

Letter to the Editor

Phylogenetic Analysis of the TATA Box Binding Protein (TBP) Gene from *Nosema locustae*: Evidence for a Microsporidia–Fungi Relationship and Spliceosomal Intron Loss

Naomi M. Fast, John M. Logsdon Jr., and W. Ford Doolittle

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

Microsporidia are obligate intracellular parasites which infect a broad range of animal hosts, including insects, fish, and mammals. Outside a host cell, microsporidia exist as chitinous spores, and infection involves the eversion of an otherwise tightly wound organelle called the polar tubule. Upon piercing the host cell's membrane, the polar tubule provides a means for the spore contents (or sporoplasm) to be injected into the host cytoplasm, where subsequent division and growth of the parasite occur. Microsporidia are unusual among eukaryotes in possessing very small genomes, along with 70S rather than 80S ribosomes, and in lacking mitochondria and peroxisomes. In addition, the 5.8S and 28S rRNA are fused in microsporidia, as they are in prokaryotes (Vossbrinck and Woese 1986). The absence of many typical eukaryotic features has been cited as evidence that the microsporidia exemplify a primitive eukaryotic condition. In fact, microsporidia have been classified as Archezoa, those protists which diverged from the rest of eukaryotes prior to mitochondrial acquisition (Cavalier-Smith 1983). Consistent with this, phylogenetic trees inferred from small-subunit ribosomal RNA, EF-1 α , and EF-2 sequences have supported a deep divergence of the microsporidia (Vossbrinck et al. 1987; Kamaishi et al. 1996a, 1996b). (Recent reviews on the subjects of microsporidia and their origin are given in Müller [1997], Embley and Hirt [1998], Keeling [1998], and Keeling and McFadden [1998]).

More recently, however, phylogenetic analyses have offered a striking alternative for the origin and phylogenetic placement of the microsporidia. Phylogenies of α - and β -tubulins (Edlind et al. 1996; Keeling and Doolittle 1996) were the first to show a specific relationship between the microsporidia and the fungi—a conclusion later supported by (mitochondrial) HSP70 (Germot, Philippe, and Le Guyader 1997; Hirt et al. 1997; Peyretailade et al. 1998b) and recent analyses of the largest subunit of RNA polymerase II (RPB1) (Hirt et al. 1999). Finally, careful analyses of large-subunit (LSU) rRNA support a more recent (“crown”) origin of the microsporidia, albeit not specifically linked with any particular group of eukaryotes (Peyretailade et al. 1998a). In addition, features shared between fungi and microsporidia have been raised as evidence supporting the sisterhood of fungi and microsporidia. These include

an insertion within the EF-1 α gene (whose length and position are conserved), separate thymidilate synthase and dihydrofolate reductase genes (otherwise fused in plants and *Plasmodium*), and certain apparently common characteristics of meiosis (see Keeling and McFadden 1998 and references therein). Due to the conflicting phylogenetic evidence regarding the origin of the microsporidia, we chose to examine another microsporidian protein-coding gene.

TBP is a universal transcription factor, involved in transcription initiation with all three eukaryotic RNA polymerases. Its role has been best characterized for pol II transcription, for which it is the primary subunit of the multimeric transcription factor TFIID. TFIID recognizes the upstream TATA element and promotes DNA bending, which in turn allows other factors of the preinitiation complex to bind (Burley and Roeder 1996). With the exception of the N-terminal region of TBP (which varies greatly in sequence and length) the sequence of the protein is highly conserved; such conservation allowed us to use degenerate PCR to amplify and clone a segment of the *Nosema locustae* TBP gene. This fragment was used as a probe to screen a genomic library (Fast et al. 1998), from which we determined the full-length gene sequence. The probe was also hybridized to a genomic Southern blot to confirm that the TBP gene is, in fact, from *N. locustae* and to determine that the gene is single-copy within the genome (data not shown).

The predicted amino acid sequence of the *N. locustae* TBP gene (259 amino acids in length) is readily alignable with known plant, animal, fungal, protist, and archaeobacterial homologs. A comprehensive alignment of all available eukaryotic TBP sequences and seven representative archaeobacterial sequences was constructed with the 181 C-terminal, unambiguously aligned amino acid characters.

Maximum-likelihood distances of this alignment were computed with PUZZLE, version 3.1 (Strimmer and von Haeseler 1996), using the JTT matrix, considering invariant sites along with eight variant gamma categories to account for site-to-site rate variation. Tree construction was then carried out using the PHYLIP package (version 3.57; Felsenstein 1993) with both the neighbor-joining and the Fitch-Margoliash algorithms. Both methods gave topologies differing only in the (unsupported) position of the *Dictyostelium* sequence; the Fitch tree is shown in figure 1. Parsimony analyses were also undertaken, and many equally parsimonious trees were recovered—presumably due to the extensive conservation of eukaryotic TBP, which limits the number of phylogenetically informative characters. Regardless, the strict consensus tree of over 500 equally parsimo-

Key words: *Nosema locustae*, microsporidia, TFIID, spliceosomal intron, phylogeny, evolution.

Address for correspondence and reprints: Naomi M. Fast, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7. E-mail: nfast@is2.dal.ca.

Mol. Biol. Evol. 16(10):1415–1419. 1999

© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

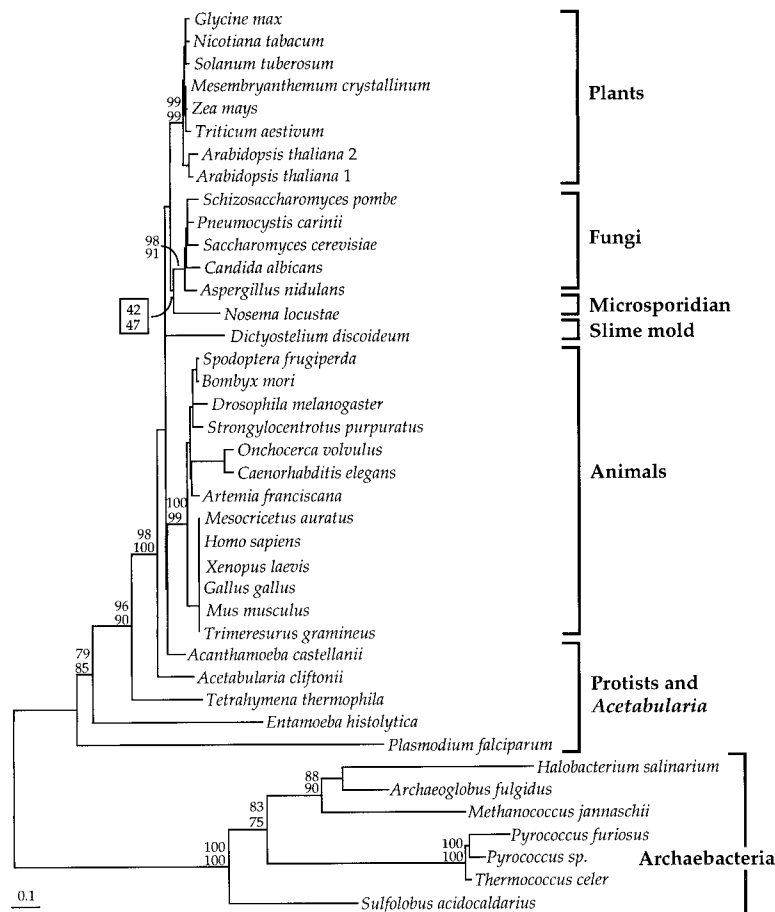


FIG. 1.—Phylogenetic analysis of TBP sequences. The Fitch-Margoliash tree shown was based on maximum-likelihood distances and constructed with global rearrangements and input sequence order jumbled 10 times. Bootstrap values based on 100 resampling replicates and PAM 250 distances calculated with PROTDIST (PHYLIP) are indicated for selected groups, with results from Fitch-Margoliash given above those for neighbor joining.

nious trees and the Fitch tree (fig. 1) agree in their support of major plant, animal, and fungal groups, while leaving other relatively long branched sequences to cluster anomalously at the base of the tree. For example, the single available green algal TBP sequence (*Acetabularia cliftonii*) branches near the base, rather than with the plant sequences. Furthermore, the single ciliate and apicomplexan sequences (*Tetrahymena* and *Plasmodium*) branch basally and separately. Most importantly, though, in all distance and parsimony analyses the *N. locustae* sequence branches together with the fungi. Interestingly, the *N. locustae* TBP sequence is not extremely divergent and possesses a relatively short branch length, a result which contrasts with many other microsporidian sequences.

We tested the significance of the microsporidia-fungi relationship versus a basal, ancient origin of the microsporidia with Kishino-Hasegawa (K-H) tests using PUZZLE, version 4.0 (Strimmer and von Haeseler 1996); we found a basal position for the microsporidia in the TBP tree to be not significantly worse than a fungal relationship (<95% confidence). An exhaustive protein maximum-likelihood (ProtML) search (Adachi and Hasegawa 1996) was also carried out by constrain-

ing as polytomies the well-defined plant, animal, and fungal groups evident in the distance and parsimony trees, while allowing *N. locustae* and the taxa possessing apparent long branches (*Acetabularia*, *Dictyostelium*, *Tetrahymena*, *Entamoeba*, and *Plasmodium*) to move freely. In addition, the archaeobacterial taxa were removed from the analysis, reducing the number of movable groups to a computationally manageable level. Although the most likely tree does not include the microsporidia-fungi clade (*Nosema* branches with *Dictyostelium*, *Acanthamoeba*, and animals), many of the trees not significantly worse than the best tree (based on K-H values) do, in fact, group the microsporidia and fungi together. Thus, the TreeCons program (Jermini et al. 1997) was used to rank the trees based on their likelihoods and standard errors (K-H values) and weight them with an exponential distribution, allowing construction of a consensus tree using this weighting scheme. In this tree, the microsporidian TBP branches with the fungal grouped at all alpha parameters tested (0.05–0.001), indicating that the microsporidia-fungi relationship was present in a majority of those trees which were not significantly worse than the best.

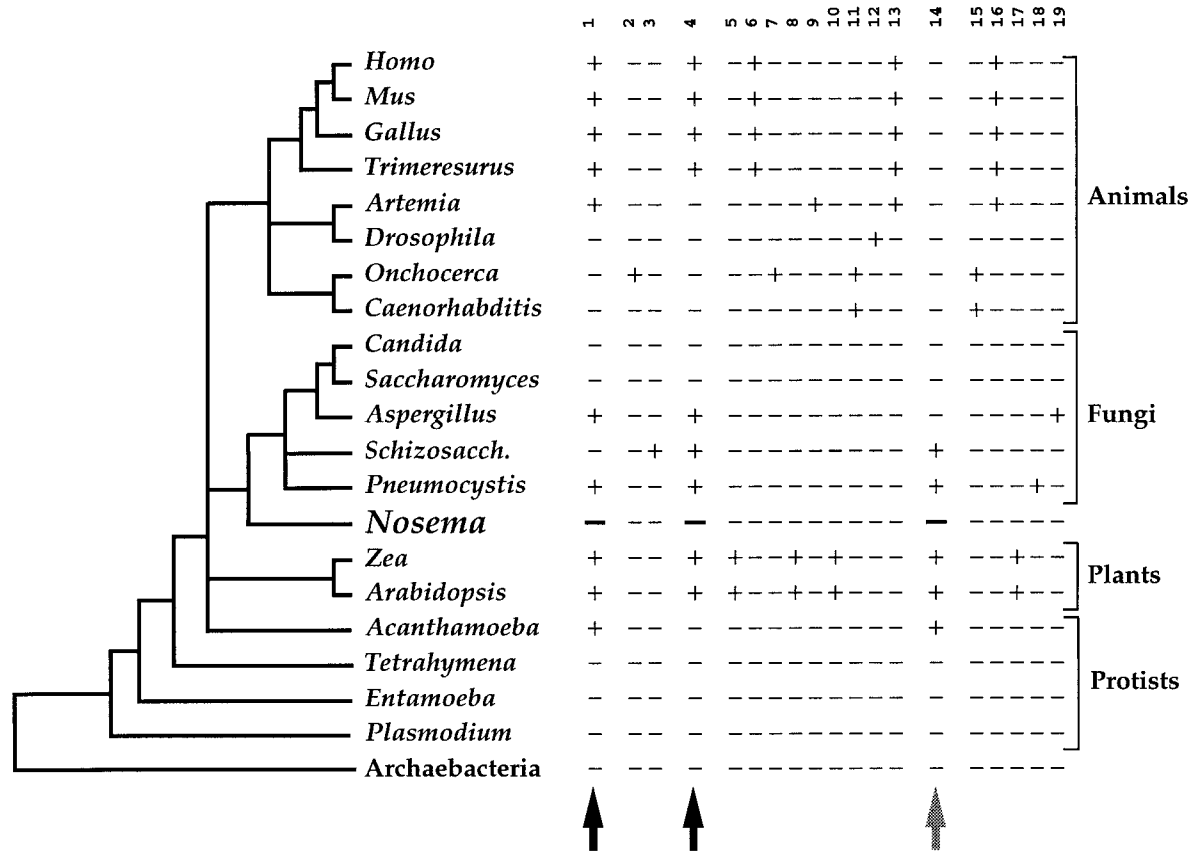


FIG. 2.—Phylogenetic distribution of introns among currently known TBP genes. The 19 intron positions present in the alignable portion of the TBP gene are numbered consecutively from 5' to 3'; the first introns of vertebrate and plant TBP genes fall outside of the alignment region; therefore, intron position 1 corresponds to the second intron in these genes. The cladogram of relationships is based mainly on the results from TBP itself (fig. 1), with resolution shown in the fungal lineages based on rRNA comparisons (Bruns et al. 1992). Major groups of eukaryotes are indicated to the right. Presence of an intron is indicated by “+” and absence by “-”. Potentially “old” introns are indicated by arrows: intron positions shared by animals, fungi, and plants (black), and the position shared by fungi, plants, and *Acanthamoeba* (gray). All sequences are available from GenBank.

It is noteworthy that the microsporidian TBP sequence suggests a recent origin for the microsporidia in all analyses, with a weak but consistent fungal affinity. Although the strength of the microsporidia–fungi relationship indicated here is not nearly as strong as the support seen in analyses with tubulins (Edlind et al. 1996; Keeling and Doolittle 1996) and RPB1 (Hirt et al. 1999), the support level is similar to that shown with the phylogenetic analyses of the microsporidian HSP 70 (Germot, Philippe, and Le Guyader 1997; Hirt et al. 1997). Therefore, TBP adds to a growing list of gene sequences that contradicts an ancient origin of the microsporidia and, instead, favors a fungal ancestry.

In addition to phylogenetic information, we also addressed questions of spliceosomal intron evolution in the microsporidia using TBP. Approximately 25 protein-coding genes have been characterized from microsporidia; among these, only recently has a single putative spliceosomal intron been found (Biderre, Méténier, and Vivarès 1998). An analysis of the intron positions in TBP genes of diverse taxa reveals that there are two introns found at identical positions in plants, animals, and fungi (fig. 2). There are very few introns conserved

at this phylogenetic depth—these “old” introns (defined in Logsdon et al. 1995) likely predate the animal/fungus/plant divergence—contrasting with all other introns in TBP and the vast majority of other genes (Palmer and Logsdon 1991). The presence of two such “old” introns in the TBP gene (as well as another likely “old” intron shared between plants, fungi, and *Acanthamoeba*) made TBP a good candidate to possess an intron in microsporidia; this was, in fact, one reason why we chose to study it. However, the full-length coding sequence of the *N. locustae* TBP gene contains no introns. If microsporidia are indeed related to fungi, as suggested here and by others, the absence of introns in the *N. locustae* TBP gene indicates that some introns were lost (rather than never gained). Indeed, all three “old” introns in TBP have also been lost at least once in other lineages. Furthermore, the presence of these three introns in some fungi indicates that intron loss in microsporidia has occurred separate from other fungal lineages (that is, unless microsporidia are specifically related to ascomycete yeasts like *Saccharomyces cerevisiae*). While it is possible that high titers of spliceosomal introns were never present in microsporidian ancestors, these analyses in-

dicates that some intron loss has certainly occurred. Such a loss of noncoding elements may be one way to reduce genome size (Keeling and McFadden 1998), concomitant with the adoption of a parasitic lifestyle. Genome size reduction is apparent in the tiny sizes of microsporidian genomes. *Nosema locustae* possesses a genome of 5.4 Mb (Streett 1994), which is very small by typical eukaryotic standards, but the reduction in genomic size of other microsporidia can be even more extreme. Indeed, the *Encephalitozoon cuniculi* genome, at 2.9 Mb, is smaller than many bacterial genomes (Biderre et al. 1995).

The sequence of the *N. locustae* TBP gene provides additional evidence that microsporidia are related to fungi, in sharp contrast to the contention that microsporidia are among the earliest eukaryotes. Fungal ancestry of the microsporidia significantly affects the interpretation of many of the unusual features they possess. While characteristics such as the apparent lack of mitochondria, peroxisomes, and stacked golgi (and formerly, spliceosomal introns), as well as the possession of tiny genomes, were once thought to embody the ancestral situation, these features should instead be taken to indicate the highly derived nature of the microsporidia.

Acknowledgments

We thank G. McFadden for the use of his laboratory facilities, P. Keeling for assistance with phylogeny, and P. Keeling and A. Roger for comments on the manuscript. N.M.F. is the recipient of postgraduate scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Killam Trust. J.M.L. was supported by postdoctoral fellowships from the NSF/Sloan Foundation and MRC (Canada). W.F.D. is a fellow of the Canadian Institute for Advanced Research. This research was supported by an MRC operating grant awarded to W.F.D. The new sequence has been deposited in GenBank with accession number AF144035. N.M.F. and J.M.L. contributed equally to this work.

LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1996. MOLPHY version 2.2: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**:1–150.
- BIDERRE, C., G. MÉTÉNIER, and C. P. VIVARÈS. 1998. A small spliceosomal-type intron occurs in a ribosomal protein gene of the microsporidian *Encephalitozoon cuniculi*. *Mol. Biochem. Parasitol.* **94**:283–286.
- BIDERRE, C., M. PAGES, G. MÉTÉNIER, E. U. CANNING, and C. P. VIVARÈS. 1995. Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. *Mol. Biochem. Parasitol.* **74**:229–231.
- BRUNS, T. D., R. VILGALYS, S. M. BARNES et al. (11 co-authors). 1992. Evolutionary relationships within the fungi: analyses of nuclear small subunit rRNA sequences. *Mol. Phylogenet. Evol.* **1**:231–241.
- BURLEY, S. K., and R. G. ROEDER. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**:769–799.
- CAVALIER-SMITH, T. 1983. A six-kingdom classification and a unified phylogeny. Pp. 1027–1034 in H. E. A. SCHENK and W. S. SCHWEMMLER, eds. *Endocytobiology II*. Walter de Gruyter, Berlin.
- EDLIND, T. D., J. LI, G. S. VISVESVARA, M. H. VODKIN, G. L. McLAUGHLIN, and S. K. KATIYAR. 1996. Phylogenetic analysis of beta-tubulin sequences from mitochondrial protozoa. *Mol. Phylogenet. Evol.* **5**:359–367.
- EMBLEY, T. M., and R. P. HIRT. 1998. Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**:624–629.
- FAST, N. M., A. J. ROGER, C. A. RICHARDSON, and W. F. DOOLITTLE. 1998. U2 and U6 snRNA genes in the microsporidian *Nosema locustae*: evidence for a functional spliceosome. *Nucleic Acids Res.* **26**:3202–3207.
- FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.
- GERMOT, A., H. PHILIPPE, and H. LE GUYADER. 1997. Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol. Biochem. Parasitol.* **87**:159–168.
- HIRT, R. P., B. HEALY, C. R. VOSSBRINCK, E. U. CANNING, and T. M. EMBLEY. 1997. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr. Biol.* **7**:995–998.
- HIRT, R. P., J. M. LOGSDON JR., B. HEALY, M. W. DOREY, W. F. DOOLITTLE, and T. M. EMBLEY. 1999. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci. USA* **96**:580–585.
- JERMIN, L. S., G. J. OLSEN, K. L. MENGERSEN, and S. EASTEAL. 1997. Majority-rule consensus of phylogenetic trees obtained by maximum-likelihood analysis. *Mol. Biol. Evol.* **14**:1296–1302.
- KAMAISHI, T., T. HASHIMOTO, Y. NAKAMURA, Y. MASUDA, F. NAKAMURA, K. OKAMOTO, M. SHIMIZU, and M. HASEGAWA. 1996a. Complete nucleotide sequences of the genes encoding translation elongation factors 1 alpha and 2 from a microsporidian parasite, *Glugea plecoglossi*: implications for the deepest branching of eukaryotes. *J. Biochem. (Tokyo)* **120**:1095–1103.
- KAMAISHI, T., T. HASHIMOTO, Y. NAKAMURA, F. NAKAMURA, S. MURATA, N. OKADA, K. OKAMOTO, M. SHIMIZU, and M. HASEGAWA. 1996b. Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. *J. Mol. Evol.* **42**:257–263.
- KEELING, P. J. 1998. A kingdom's progress: Archezoa and the origin of eukaryotes. *BioEssays* **20**:87–95.
- KEELING, P. J., and W. F. DOOLITTLE. 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.* **13**:1297–1305.
- KEELING, P. J., and G. I. McFADDEN. 1998. Origins of microsporidia. *Trends Microbiol.* **6**:19–23.
- LOGSDON, J. M. JR., M. G. TYSHENKO, C. DIXON, J. D. JAFARI, V. K. WALKER, and J. D. PALMER. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. USA* **92**:8507–8511.
- MÜLLER, M. 1997. What are the microsporidia? *Parasitol. Today* **13**:455–456.
- PALMER, J. D., and J. M. LOGSDON JR. 1991. The recent origins of introns. *Curr. Opin. Genet. Dev.* **1**:470–477.
- PEYRETAILLADE, E., C. BIDERRE, P. PEYRET, F. DUFFIEUX, G. MÉTÉNIER, M. GOUY, B. MICHOT, and C. P. VIVARÈS. 1998a. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. *Nucleic Acids Res.* **26**:3513–3520.

- PEYRETAILLADE, E., V. BROUSSOLLE, P. PEYRET, G. MÉTÉNIER, M. GOUY, and C. P. VIVARÈS. 1998*b*. Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. *Mol. Biol. Evol.* **15**:683–689.
- STRETT, D. A. 1994. Analysis of *Nosema locustae* (Microsporidia: Nosematidae) chromosomal DNA with pulsed-field-gel-electrophoresis. *J. Invertebr. Pathol.* **63**:301–303.
- STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**:964–969.
- VOSSBRINCK, C. R., J. V. MADDOX, S. FRIEDMAN, B. A. DEBRUNNER-VOSSBRINCK, and C. R. WOESE. 1987. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**:411–414.
- VOSSBRINCK, C. R., and C. R. WOESE. 1986. Eukaryotic ribosomes that lack a 5.8S RNA. *Nature* **320**:287–288.

MANOLO GOUY, reviewing editor

Accepted June 16, 1999