Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 31 (2004) 572–580
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Molecular phylogenies of Parabasalia inferred from four protein genes and comparison with rRNA trees

Delphine Gerbod,[a,b] Emily Sanders,[c] Shigeharu Moriya,[b,d] Christophe Noël,[a] Hirotoshi Takasu,[d] Naomi M. Fast,[e] Pilar Delgado-Viscogliosi,[f] Moriya Ohkuma,[b] Toshiaki Kudo,[b] Monique Capron,[a] Jeffrey D. Palmer,[c] Patrick J. Keeling,[e] and Eric Viscogliosi[a,*]

[a] *Institut Pasteur, Unité INSERM U547, 1 Rue du Professeur Calmette, B. P. 245, 59019 Lille Cedex, France*
[b] *Laboratory of Microbiology, RIKEN (The Institute of Physical and Chemical Research) and Japan Science and Technology Corporation (JST), Wako, Saitama 351-0198, Japan*
[c] *Department of Biology, Indiana University, Bloomington, IN 47405, USA*
[d] *Laboratory of Environmental Molecular Biology, Graduate School of Yokohama City University, Suehiro-cho, Tsurumi-ward, Kanagawa 230-0045, Japan*
[e] *Department of Botany, Canadian Institute for Advanced Research, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4*
[f] *Département Eaux et Environnement, Institut Pasteur, Laboratoire de Recherches Appliquées, 1 Rue du Professeur Calmette, B.P. 245, 59019 Lille Cedex, France*

Received 1 May 2003; revised 1 September 2003

Abstract

The molecular phylogeny of parabasalids has mainly been inferred from small subunit (SSU) rRNA sequences and has conflicted substantially with systematics based on morphological and ultrastructural characters. This raises the important question, how congruent are protein and SSU rRNA trees? New sequences from seven diverse parabasalids (six trichomonads and one hypermastigid) were added to data sets of glyceraldehyde-3-phosphate dehydrogenase (GAPDH), enolase, α-tubulin and β-tubulin and used to construct phylogenetic trees. The GAPDH tree was well resolved and identical in topology to the SSU rRNA tree. This both validates the rRNA tree and suggests that GAPDH should be a valuable tool in further phylogenetic studies of parabasalids. In particular, the GAPDH tree confirmed the polyphyly of Monocercomonadidae and Trichomonadidae and the basal position of *Trichonympha agilis* among parabasalids. Moreover, GAPDH strengthened the hypothesis of secondary loss of cytoskeletal structures in Monocercomonadidae such as *Monocercomonas* and *Hypotrichomonas*. In contrast to GAPDH, the enolase and both tubulin trees are poorly resolved and rather uninformative about parabasalian phylogeny, although two of these trees also identify *T. agilis* as representing the basal-most lineage of parabasalids. Although all four protein genes show multiple gene duplications (for 3–6 of the seven taxa examined), most duplications appear to be relatively recent (i.e., species-specific) and not a problem for phylogeny reconstruction. Only for enolase are there more ancient duplications that may confound phylogenetic interpretation.

Keywords: Evolution; Hypermastigid; Molecular phylogeny; Parabasalia; Protein coding gene; Small subunit rRNA gene sequence; Trichomonad; Taxonomy

1. Introduction

The Parabasalia, or commonly parabasalids, are anaerobic flagellated protists typified by the presence of hydrogenosomes, a characteristic mastigont including a well-developed flagellar apparatus, and one or more parabasal apparatus consisting of a parabasal body (Golgi complex) and a parabasal filament (Brugerolle, 1976; Honigberg, 1963; Honigberg and Brugerolle, 1990). The current taxonomy of Parabasalia is based on morphological/ultrastructural characters, mostly linked

* Corresponding author. Fax: +33-3-20-87-78-88.
E-mail address: eric.viscogliosi@pasteur-lille.fr (E. Viscogliosi).

1055-7903/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2003.09.013

to the structure and development of the cytoskeleton, which can range from rudimentary to very complex. More than 80 genera and 400 parabasalid species were identified as early as the 1970s (Yamin, 1979). These were separated into two classes (Cavalier-Smith, 2002): the Trichomonadea (or commonly trichomonads) and the Hypermastigea (or hypermastigids). The Trichomonadea have been subdivided into four main families: Monocercomonadidae, Trichomonadidae, Devescovinidae, and Calonymphidae (Brugerolle, 1976; Honigberg, 1963). The Monocercomonadidae, which includes species exhibiting a rudimentary cytoskeleton, have been thought to occupy a basal position in the parabasalid tree. The other trichomonad families, with their more complex cytoskeleton, were thought to be derived from monocercomonad ancestors. Finally, the hypermastigids, which are characterized by hyperdevelopment of the cytoskeleton and multiplication of flagella, were positioned at the apex of the parabasalid tree, evolving from the simpler types. Thus, the presumed evolution of the parabasalids reflected the traditional view regarding polarization of cytoskeletal development from simple to complex.

In recent years, phylogenetic studies of parabasalids have begun to use molecular sequence data, primarily small subunit (SSU) rRNA gene sequences (for review see Gerbod et al., 2002; Keeling, 2002; Keeling et al., 1998; Viscogliosi et al., 1999). The rRNA analyses have provided new insights into the evolution of this protist group and have often conflicted with phylogenetic hypotheses based on morphology (Brugerolle, 1976; Honigberg, 1963). A pair of logically related major conflicts involves the Monocercomonadidae and certain hypermastigids. Whereas morphology suggests that the Monocercomonadidae is monophyletic and basal among parabasalids, rRNA suggests that members of the group arose at least three times separately from organisms with a complex cytoskeleton. If rRNA trees are correct on this issue, then the Monocercomonadidae have lost cytoskeletal structures during evolution and their apparent simplicity is misleading both as a primitive character and as a significant taxonomic criterion. Conversely, certain morphologically complex and putatively phylogenetically derived hypermastigids such as *Trichonympha* represent the earliest lineage of parabasalids in rRNA trees (Dacks and Redfield, 1998; Delgado-Viscogliosi et al., 2000; Gunderson et al., 1995; Keeling et al., 1998; Silberman et al., 1996), thus also conflicting with the simple-to-complex polarisation of morphological evolution proposed in traditional systematics of parabasalids.

Comparative phylogenetic studies have demonstrated that single gene phylogenies based on either RNA or protein can, depending on the group in question, be very misleading. It is therefore critical to compare and assess the parabasalid SSU rRNA tree with those of multiple protein genes. Certain proteins have already been used as phylogenetic markers for parabasalids, albeit with limited taxonomic sampling. Iron-containing superoxide dismutase (FeSOD), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and class II fumarase have all been used to infer relationships within a restricted sampling of trichomonad species (Gerbod et al., 2001a; Viscogliosi and Müller, 1998; Viscogliosi et al., 1996), but no gene other than SSU rRNA has been sampled from a wide variety of parabasalids. This is especially true with respect to hypermastigids, which are found exclusively in the guts of insects, making them difficult to include in molecular studies. However, it is now feasible to obtain hypermastigid rRNA sequences from the whole-gut content of termites by polymerase chain reaction (PCR) approaches and to then assign the sequences to the corresponding species by whole-cell in situ hybridization using sequence-specific probes (Gerbod et al., 2002; Moriya et al., 2001; Ohkuma et al., 1998, 2000).

In this work, we have extended previous studies in which GAPDH (Markos et al., 1993; Viscogliosi and Müller, 1998), enolase (Keeling and Palmer, 2000), α-tubulin (Keeling and Doolittle, 1996a; Moriya et al., 2001; Noël et al., 2001), and β-tubulin (Katiyar and Edlind, 1994) gene sequences were obtained from a few parabasalids, with different sampling for each gene. We have now sequenced these genes from each of seven diverse parabasalids representing all four major clades as defined by SSU rRNA (Gerbod et al., 2002). For the first time in parabasalids, this allows phylogenetic trees with the same species sampling to be inferred from several molecular markers, as well as comparison of protein- and rRNA-based trees.

2. Materials and methods

2.1. Origin, cultivation and DNA extraction of trichomonads

The origins of strains were as follows: *Trichomonas vaginalis* strain NIH-C1 (ATCC 30001) from *Homo sapiens*; *Tetratrichomonas gallinarum* strain A6 from *Anas platyrhinchos*; *Tritrichomonas foetus* strain KV1 (ATCC 30924) from *Bos primigenus*; *Monocercomonas* sp. strain NS-1PRR (ATCC 50210) from *Natrix sipedon*; *Trichomitus batrachorum* strain G11 (ATCC 30066) from *Elaphe obsoleta*; *Hypotrichomonas acosta* strain 3L3 (ATCC 30069) from *Drymarchon corais couper*. All isolates used were grown axenically at 37 or 27 °C in trypticase–yeast extract–maltose (TYM) medium (Diamond, 1957) without agar supplemented with 10% (v/v) heat-inactivated horse serum (Gibco-BRL), 100 U/ml of penicillin G, and 50 μg/ml of streptomycin sulfate.

Genomic DNA was isolated as described (Riley and Krieger, 1992).

2.2. Protein gene sequences of trichomonads

A GAPDH gene from *H. acosta* was amplified by PCR from genomic DNA using degenerate primers and conditions described previously (Viscogliosi and Müller, 1998). Other GAPDH sequences used in this study have been previously obtained by Markos et al. (1993) and Viscogliosi and Müller (1998). Database accession numbers of these sequences are: *T. vaginalis gap1* and *gap2*, L11394 and AF022414, respectively; *T. gallinarum gap*, AF022419; *T. foetus gap1* and *gap2*, AF022415 and AF022416, respectively; *Monocercomonas* sp. *gap*, AF022420; *T. batrachorum gap1* and *gap2*, AF022417 and AF022418, respectively.

Enolase genes from *H. acosta*, *T. foetus*, and *T. gallinarum* were amplified by PCR using primers and conditions described previously (Keeling and Palmer, 2000). Other enolase sequences were reported previously (Keeling and Palmer, 2000). Database accession numbers of these sequences are: *Monocercomonas* sp. *eno1* and *eno2*, AF159523 and AF159522, respectively; *T. vaginalis eno1* to *eno5*, AF159524–AF159528; *T. batrachorum eno1* and *eno2*, AF159521 and AF159520, respectively.

α-Tubulin-coding genes from *T. gallinarum*, *H. acosta*, and *T. foetus* were amplified by PCR using primers and conditions described previously (Keeling and Doolittle, 1996b). Other α-tubulin sequences from trichomonads were reported previously (Keeling and Doolittle, 1996b; Noël et al., 2001). Database accession numbers of these sequences are: *Monocercomonas* sp. *alpha1* and *alpha2*, U66902 and U66903, respectively; *T. batrachorum alpha*, U66904; *T. vaginalis alpha1* and *alpha2*, AF327847 and AF327848, respectively.

β-Tubulin-coding regions from *T. foetus*, *T. batrachorum*, *Monocercomonas* sp., *H. acosta*, and *T. gallinarum* were amplified using the primers GCCTGCAGGN CARTGYGGNAAYCA and GTCCATGCCYTCRTT NACYTACCAYTG. Amplifications consisted of 35 cycles of 1 min at 94 °C, 2 min at 50 °C and 3 min at 72 °C. The β-tubulin sequence from *T. vaginalis* was reported previously (Katiyar and Edlind, 1994), and is under the Accession No. L05468.

PCR products were separated by agarose electrophoresis, purified, and cloned in the T-vector, pCR 2.1-TOPO (Invitrogen). Clones containing inserts of approximately the expected size were picked and sequenced. From each species, several clones were sequenced on both strands by primer walking using ABI dye-terminator chemistry. Trichomonad sequences obtained in this study have been deposited in GenBank under the Accession Nos. AY272770–AY272775 and AY272777–AY272791.

2.3. Construction of a cDNA library from symbionts of Reticulitermes speratus and assignment of protein sequences from Trichonympha agilis

The lower termite *Reticulitermes speratus* (order Isoptera, family Rhinotermitidae) was collected in the vicinity of Ogose, Saitama prefecture, Japan, and maintained on a sterile artificial diet (cellulose powder:pine wood chips:water = 1:1:3) in 6-cm plastic dishes at room temperature for 1 to 2 weeks. Briefly, a cDNA library was constructed by the biotinylated CAP trapper method (Carninci and Hayashizaki, 1999) from the mixed protist population found in the gut of *R. speratus*. The sequencing project of this cDNA library is completed, and the analysis of these data is now in progress (Moriya et al., in preparation).

Full length sequences coding for GAPDH, enolase, and β-tubulin were obtained from this library. Using BLASTP, we identified some of these clones that shared significant similarity with those of parabasalids reported so far. To confirm the origin of these GAPDH and enolase sequences, in situ hybridization was performed on *R. speratus* gut biota. The hindgut content was fixed and incubated as previously described (Moriya et al., 1998) with some minor modifications. For fluorescence in situ hybridization (FISH) and enzymatic amplified immunohybridization studies, we used two different oligonucleotides 5′-labeled with FITC. Probe A (5′-AACCTTCTCT GAAAGGTTTTG-3′) is complementary to a region of the specific GAPDH sequence of clone 2038A-85, whereas probe B (5′-GACAATATTTAACATCATCA A-3′) is complementary to a region of the specific enolase sequence of clone 2030A-07. The SSU rRNA probe (5′-GGRCATCACRGACCTGTTAT-3′) used as a positive control hybridizes to most eukaryotes (Amann et al., 1995) and recognizes sequences within helix E-38 according to the *Saccharomyces cerevisiae* SSU rRNA secondary structure (Neefs et al., 1993). The hybridization solution (100 pmol/ml of FITC-labeled probe oligonucleotide) was applied onto the specimen, sealed in a cover well chamber, and incubated at 37 °C for 6 h. The specimen was immunologically stained with alkaline phosphatase-conjugated anti-FITC antibodies (1/250 dilution: Boehringer–Mannheim) for 1 h at 37 °C. Signals were detected using the NBT/BCIP system (Boehringer–Mannheim). For signal detection, the specimen was washed with water and mounted in 50% glycerol with 0.5% triethylenediamine and observed by light microscopy.

The origin of two closely related β-tubulin sequences obtained from the cDNA library made from symbionts of *R. speratus* was confirmed by PCR. Sense (Betub-para-200F: 5′-AAAGTGCAGATGAAGTTTT C-3′) and antisense (Betub-para-410R: 5′-CTTCAT CAAAT TCAACTGTT-3′) primers were designed from the sequences of both cDNA clones. The hypermastigids *T. agilis* and *Teranympha mirabilis* and the oxymonads

Pyrsonympha grandis and *Dinenympha exilis* were collected by micropipetting as already described (Gerbod et al., 2000) from the hindgut of *R. speratus*. cDNAs were prepared from each sample of micro-manipulated cells according to Moriya et al. (2001). PCR was done using each of these cDNA samples as template. Amplifications consisted of 30 cycles of 20 s at 94 °C, 30 s at 55 °C, and 1 min at 72 °C. The resultant PCR products were cloned into a pCR 2.1-TOPO vector and sequenced. Full length GAPDH, enolase, and β-tubulin sequences from *T. agilis* obtained in this study have been deposited in GenBank under Accession Nos. AB107786–AB107789. α-Tubulin sequences were already available for *T. agilis* under Accession Nos. AF230347 and AF230348.

2.4. Phylogenetic analyses

Amino acid sequences from parabasalids and other organisms were aligned with the use of the BioEdit v5.0.9 package (http://www.mbio.ncsu.edu:BioEdit/bioedit.html). Alignment was facilitated by the fact that, like all previously characterized genes from parabasalids, none of these genes appear to contain any introns. Unambiguously alignable sites were chosen for phylogenetic inference (296, 306, 377, and 372 sites from GAPDH, enolase, α-tubulin, and β-tubulin, respectively). Full-length alignments and sites used in analyses are available upon request to the corresponding author. Distance matrices were calculated with TREE-PUZZLE 5.0 (Strimmer and Von Haeseler, 1996) using the WAG substitution matrix and site-to-site rate variation modelled on a gamma distribution with invariable sites and eight rate categories, and the shape parameter α was estimated from the data. Distance trees were constructed using weighted neighbour-joining using WEIGHBOR (Bruno et al., 2000). Bootstrap resampling (100 replicates) was carried out using the shell script PUZZLEBOOT (by Holder and Roger: http://www.tree-puzzle.de). A protein maximum likelihood analysis was also conducted using ProML (Felsenstein, 1993). Site-to-site variation was corrected using the R option by entering the nine categories estimated by TREE-PUZZLE and their respective frequencies (eight rates plus invariable sites). Protein maximum-likelihood bootstrapping (100 replicates) was performed as above, with the rates and rate categories from the original dataset enforced on each replicate. Phylogenetic analyses were also carried out using MrBAYES 2.01 (Huelsenbeck and Ronquist, 2001). Bayesian analyses were performed using the JTT amino acid replacement model (Jones et al., 1992). In all Bayesian analyses, starting trees were random, four simultaneous Markov chains were run for 500,000 generations, burn-in values were set at 30,000 generations (based on empirical values of stabilizing likehoods), and trees were sampled every 100 generations. Bayesian posterior probabilities were calculated using a Markov chain Monte Carlo (MCMC) sampling approach (Green, 1995) implemented in MrBAYES 2.01. All analyses were carried out both including and excluding outgroups to determine if this was substantially affecting relationships within the Parabasalia.

3. Results and discussion

3.1. Taxon sampling based on rRNA phylogeny

The most recent large-scale phylogenetic analysis of parabasalids included all known SSU rRNA gene sequences from more than 40 parabasalid taxa (Gerbod et al., 2002). This analysis showed that all trichomonad families and both classes of parabasalids (trichomonads and hypermastigids) formed polyphyletic groups, suggesting that a revision of the traditional systematic scheme is necessary. Maximum likelihood analysis of SSU rRNA identified four main clades (Fig. 1) that were consistent with those described in previous phylogenetic analyses (Delgado-Viscogliosi et al., 2000; Gerbod et al., 2000, 2001b; Keeling et al., 1998; Keeling, 2002). Clade 1 arises basally and corresponds to the hypermastigid suborder Trichonymphina, represented in this study by *T. agilis*. Next-basal is clade 2, consisting of certain Trichomonadidae (here represented by *T. batrachorum*) and certain Monocercomonadidae (*H. acosta*). Most parabasalids belong to clades 3 and 4. Clade 3 is a very heterogeneous group composed of certain Trichomonadidae (represented by *T. foetus*), certain Monocer-

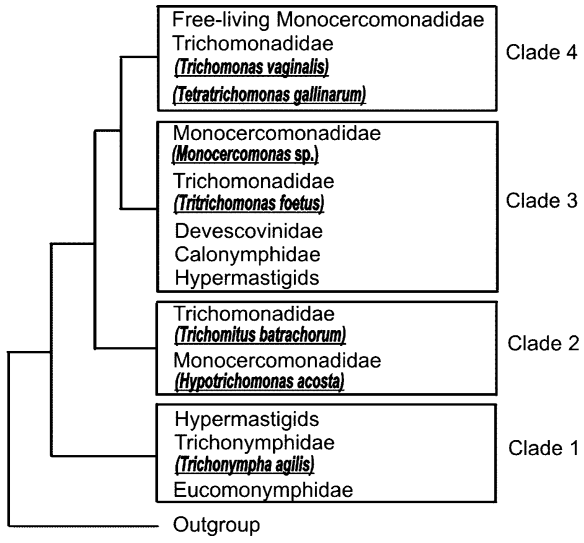


Fig. 1. Overview of phylogenetic relationships among parabasalids based on SSU rRNA analyses (Gerbod et al., 2002). The parabasalid species examined in this study are in boldface and in brackets. Note that this species sampling represents the four main clades identified from SSU rRNA based-trees.

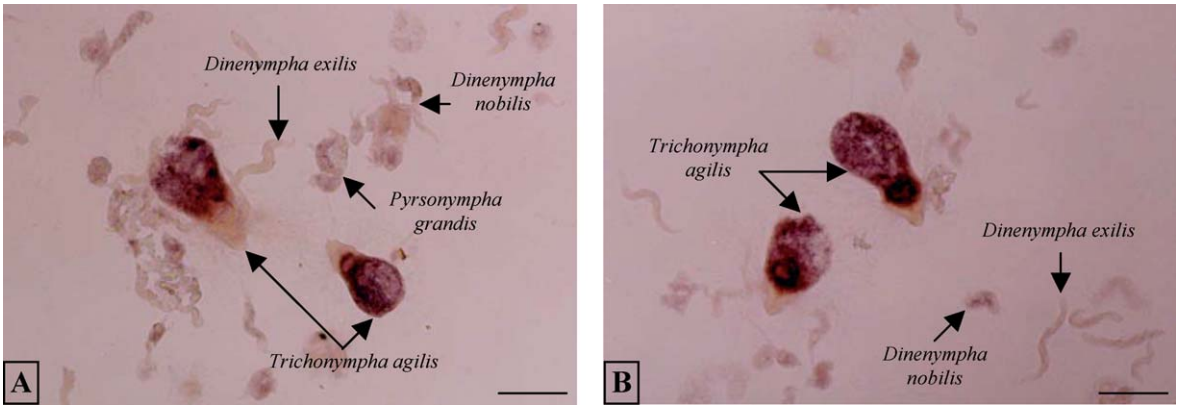


Fig. 2. Verification of the origin of two clones obtained from the cDNA library made from symbionts of *R. speratus*. In situ hybridization was performed using probe A specific for the GAPDH sequence of the cDNA clone 2038A-85 (A) and probe B specific for the enolase sequence of the cDNA clone 2030A-07 (B). Shown are phase-contrast micrographs of termite intestinal symbiotic protists stained with alkaline phosphatase-conjugated anti-FITC antibodies. Note that both probes hybridized only with *T. agilis* cells. Scale bars = 100 μm.

comonadidae (*Monocercomonas* sp.), all examined Devescovinidae and Calonymphidae, and certain hypermastigids. Clade 4 comprises all free-living Monocercomonadidae and nearly all species belonging to one of two main Trichomonadidae subfamilies (Trichomonadinae; represented by *T. vaginalis* and *T. gallinarum*). The taxa chosen for this study represent the phylogenetic diversity of parabasalids according to rRNA phylogeny about as well as is possible with seven taxa, i.e., all four major rRNA-based clades are represented, three of them by two taxa each.

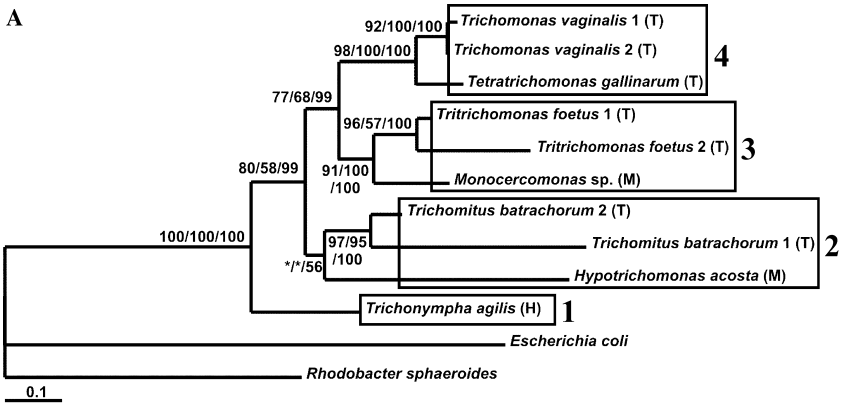
3.2. Assignment of protein genes from T. agilis

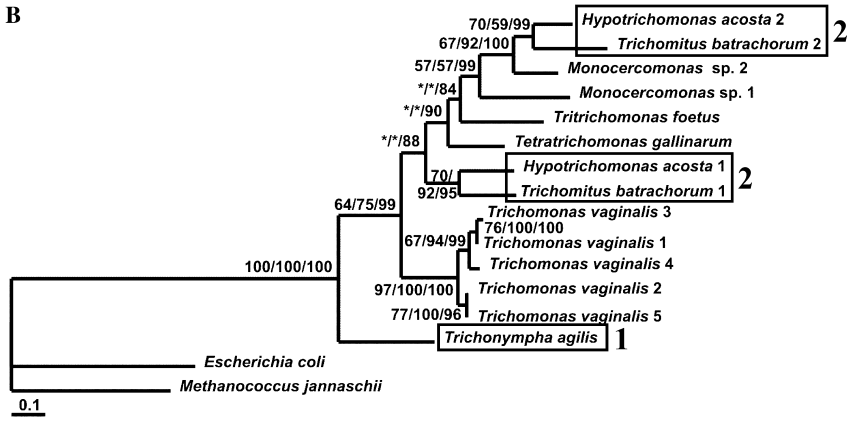
The termite *Reticulitermes speratus* harbors in its hindgut at least 13 flagellates (Koidzumi, 1921; Yamin, 1979): nine oxymonad species belonging to the genera *Dinenympha* and *Pyrsonympha* and four species of hypermastigids (*Holomastigotes elongatum*, *Spironympha porteri*, *T. mirabilis*, and *T. agilis*). All of these microorganisms were found in our cell suspensions (data not shown). In particular, the hypermastigid *T. agilis* was found to be a relatively dominant member of the microfauna in the hindgut of *R. speratus* and is easily recognizable on the basis of its large size and morphological features (Grassé, 1952; Hollande and Carruette Valentin, 1971; Kirby, 1932, 1944; Koidzumi, 1921). *T. agilis* is approximately 150–200 μm long and 150–180 μm in width at its widest point and exhibits a typical "pear-shaped" cell morphology (Figs. 2A and B). The body of this hypermastigid is subdivided into three regions: the rostrum at the anterior part of the cell; the flagella-bearing region of the body behind the rostrum; and the non-flagellated posterior region. The flagella are exceedingly numerous, estimated at several hundreds.

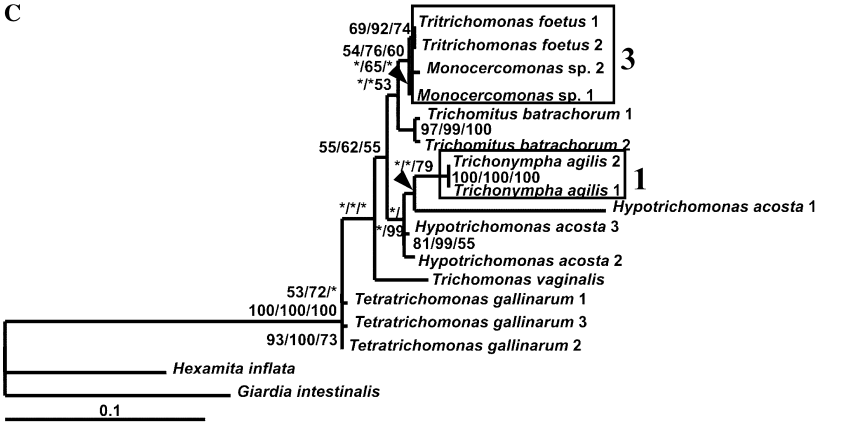
A full length cDNA library was constructed from the mixed protist population found in the gut of *R. speratus*. The deduced GAPDH, enolase, and β-tubulin amino acid sequences of certain clones obtained from this library were highly similar in sequence to those described in other parabasalids. Since four species of hypermastigids can be found in the hindgut of this termite, we applied a whole-cell FISH technique using sequence-specific probes in order to confirm the source of these unassigned cDNA sequences. The positive control for FISH experiments was a Texas-Red labelled probe complementary to all eukaryotic SSU rDNA sequences. Hybridization with this probe resulted in a fluorescence signal for each flagellate (data not shown). The FITC-labeled GAPDH- (Fig. 2A) and enolase-specific (Fig. 2B) probes hybridized strongly and specifically to only those cells with the size and morphology of *T. agilis* (Fig. 2A). These results indicate that the GAPDH and enolase sequences obtained from the *R. speratus* symbiont library were in all probability derived from *T. agilis* and not from other hypermastigid cells.

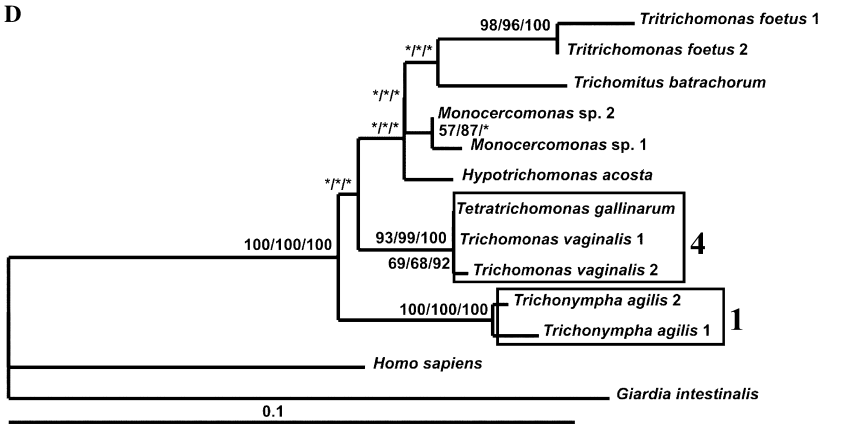
The assignment of two closely related β-tubulin sequences obtained from this cDNA library was not convincing by FISH experiments. Therefore, cDNA was prepared from micro-manipulated cells of *T. agilis* and

Fig. 3. Maximum-likelihood trees based on GAPDH (A), enolase (B), β-tubulin (C), and α-tubulin (D). Choices of outgroups were according to previous phylogenetic analyses (Edlind et al., 1996; Keeling and Doolittle, 1996b; Keeling and Palmer, 2000; Keeling et al., 2000; Moriya et al., 2001; Viscogliosi and Müller, 1998). The systematic assignment of each parabasalid taxon is indicated as established from morphological data: M, Monocercomonadidae; T, Trichomonadidae; H, Hypermastigid. Clades identified from SSU rRNA trees (1–4) and recovered in protein based-trees are boxed. Numbers near the individual nodes indicate bootstrap values and Bayesian posterior probabilities given as percentages by the three different tree reconstruction methods (maximum likelihood/distance/MrBAYES). Asterisks designate nodes with values below 50%. The scale bars indicate 0.1 substitutions (corrected) per site.


A
92/100/100
Trichomonas vaginalis 1 (T)
98/100/100
Trichomonas vaginalis 2 (T)
Tetratrichomonas gallinarum (T)
4
77/68/99
96/57/100
Tritrichomonas foetus 1 (T)
Tritrichomonas foetus 2 (T)
3
80/58/99
91/100
/100
Monocercomonas sp. (M)
Trichomitus batrachorum 2 (T)
100/100/100
97/95
/100
Trichomitus batrachorum 1 (T)
2
//56
Hypotrichomonas acosta (M)
Trichonympha agilis (H)
1
Escherichia coli
Rhodobacter sphaeroides
0.1


B
70/59/99
Hypotrichomonas acosta 2
2
67/92/100
Trichomitus batrachorum 2
57/57/99
Monocercomonas sp. 2
//84
Monocercomonas sp. 1
//90
Tritrichomonas foetus
//88
Tetratrichomonas gallinarum
Hypotrichomonas acosta 1
70/
92/95
2
Trichomitus batrachorum 1
64/75/99
Trichomonas vaginalis 3
76/100/100
67/94/99
Trichomonas vaginalis 1
Trichomonas vaginalis 4
100/100/100
97/100/100
Trichomonas vaginalis 2
77/100/96
Trichomonas vaginalis 5
Trichonympha agilis
1
Escherichia coli
Methanococcus jannaschii
0.1


C
69/92/74
Tritrichomonas foetus 1
54/76/60
Tritrichomonas foetus 2
/65/
Monocercomonas sp. 2
3
*/*53
Monocercomonas sp. 1
Trichomitus batrachorum 1
97/99/100
Trichomitus batrachorum 2
55/62/55
Trichonympha agilis 2
//79
100/100/100
1
Trichonympha agilis 1
//*
Hypotrichomonas acosta 1
*/
*/99
Hypotrichomonas acosta 3
81/99/55
Hypotrichomonas acosta 2
Trichomonas vaginalis
53/72/*
Tetratrichomonas gallinarum 1
100/100/100
Tetratrichomonas gallinarum 3
93/100/73
Tetratrichomonas gallinarum 2
Hexamita inflata
Giardia intestinalis
0.1


D
98/96/100
Tritrichomonas foetus 1
//*
Tritrichomonas foetus 2
Trichomitus batrachorum
//*
Monocercomonas sp. 2
//*
57/87/*
Monocercomonas sp. 1
Hypotrichomonas acosta
//*
Tetratrichomonas gallinarum
93/99/100
Trichomonas vaginalis 1
4
100/100/100
69/68/92
Trichomonas vaginalis 2
Trichonympha agilis 2
100/100/100
1
Trichonympha agilis 1
Homo sapiens
Giardia intestinalis
0.1

also from the hypermastigid *T. mirabilis* and the oxymonads *D. exilis* and *P. grandis* and used for PCR experiments. Using specific primers designed from the sequences of both β-tubulin cDNA clones, we succeeded in amplifying a 620 bp fragment from cDNAs of *Trichonympha* and *Teranympha* (data not shown), whereas no product was obtained from both oxymonad cDNAs. Several independent clones were completely sequenced from both PCR products. The two β-tubulin sequences obtained from the *R. speratus* mixed library were much more similar to those from micro-manipulated *Trichonympha* cells than to those from *Teranympha* cells. This strongly suggests that *T. agilis* was the source of both β-tubulin clones derived from the *R. speratus* library.

3.3. Phylogenies of parabasalids inferred from protein sequences

As stated above, SSU rRNA-based trees have deeply modified the classic/intuitive systematic of parabasalids. However, congruence between independent data sets is the strongest argument in favor of a given topology, so the SSU rRNA tree of Parabasalia needs to be compared with phylogenies based on other molecules. To this end, we sought to examine the relationships among parabasalids using four protein coding genes: GAPDH, enolase, α-tubulin, and β-tubulin.

The maximum likelihood tree of parabasalid GAPDH sequences (Fig. 3A) is identical in topology for the relevant taxa to rRNA trees (Fig. 1, Gerbod et al., 2002), with all but one of the GAPDH branches receiving moderate to high support. In contrast, the enolase tree (Fig. 3B) and both the α- and β-tubulin trees (Figs. 3C and D) are poorly resolved, with very few well-supported groups of taxa (as opposed to a number of well-supported groups of duplicated genes all belonging to the same species; see below). It is important to emphasize that the enolase and tubulin trees are not strongly inconsistent with the well-supported (and identical) GAPDH and rRNA trees; the issue is simply one of lack of resolution. In analyses where outgroups were excluded, no significant deviation from these topologies was noted.

The one relationship from the GAPDH and rRNA trees that is also recovered with at least two of the other three proteins is the placement of the hypermastigid *Trichonympha* as the earliest-diverging parabasalid (this receives moderate support with GAPDH and enolase, and weak support with α-tubulin). This emerging molecular consensus that hypermastigids with a hyperdeveloped cytoskeleton, classically considered to be a late-evolving lineage, instead arose early, possibly basally, in parabasalid evolution suggests that the traditional, simple-to-complex evolutionary polarization of characters in comparative morphological analyses should be reconsidered. However, as suggested by Brugerolle and Patterson (2001), the hypermastigids might have been preceded by less complex parabasalid ancestors, whose other descendants have either become extinct or not yet been identified. In any event, this major upheaval of parabasalid phylogeny emphasizes the need for caution when attempting to polarise trends such as morphological "complexity."

Three relationships seen with GADPH and rRNA are also recovered with one of the other three proteins. These are the clustering of (1) *Hypotrichomonas* and *Trichomitus* (enolase, Fig. 3B), (2) *Monocercomonas* and *Tritrichomonas* (β-tubulin, Fig. 3C), and (3) *Trichomonas* and *Tetratrichomonas* (α-tubulin, Fig. 3D).

The GAPDH tree (Fig. 3A) confirms the rRNA-based hypothesis of polyphyly (Fig. 1, Gerbod et al., 2002) of two trichomonad families, the Monocercomonadidae (here represented by *Monocercomonas* and *Hypotrichomonas*) and the Trichomonadidae (*Trichomonas*, *Tetratrichomonas*, *Tritrichomonas*, and *Trichomitus*). Moreover, the positions of *Monocercomonas* and *Hypotrichomonas* strengthen the hypothesis that the relative morphological simplicity of these Monocercomonadidae might have arisen through the loss of cytoskeletal structures, and that they are descended from more complex forms.

Multiple sets of gene duplicates were recovered for all four proteins (Fig. 3). At least three (Fig. 3A) and as many as six (Fig. 3C) of the seven parabasalids examined were found to contain either two copies of a given gene (in a total of 14 cases), three copies (two cases), or even five copies (one case). As previously seen for FeSOD (Viscogliosi et al., 1996), in almost all cases the duplicates from a particular species are closely related and group as sister sequences in the gene trees (or, if not, are part of a poorly resolved cluster of sequences in a manner not significantly inconsistent with a sister-sequence relationship). In other words, in most cases the duplications are recent and therefore do not confound phylogeny reconstruction. Only with enolase (Fig. 3B) is there evidence of deeper duplications that might cause problems with the resulting gene tree. Enolase was clearly duplicated in a common ancestor of the sister-taxa (Figs. 1 and 3A) *Hypotrichomonas* and *Trichomitus*, but what is unclear is whether this duplication predates their common ancestry with any of the other examined parabasalids. The positions of the two enolase genes from *Monocercomonas* is part of this uncertainly, as they group apart in a manner consistent with the possibility of enolase duplication at or near of the base of parabasalids.

Despite this (these) problematic duplication(s) of the enolase gene, much of the problem with the enolase tree (Fig. 3B) is probably unrelated to gene duplication and instead the result of a fundamental lack of phylogenetic resolution in the data set (note the relatively low bootstrap values on a number of nodes), resulting from some

combination of homoplasy and just too few informative characters. In the case of both tubulins, the major cause of poor resolution is probably their low rates of evolution within Parabasalia (as in most eukaryotes) and a corresponding paucity of informative characters. Excluding the unusually divergent β-tubulin of *Hypotrichomonas acosta 1*, there is only a maximum of 6 and 7% divergence among parabasalid β- and α-tubulins, respectively.

In summary, phylogenetic trees of parabasalids based on GADPH and SSU rRNA sequences are well-supported and entirely congruent with each other. For enolase and both α- and β-tubulins, however, parabasalid trees are relatively poorly resolved. Nonetheless, these trees do recover some of the same groupings that are present in the GAPDH and rRNA trees, and in no case is there strongly supported conflict among any of the five gene trees. Numerous gene duplications are evident in all four protein trees. Most duplications are very recent and only certain of those for enolase appear to hamper phylogeny reconstruction. Overall, our data suggest that GAPDH should be a valuable complement to SSU rRNA to elucidate phylogenetic relationships among parabasalids and that other molecules should be assessed in order to find other good phylogenetic markers for the group.

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, and the Institut Pasteur de Lille to E.V., Bioarchitect Research Program and EcoMolecular Science Research Program grants to T.K, Natural Science and Engineering Research Council of Canada (227301-00) to P.J.K., and National Institutes of Health RO1-GM-35087 to J.D.P. E.V. was the recipient of a RIKEN Eminent Scientist Invitation Program (RESIP) appointment as a Visiting Professor. D.G. was supported by a postdoctoral fellowship from the Japanese STA-JST. C.N. was supported by a grant from the Ministère Français de l'Education Nationale, de la Recherche et de la Technologie.

References

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

Brugerolle, G., 1976. Cytologie ultrastructurale, systématique et évolution des Trichomonadida. Ann. Station Biologique de Besse-en-Chandesse 10, 1–90.

Brugerolle, G., Patterson, D.J., 2001. Ultrastructure of *Joenina pulchella* Grassi, 1917 (Protista, Parabasalia), a reassessment of evolutionary trends in the parabasalids, and a new order Cristamonadida for devescovinid, calonymphid and lophomonad flagellates. Org. Divers. Evol. 1, 147–160.

Bruno, W.J., Socci, N.D., Halpern, A.L., 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol. Biol. Evol. 17, 189–197.

Carninci, P., Hayashizaki, Y., 1999. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19–44.

Cavalier-Smith, T., 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int. J. System Evol. Microbiol. 52, 297–354.

Dacks, J.B., Redfield, R.J., 1998. Phylogenetic placement of *Trichonympha*. J. Eukaryot. Microbiol. 45, 445–447.

Delgado-Viscogliosi, P., Viscogliosi, E., Gerbod, D., Kulda, J., Sogin, M.L., Edgcomb, V.P., 2000. Molecular phylogeny of parabasalids based on small subunit rRNA sequences, with emphasis on the Trichomonadinae subfamily. J. Eukaryot. Microbiol. 47, 70–75.

Diamond, L.S., 1957. The establishment of various trichomonads of animals and man in axenic cultures. J. Parasitol. 43, 488–490.

Edlind, T.D., Li, J., Visvesvara, G.S., Vodkin, M.H., McLaughlin, G.L., Katiyar, S.K., 1996. Phylogenetic analysis of β-tubulin sequences from amitochondrial protozoa. Mol. Phylogenet. Evol. 5, 359–367.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package). Distributed by the author, Department of Genetics, University of Washington, Seattle.

Gerbod, D., Edgcomb, V.P., Noël, C., Delgado-Viscogliosi, P., Viscogliosi, E., 2000. Phylogenetic position of parabasalid symbionts from the termite *Calotermes flavicollis* based on small subunit rRNA sequences. Int. Microbiol. 3, 165–172.

Gerbod, D., Edgcomb, V.P., Noël, C., Vanacova, S., Wintjens, R., Tachezy, J., Sogin, M.L., Viscogliosi, E., 2001a. Phylogenetic relationships of class II fumarase genes from trichomonad species. Mol. Biol. Evol. 18, 1574–1584.

Gerbod, D., Edgcomb, V.P., Noël, C., Zenner, L., Wintjens, R., Delgado-Viscogliosi, P., Holder, M.E., Sogin, M.L., Viscogliosi, E., 2001b. Phylogenetic position of the trichomonad parasite of turkeys, *Histomonas meleagridis* (Smith) Tyzzer, inferred from small subunit rRNA sequence. J. Eukaryot. Microbiol. 48, 498–504.

Gerbod, D., Noël, C., Dolan, M.F., Edgcomb, V.P., Kitade, O., Noda, S., Dufernez, F., Ohkuma, M., Kudo, T., Capron, M., Sogin, M.L., Viscogliosi, E., 2002. Molecular phylogeny of parabasalids inferred from small subunit rRNA sequences, with emphasis on the Devescovinidae and Calonymphidae (Trichomonadea). Mol. Phylogenet. Evol. 25, 545–556.

Grassé, P.P., 1952. Phylogénie. Protozoaires: Généralités. Flagellés. Ordre des Trichonymphines. In: Grassé, P.P. (Ed.), Traité de Zoologie. Masson et Cie, Paris, pp. 863–915.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Gunderson, J., Hinkle, G., Leipe, D., Morrison, H.G., Stickel, S.K., Odelson, D.A., Breznak, J.A., Nerad, T.A., Müller, M., Sogin, M.L., 1995. Phylogeny of trichomonads inferred from small-subunit rRNA sequences. J. Eukaryot. Microbiol. 42, 411–415.

Hollande, A., Carruette Valentin, J., 1971. Les atractophores, l'induction du fuseau et la division cellulaire chez les hypermastigines. Etude infrastructurale et révision systématique des Trichonymphines et des Spirotrichonymphines. Protistologica 7, 5–100.

Honigberg, B.M., 1963. Evolutionary and systematic relationships in the flagellate order Trichomonadida Kirby. J. Protozool. 10, 20–63.

Honigberg, B.M., Brugerolle, G., 1990. Structure. In: Honigberg, B.M. (Ed.), Trichomonads Parasitic in Humans. Springer-Verlag, New York, pp. 5–35.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. CABIOS 8, 275–282.

Katiyar, S.K., Edlind, T.D., 1994. Beta-tubulin genes of *Trichomonas vaginalis*. Mol. Biochem. Parasitol. 64, 33–42.

Keeling, P.J., 2002. Molecular phylogenetic position of *Trichomitopsis termopsidis* (Parabasalia) and evidence for the Trichomitopsiinae. Eur. J. Protistol. 38, 279–286.

Keeling, P.J., Doolittle, W.F., 1996a. A non-canonical genetic code in an early diverging protist lineage. EMBO J. 15, 2285–2290.

Keeling, P.J., Doolittle, W.F., 1996b. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol. Biol. Evol. 13, 1297–1305.

Keeling, P.J., Palmer, J.D., 2000. Parabasalian flagellates are ancient eukaryotes. Nature 405, 635–637.

Keeling, P.J., Poulsen, N., McFadden, G.I., 1998. Phylogenetic diversity of parabasalian symbionts from termites, including the phylogenetic position of *Pseudotrichonympha* and *Trichonympha*. J. Eukaryot. Microbiol. 45, 643–650.

Keeling, P.J., Luker, M.A., Palmer, J.D., 2000. Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. Mol. Biol. Evol. 17, 23–31.

Kirby, H., 1932. Flagellates of the genus *Trichonympha* in termites. Univ. Calif. Publ. Zool. 37, 349–476.

Kirby, H., 1944. The structural characteristics and nuclear parasites of some species of *Trichonympha* in termites. Univ. Calif. Publ. Zool. 49, 185–282.

Koidzumi, M., 1921. Studies on the intestinal protozoa found in the termites of Japan. Parasitology 13, 235–309.

Markos, A., Miretsky, A., Müller, M., 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote, *Trichomonas vaginalis*. J. Mol. Evol. 37, 631–643.

Moriya, S., Ohkuma, M., Kudo, T., 1998. Phylogenetic position of symbiotic protist *Dinenympha exilis* in the hindgut of the termite *Reticulitermes speratus* inferred from the protein phylogeny of elongation factor 1α. Gene 210, 221–227.

Moriya, S., Tanaka, K., Ohkuma, M., Sugano, S., Kudo, T., 2001. Diversification of the microtubule system in the early stage of eukaryote evolution: elongation factor 1α and α-tubulin protein phylogeny of termite symbiotic oxymonad and hypermastigote protists. J. Mol. Evol. 52, 6–16.

Neefs, J.M., Van de Peer, Y., De Rijk, P., Chapelle, P., De Wachter, R., 1993. Compilation of small ribosomal subunit RNA structures. Nucleic Acids Res. 21, 3025–3049.

Noël, C., Gerbod, D., Fast, N.M., Wintjens, R., Delgado-Viscogliosi, P., Doolittle, W.F., Viscogliosi, E., 2001. Tubulins in *Trichomonas vaginalis*: molecular characterization of α-tubulin genes, posttranslational modifications, and homology modelling of the tubulin dimer. J. Eukaryot. Microbiol. 48, 647–654.

Ohkuma, M., Ohtoko, K., Grunau, C., Moriya, S., Kudo, T., 1998. Phylogenetic identification of the symbiotic hypermastigote *Trichonympha agilis* in the hindgut of the termite *Reticulitermes speratus* based on small-subunit rRNA sequence. J. Eukaryot. Microbiol. 45, 439–444.

Ohkuma, M., Ohtoko, K., Iida, T., Tokura, M., Moriya, S., Usami, R., Horikoshi, K., Kudo, T., 2000. Phylogenetic identification of hypermastigotes, *Pseudotrichonympha*, *Spirotrichonympha*, *Holomastigotoides*, and parabasalian symbionts in the hindgut of termites. J. Eukaryot. Microbiol. 47, 249–259.

Riley, D.E., Krieger, J.N., 1992. Rapid and practical DNA isolation from *Trichomonas vaginalis* and other nuclease-rich protozoa. Mol. Biochem. Parasitol. 51, 161–164.

Silberman, J.D., Clark, C.G., Sogin, M.L., 1996. *Dientamoeba fragilis* shares a recent common evolutionary history with the trichomonads. Mol. Biochem. Parasitol. 76, 311–314.

Strimmer, K., Von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Viscogliosi, E., Müller, M., 1998. Phylogenetic relationships of the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase, from parabasalid flagellates. J. Mol. Evol. 47, 190–199.

Viscogliosi, E., Durieux, I., Delgado-Viscogliosi, P., Bayle, D., Dive, D., 1996. Phylogenetic implication of iron-containing superoxide dismutase genes from trichomonad species. Mol. Biochem. Parasitol. 80, 209–214.

Viscogliosi, E., Edgcomb, V.P., Gerbod, D., Noël, C., Delgado-Viscogliosi, P., 1999. Molecular evolution inferred from small subunit rRNA sequences: what does it tell us about phylogenetic relationships and taxonomy of the parabasalids? Parasite 6, 279–291.

Yamin, M.A., 1979. Flagellates of the orders Trichomonadida Kirby, Oxymonadida Grassé, and Hypermastigida Grassi & Foà reported from lower termites (Isoptera families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, and Serritermitidae) and from the wood-feeding roach *Cryptocercus* (Dictyoptera: Cryptocercidae). Sociobiology 4, 5–119.