxa in the data set. This permitted the inclusion of a larger number of aligned sites and the use of more rigorous, computationally demanding methods of analysis.
Phylogenetic Analysis of a Restricted Data Set

Based on the preliminary analysis, we created a smaller data set for more intensive analysis by selecting taxa that branched near the Giardia sequence. These taxa included mitochondrial and proteobacterial sequences, with the Chlamydia and spirochetes as immediate outgroups. We performed likelihood and distance analyses on this data set.

In analyses without a correction for among-sites rate variation (protein ML, quartet puzzling ML, and ML distance), the Giardia sequence grouped with HSP70 sequences from the microsporidia, *Nosema* *loccustae*, and *Encephalitozoon cuniculi*. This group gained strong bootstrap support from PROTML, distance, and quartet puzzling analyses (fig. 4A). However, the exact placement of this group in the HSP70 tree was poorly resolved—the ML tree displayed the microsporidia/Giardia group outside the α-proteobacterial/mitochondrial clade (not shown), while distance analyses placed it at the base of mitochondria (fig. 4A). Each of these placements was weakly supported by the respective methods; bootstrap values and puzzling support values were below 50% in all cases. Indeed, the quartet puzzling consensus tree showed the Giardia, microsporidian, mitochondrial, and α-proteobacterial sequences in an unresolved polytomy (not shown).

Not correcting for among-sites rate variation can lead to problems whereby sequences with higher rates of evolution artifactually group together (the "long-branch attraction" problem; Lockhart et al. 1996; Yang 1996). Since the microsporidian and Giardia sequences formed extremely long branches in our analyses (fig. 4A), we suspected that their strong association could be due to this kind of artifact. To test this hypothesis, we performed ML distance and quartet puzzling ML analyses using the gamma model of among-sites rate variation implemented in PUZZLE, version 4.0. Consistent with the artifact explanation, trees recovered from these analyses were not supported by the respective methods.
analyses did not show a specific association between the microsporidian and the Giardia sequences (fig. 4B). However, the exact placement of Giardia and microsporidia in these trees was again not well resolved. In the distance analysis, Giardia fell outside the α-proteobacterial/mitochondrial clade, with microsporidia emerging at the base of mitochondria. Bootstrap and quartet puzzling support values for branches separating Giardia from the mitochondrial sequences were all below 50%. One source of instability in these analyses was the two Ehrlichia sequences. Although Ehrlichia are well known to be a sister group of the Rickettsia (Weisburg et al.
To examine the placement of the Giardia sequence without the confounding bias and instability introduced by the microsporidian and Ehrlichia sequences, we removed them from the alignment and applied gamma-corrected ML distance and ML quartet puzzling methods to the data. The quartet-puzzling consensus tree

1989; Roger et al. 1998), in our various HSP70 analyses the two Ehrlichia sequences moved around, sometimes grouping together weakly (53% bootstrap support in PROTML analysis) and sometimes splitting apart, not always specifically allied with the Rickettsia sequence (fig. 4A).
showed an unresolved polytomy containing mitochondrial plus α-proteobacterial, Giardia, and γ-proteobacterial sequences indicating little resolution in the data. Although the support for the Giardia/protectobacteria/mitochondria clade was strong (quartet puzzling support = 90%, ML distance bootstrap = 90%) both trees showed Giardia being outside the mitochondrial/α-proteobacterial clade, which was moderately supported (quartet puzzling support = 74%, bootstrap support = 70%). To determine whether this placement outside of the mitochondrial/α-proteobacterial clade was significant, we moved the position of the Giardia sequence to all possible branching positions in the backbone ML distance Fitch-Margoliash tree and evaluated the likelihood score for each position under the JTT+Γ model. The optimal position in ML analysis, shown in figure 5, displayed the Giardia sequences as an immediate outgroup to the α-proteobacteria plus mitochondria. However, the next most likely position of Giardia in this tree was as a sister group to mitochondria, with a difference in log likelihood of 0.4 compared with the optimal tree (fig. 5). To test whether the alternative positions for Giardia were significantly worse, we used Kishino-Hasegawa tests (Kishino and Hasegawa 1989) to evaluate the significance of differences in likelihood for alternative topologies. These tests indicated that of all the possible placements of Giardia in the tree, only two could be significantly excluded at the 5% level (indicated branches in fig. 5).

Therefore, since the branching point of the Giardia HSP70 can be moved to almost anywhere in the tree without a significant decrease in the likelihood, it appears that the data cannot resolve its true position. This is at least partly due to the extremely long branch leading to Giardia HSP70 in trees. Most phylogenetic signal that would allow a clear placement of this sequence in the HSP70 tree has been erased by the extreme divergence of this sequence. This is in sharp contrast to the Giardia chaperonin 60 that shows a strongly supported relationship to the mitochondrial lineage (Roger et al. 1998). HSP70 and cpn60 work sequentially in refolding proteins during organellar import (Stuart et al. 1994), and both are of mitochondrial origin in all eukaryotes studied so far, including the amitochondriate parabasalids (Bui, Bradley, and Johnson 1996; Germot, Philippe, and Le Guyader 1996; Horner et al. 1996; Roger et al. 1998). While a mitochondrial origin for the Giardia DnaK is not specifically supported, it nevertheless remains a possibility.

The Impact of Taxon Deletion on the Gamma Shape Parameter

The gamma distribution model of sequence evolution assumes that sites in a gene evolve at different rates.
but that those rates do not change over the tree. The gamma shape parameter, alpha, ranges from $\alpha = \infty$ (no rate heterogeneity) to $\alpha < 1$ (strong heterogeneity). The value of alpha estimated for the whole data set was 0.6, indicating strong heterogeneity in rates of evolution at different sites. The highly divergent nature of the Giardia sequence (see fig. 2) suggested that the pattern of rates at sites may have changed in this lineage. If so, the value of alpha should change if Giardia is excluded from the data set. To test this hypothesis, we investigated the impact of the presence or absence of sequences on the estimate of the gamma shape parameter by systematically deleting proteobacterial or mitochondrial taxa from our analyses and reestimating the shape parameter (table 1). The deletion of most of these taxa had little effect on the estimate of the shape parameter, which remained close to 0.6, the value estimated for the whole data set. However, deletion of several groups had a much larger effect. As expected, removing Giardia produced the largest effect, yielding an alpha estimate of 0.55, while deletion of trypanosomes and the microsporidia had the next largest effects, also yielding lower alpha estimates of 0.56 and 0.57, respectively. Since the microsporidian and Giardia HSP70 sequences appear to artefactually group together in some analyses (see above), we examined the effect of deleting both. This caused the largest decrease in the alpha estimate, with $\alpha = 0.51$.

Combining sets of sequences with different patterns of among-sites rate variation leads to an “averaging” effect on among-sites rate variation for the whole data set. Under these conditions, the estimate for the gamma shape parameter, alpha, is increased, falsely indicating less extreme rate variation among sites for the whole data set than is displayed by either set of sequences alone (Lockhart et al. 1998). Conversely, in our analyses, the decreases in the alpha parameter estimates upon removal of the Giardia, microsporidia, and trypanosome HSP70s indicate that these sequences have a different pattern of rates of sites than the rest of the HSP70 data set. Consideration of the alignment (fig. 2) suggests that the Giardia and microsporidia sequences, in particular, have fewer sequence constraints and are more variable in general than other proteobacterial and mitochondrial HSP70s (as might be expected for a gene whose functional constraints have been relaxed). It is possible that the decrease in constraints in the HSP70s from these amitochondriate groups, coupled with an elevated rate of evolution, contributes to their artefactual grouping in some of our phylogenetic analyses (fig. 4A). Furthermore, the changes in the pattern of substitution in these taxa violate both the equal-rates and the gamma distribution models we used and may be partly responsible for the difficulties encountered in discerning the true phylogenetic placement of the Giardia HSP70 sequence.

### Table 1

<table>
<thead>
<tr>
<th>Excluded Taxa</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.60</td>
</tr>
<tr>
<td>Rickettsia, Ehrlichia sp.</td>
<td>0.62</td>
</tr>
<tr>
<td>Animals</td>
<td>0.61</td>
</tr>
<tr>
<td>Plants, Eimeria</td>
<td>0.61</td>
</tr>
<tr>
<td>Ehrlichia</td>
<td>0.60</td>
</tr>
<tr>
<td>$\alpha$-proteobacteria</td>
<td>0.60</td>
</tr>
<tr>
<td>Rhodobacter, Caulobacter</td>
<td>0.60</td>
</tr>
<tr>
<td>Spirochetes</td>
<td>0.60</td>
</tr>
<tr>
<td>Agrobacterium, Bradyrhizobium</td>
<td>0.59</td>
</tr>
<tr>
<td>Chlamydia</td>
<td>0.59</td>
</tr>
<tr>
<td>Fungi</td>
<td>0.59</td>
</tr>
<tr>
<td>Trichomonas</td>
<td>0.58</td>
</tr>
<tr>
<td>Microsporidia</td>
<td>0.57</td>
</tr>
<tr>
<td>Trypanosomes</td>
<td>0.56</td>
</tr>
<tr>
<td>Giardia</td>
<td>0.55</td>
</tr>
<tr>
<td>Giardia, Microsporidia</td>
<td>0.51</td>
</tr>
</tbody>
</table>

**Note.**—To ensure that missing data did not affect the parameter estimates, alignment positions containing missing data were removed from the data set, leaving 47 sites for this analysis. For each data set, $\alpha$ was estimated by maximum-likelihood optimization on neighbor-joining topology.

### Discussion

Genes encoding HSP70 have been extensively used as phylogenetic markers. They meet several criteria for useful markers: they are conserved in all domains of life, they can be aligned based on highly conserved sites and functional domains, they contain enough variable sites to provide phylogenetic signal, and they have undergone gene duplication events, so that gene trees can be reciprocally rooted. However, it is becoming increasingly evident that HSP70 genes do not have the power to resolve distant phylogenetic relationships, particularly in cases where the genes are serving different functions in different organisms. For example, in recently published analyses of HSP70 (Germot and Philippe 1999), the position of T. vaginalis could not be reliably determined, appearing in some analyses as a sister group to Giardia, in others as a sister group to fungi/metazoa. The DnaK HSP70 homologs, in particular, appear unable to resolve the relationships between lineages that are either deep-branching or fast-evolving. Ideally, genomic sequence for many of these lineages will one day be available to clarify these relationships.

An objective look at the data cited as evidence that Giardia once had mitochondria reveals that they are not robust. Henze et al. (1995), in a study of glyceraldehyde-3-phosphate dehydrogenase, considered the possibility that Giardia was secondarily amitochondriate. However, they correctly point out that the argument is based solely on the presence of a eubacterial GAPDH gene. Keeling and Doolittle (1997) concluded that Giardia’s triosephosphate isomerase is of $\alpha$-proteobacterial origins. While their conclusion may be correct, it is a strong statement to make based on a problematic analysis that showed (1) no bootstrap support for the monophyly of the eukaryotes, yet subsequent treatment of the eukaryotes as a single group; (2) a single representative of the $\alpha$-proteobacteria in the data set used, and (3) non-significantly different tree topologies when the outgroup to the eukaryotes was all proteobacteria versus $\alpha$-proteobacteria.

Other analyses, in which a sister taxon relationship has been observed between Giardia and Trichomonas,
i.e., analyses of valyl-tRNA synthetase (Hashimoto et al. 1998), beta-tubulin (Keeling and Doolittle 1996), and cpn60 (Roger et al. 1998), have been taken as evidence that if Trichomonas once harbored the mitochondrial endosymbiont, Giardia must have done so as well (Roger 1999; Roger, Morrison, and Sogin 1999). However, a sister relationship of the diplomonads and parabasalids is not seen in analyses of other proteins, such as alpha tubulins (Keeling and Doolittle 1996) or the cytosolic and ER forms of HSP70 (Germot and Philippe 1999).

The credibility of the sister relationship seen in the analysis of ValRS is weakened by the limited number of taxa included and the observed specific relationship of the Trichomonas/Giardia grouping to the green plants. This is not consistent with any other published molecular analysis, nor is there any phenotypic basis for describing diplomonads as being specifically related to photosynthetic chlorophytes or green plants. The sister group relationship might be artificial, an example of the long-branch attraction problem, in which case this line of reasoning becomes invalid. The sister relationship hypothesis should be tested using additional protein-coding sequences as such data become available for both diplomonads and parabasalids.

The strongest evidence thus far that Giardia once harbored a mitochondrial symbiont comes from analysis of cpn60 (Roger et al. 1998). If this is true, the simplest explanation for the origin of the DnaK-like HSP 70, given its affinity for the proteobacterial DnaK homologs, is that it is also a mitochondrial relic. The cpn60 phylogeny indicates a specific relationship of Giardia’s cpn60 to the alpha-proteobacterial and mitochondrial cpn60. The gene tree also shows evidence of a sister group relationship between Trichomonas and Giardia, although Entamoeba also falls within this clade. The inclusion of the Entamoeba sequence was considered an artifact caused by long-branch attraction and by a statistically significant amino acid composition bias shared by Giardia and Entamoeba (Roger et al. 1998). If the putative sister relationship is true, it is interesting that the cpn60 phylogeny apparently can resolve deep relationships while HSP70 cannot. Neither the present function nor the cellular localization is known for either cpn60 or this DnaK-like HSP70. If both are of mitochondrial origin, presumably they have been evolving for the same amount of time but not under the same constraints, which suggests that they function independently.

Genes have been discovered in amitochondriate lineages that strongly suggest the occurrence of multiple lateral transfer or endosymbiotic events from bacteria other than the alpha-proteobacterial symbiont that gave rise to the mitochondrion. One example in Giardia is a class 2 3-hydroxy-3-methylglutaryl coenzyme A reductase (Boucher and Doolittle 2000) which appears to have been acquired from a bacterial source distinct from the presumptive alpha-proteobacterial symbiont. The discovery in Giardia of an iron-hydrogenase gene most closely related to a gene found in Entamoeba argues that a lateral gene transfer event may have occurred, with the possible source being an anaerobic bacterium living in animal digestive tracts (J. Nixon et al., personal communication). It has been noted by others that Giardia’s residence in the intestinal environment would make it easy to acquire genes from bacteria (Boucher and Doolittle 2000). It is likely that many other genes of bacterial origin will be evident in Giardia’s genome when the genome sequence data are complete. Even if Giardia is secondarily amitochondriate, not all of its bacterial-like genes are necessarily mitochondrial relics. It will be interesting and informative to determine what proportion of Giardia’s bacterial relics are specifically related to alpha-proteobacterial homologs.

The existing data, including the results reported here, clearly demonstrate that the genomes of amitochondriate organisms contain genes of endosymbiotic origin. However, phylogenetic inferences do not strongly support the most parsimonious hypothesis that a single early endosymbiotic event (the mitochondria-early hypothesis) accounts for the origins of both cpn60 and DnaK-like HSP70s in amitochondrial lineages. Only the cpn60 gene shows a specific affinity with homologs from the mitochondria. The DnaK-like HSP70 genes merely suggest an unresolved phylogenetic affinity with proteobacterial and mitochondrial DnaK forms. Evolutionary history has been obscured by rapid evolution of the genes in amitochondriate taxa, and the present functions of these genes in these taxa are not known.

**Supplementary Material**

The complete amino acid alignment has been submitted to EMBL under alignment number DS44922.

**Acknowledgments**

Polyclonal antibodies to yeast ssc1 were a kind gift from Dr. Elizabeth Craig, University of Chicago. This work was supported by the G. Unger Vetlesen Foundation; grants GM32964 and AI43273 from the National Institutes of Health, awarded to M.L.S; and grant AI42488 from the NIH, awarded to F.D.G. Additional support for the Giardia genome project was provided by LI-COR Biotechnology, Inc.

**LITERATURE CITED**


JULIAN ADAMS, reviewing editor

Accepted November 24, 2000