J Mol Evol (2004) 58:550–556
DOI: 10.1007/s00239-003-2577-4

JOURNAL OF MOLECULAR EVOLUTION

Polymorphic Insertions and Deletions in Parabasalian Enolase Genes

Patrick J. Keeling

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, 3529-6270 University Blvd., Vancouver, British Columbia V6T 1Z4, Canada

Received: 16 July 2003 / Accepted: 14 November 2003

Abstract. Insertions and deletions in gene sequences have been used as characters to infer phylogenetic relationships and, like any character, the information they contain varies in utility between different levels of evolution. In one case, the absence of two otherwise highly conserved deletions in the enolase genes of parabasalian protists has been interpreted as a primitive characteristic that suggests these were among the first eukaryotes. Here, semi-environmental 3′-RACE was used to sample enolases from parabasalia in the hindgut of the termite *Zootermopsis angusticolis* to examine the conservation of this character within the parabasalia. Parabasalian homologues were found to be polymorphic for these deletions, and the phylogeny of parabasalian enolases shows that the deletion-possessing genes branch within deletion-lacking genes (i.e., they did not form two clearly distinct groups). Phylogenetic incongruence was detected in the carboxy-terminal third of the sequence (in the region of the deletions), but there is no unambiguous evidence for recombination. The polymorphism of this character discredits these deletions as strong evidence for the early origin of parabasalia, although the complex distribution makes it impossible to state whether parabasalian enolases were ancestrally like those of other eukaryotes. These observations stress the importance of strong corroborating evidence when considering insertion and deletion data, and raises some interesting questions about the apparent variation in degree of conservation of these deletions between different eukaryotic groups.

Key words: Parabasalia — Enolase — Insertion — Deletion — Phylogeny

Introduction

Reconstructing an accurate representation of eukaryotic relationships has proven to be a difficult problem for a number of reasons. The first phylogenies based on small subunit ribosomal RNA (SSU rRNA) sequences appeared to provide a fully resolved and reproducible picture of eukaryotic evolution (Sogin 1991) that was congruent in many ways with the first sampled protein coding gene phylogenies (Brown and Doolittle 1995; Hashimoto et al. 1994; Kamaishi et al. 1996). Since then, however, phylogenies based on several different protein coding genes have identified a number of serious problems with this picture. One of the most widely studied problems is the position of the root of eukaryotes and the effects that divergent sequences have on the placement of several putatively "ancient" lineages. A number of gene trees originally identified diplomonads, microsporidia, and parabasalia as being the first-branching eukaryotes (Sogin 1991), but as the sequences placing these lineages deep are all relatively divergent, serious concerns were raised about the validity of this position (Baldauf 2003; Embley and Hirt 1998; Philippe and Adoutte 1998; Simpson and Roger 2002). This is perhaps best illustrated by the

email: pkeeling@interchange.ubc.ca

microsporidia, which have been demonstrated to be highly derived fungi rather than the primitive protists suggested by SSU rRNA and several other genes (reviewed by Keeling and Fast 2002).

The problems associated with phylogenetic reconstruction at this level have led to several searches for molecular characteristics that could be used independently of phylogenetic reconstruction to determine evolutionary relationships and even identify the earliest-diverging lineages of eukaryotes (Keeling and Palmer 2000; Stechmann and Cavalier-Smith 2002). These characteristics can include the presence or absence of introns, fused genes, gene duplications, or, most commonly, insertions and deletions in gene sequences. If the sequences of insertions and the surrounding areas are highly conserved, these events are considered to provide powerful and seemingly easily interpreted markers for major evolutionary events (e.g., Archibald et al. 2002; Baldauf and Palmer 1993; Baldauf et al. 1996; Rivera and Lake 1992). In one such case, a pair of deletions in the gene for enolase has been used to suggest a deep-branching position for the parabasalia (Keeling and Palmer 2000). All eukaryotes other than parabasalia possess two closely spaced deletions (for simplicity, these characters are referred to as deletions in eukaryotes since eubacteria and archaea have amino acids at these positions). These deletions are absent in all prokaryotes with a few exceptions: in a small number of phylogenetically isolated taxa scattered across the archaea and eubacteria, deletions or insertions have occurred in the same region of the protein (Bapteste and Philippe 2002; Hannaert et al. 2000; Keeling and Palmer 2000). It is important to note, however, that every prokaryotic enolase with an insertion or deletion in this region has close relatives with enolase genes resembling "normal" prokaryotes. Accordingly, the region can reasonably be interpreted as being able to tolerate changes in size (it maps to an external loop) but is still highly conserved when eukaryotic and prokaryotic biodiversity is considered as a whole. Interestingly, however, these deletions were found to be uniformly absent in genes from diverse parabasalia. Enolase phylogeny gave no indication that the parabasalian genes were derived from a prokaryote by lateral gene transfer, so this distribution was interpreted as the result of deletions that took place after parabasalia diverged but before the divergence of all other known eukaryotes (Keeling and Palmer 2000). Another recently described character, a fusion of thymadylate synthase and dihydrofolate reductase, has led to the alternative suggestion that the root of the eukaryotes is near the branch leading to animals, fungi, and their close relatives (Stechmann and Cavalier-Smith 2002). These characters are irreconcilable, so one or both of these characters must be misleading.

Here we use semienvironmental 3′-RACE (rapid amplification of cDNA ends) to show that this region of parabasalian enolase genes is polymorphic for the presence of the deletions, and the distribution of this character within parabasalia is not apparently congruent with the phylogeny of the enzyme. This situation is reminiscent of other cases reported recently where insertions and deletions do not follow the same pattern as phylogenetic trees based on the genes where they are found (Bapteste and Philippe 2002; Hannaert et al. 2000), which has been interpreted as resulting from recombination in some cases (Archibald and Roger 2002b; Keeling and Palmer 2001). Parabasalian enolase provides an interesting case where insertion and deletion data are apparently highly conserved at the macroevolutionary scale, but are polymorphic between close relatives in this group, and so may be just as misleading as phylogenetic reconstruction. Specifically, these data undermine the weight of this character as evidence for the deep branching position of parabasalia.

Materials and Methods

RNA and Extraction and 3′-RACE

Zootermopsis angusticolis was collected from damp logs at Jericho Beach, Vancouver, and maintained in the lab. Hindguts were dissected from 10 termites and the contents were recovered in Trager's (1934) Medium U. Hindgut contents were precipitated by centrifugation, resuspended in Trizol (Invitrogen), and transferred to a Knotes Duall 20 tissue homogenizer, and RNA was purified as described (Keeling and Leander 2003). Whole hindgut RNA was used as a template for poly (A)-primed first strand synthesis using a 3′-RACE adapter primer, GCGAGCACAGAATTAATACGA CTCACTATAGGT$_{12}$VN (Ambion). Enolase transcripts were amplified using the enolase-specific primer AGCGGCAAC CCGACNGTNGARGTNGA and a 3′ anchor primer, GCGAGCACAGAATTAATACGACT. Products of the expected size were isolated and cloned into pCR2.1 (Invitrogen). Fourteen individual clones were sequenced, nine of which were unique and were completely sequenced on both strands. All clones were parabasalian enolases. New sequences have been deposited in GenBank under accession numbers AY393926–AY393934.

Phylogenetic Analyses

New enolase sequences were added to an existing alignment of amino acid sequences (Keeling and Palmer 2001) and phylogenies inferred using distance and maximum likelihood (alignment is available upon request). Distances were calculated for a 97-sequence data set consisting of 301 alignable characters using TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1996) with the WAG substitution frequency matrix, and amino acid frequencies estimated from the data. Site-to-site rate variation was modeled on a discrete Γ distribution with eight rate categories and an invariable sites category, with the shape parameter α and the proportion of invariable sites estimated from the data by TREE-PUZZLE (parameters were 1.07 and 0.04, respectively). Trees were inferred from distances using weighted neighbor joining with WEIGHBOR 1.0.1a (Bruno et al.

	26–48	237–253	
Zea mays	VGLSD-GSYRGAV----PSGASTGIYE	IEKAGYT-GK-VVIGMDVA	Other Eukaryotes
Arabidopsis thaliana	IHTSN-GIKTAAV----PSGASTGIYE	IEKAGYT-GK-VVIGMDVA	
Chlamydomonas reinhardtii	VYTRK-GMFRAAV----PSGASTGVHE	IEKAGYT-GK-VKIGMDVA	
Mastocarpus papillatus 1	VATAG-GKFSAMV----PSGASTGIYE	IEMAGYT-GK-IKVGMDVA	
Plasmodium falciparum	LETNL-GIFRAAV----PSGASTGIYE	IKSAGYE-GK-VKIAMDVA	
Tetrahymena thermophila	LTVDN-GVFRAAV----PSGASTGIYE	IKKAGHE-GK-IKISMDVA	
Mastigamoeba balamuthi	LTTEK-GLFRSAV----PSGASTGIYE	IKQAGYT-GK-IEIGMDVA	
Trypanosoma brucei	VTTER-GVFRSAV----PSGASTGVYE	IEEAGHR-GK-FAICMDCA	
Entamoeba histolytica	ITTGK-GMFRSCV----PSGASTGVHE	IAKAGYT-GK-IEIAMDCA	
Homo sapiens alpha	LFTSK-GLFRAAV----PSGASTGIYE	IGKAGYT-DK-VVIGMDVA	
Caenorhabditis elegans	LFTEK-GVFRAAV----PSGASTGVHE	IDKAGYT-GK-ISIGMDVA	
Saccharomyces cerevisiae	LTTEK-GVFRSIV----PSGASTGVHE	IKAAGHD-GK-VKIGLDCA	
Neocallimastix frontalis	VTTDK-GLFRAAV----PSGASTGVHE	IAKAGYT-GK-VKIGMDVA	
Trichomonas vaginalis 1	VYVKYLGRITLAGRSSAPSGASTGVGE	IKAAGYEAGKDVFIAMDVA	Parabasalia
Trichomonas vaginalis 2	VYVKYLGRITLAARSSAPSGASTGVGE	IKAAGYEAGKDIMIGMDVA	
Trichomitus batrachorum 1	VYADYLGQVIFAGRSSAPSGASTGVGE	MREAGYEPGVDMFLGLDAA	
Trichomitus batrachorum 2	VYAKQFGEVVFAGRSSAPSGASTGSGE	IVEAGLKVGEDIRLGLDAA	
Monocercomonas ATCC 50210-1	VYANYCGEVVFAGRSSAPSGASTGSNE	ITEAGYKPGEEIRIGLDAA	
Monocercomonas ATCC 50210-2	VYATQYGQVNLAGRSSAPSGASTGSNE	IKDAGYTPGTDIRLGLDAA	
Tritrichomonas foetus	VYVHYLGEEQFAGRSSAPSGASTGSNE	VKDAGYEPGVDVNICLDAA	
Tetratrichomonas gallinarum	VYANYMGGVRFAGRSSAPSGASTGSGE	VKAAGYEPGKDIFFGLDAA	
Hypotrichomonas acosta 1	VYADYLGEVIFAGRSSAPSGASTGSGE	MTEAGYTPAKDMFLAFDAA	
Hypotrichomonas acosta 2	VYAKRFGDVVFAGRSSAPSGASTGSNE	IQAAGLKVGEDIRLGLDAA	
Zootermopsis symbiont 29	IYAKYCGEVIFAGRSSAPSGASTGSNE	IENLGYT-GQ-VKIALDAA	
Zootermopsis symbiont 22	VYANYLGGVDLVARSSAPSGASIGSGE	IAKANYE-GQ-VKIAMDAA	
Zootermopsis symbiont 31	VYAKYCGEVIFAGRSSAPSGASTGSNE	IDAAGYA-GQ-VNIALDAA	
Zootermopsis symbiont 33	VYANYLGGVDLIARSSAPSGASTGSGE	INDAGYF-GQ-VNIALDAA	
Zootermopsis symbiont 13	VYATVLSTEKFVARSSAPSGASTGVGE	IKDSGYEPAVDVFIALDPA	
Zootermopsis symbiont 17	VYATVLSSEKFVARSSAPSGASTGVGE	ITDSGYTPAVDVFIALDPA	
Zootermopsis symbiont 30	VYATVLSTEKFVARSSAPSGASTGVGE	ITDSGYVPAVDVFIALDPA	
Trichonympha agilis	VYAKHLGEVVLVARSSAPSGASTGSGE	VKAAGYVPVDDVKYCLDCA	
Pyrococcus horikoshii	VHTPI-AMGRAAV----PSGASTGTHE	IEEAGYKPGDEIALAIDAA	Archaea
Methanopyrus kandleri	VELED-GVGRAMV----PSGASTGTYE	IEEAGYAPGKEIALALDAA	
Methanocaldococcus jannaschii	VITKGNGYGSAIV----PSGASTGTHE	VKKAGYE--DEVVFALDAA	
Methanosarcina acetivorans	VFTPK-GFGRASV----PSGASTGTNE	IEEAGYTES-EVTIGLDAA	
Pyrococcus abyssi	VATDE-GFGRFAS----PIEENPMLHI	EDNNVAYIKPLGPPELFLE	
Sulfolobus solfataricus	IRTSD-GESFGDA----PAGASKGTRE	INNAGYEG--KIYMGMDAA	
Thermoplasma acidophilum	VYIPG-GFGRTSA----PAGASTGETE	VNEVSSETKVKIYTGLDFA	
Halobacterium NRC-1	VTTDD-GFGRAAA----PSGASTGEHE	AAEVADEFGFDVRLGLDLA	
Helicobacter pylori	VVLSD-NKASAIV----PSGASTGKRE	IEKAGYKLGEEIALALDVA	Eubacteria
Campylobacter jejuni	VTLSD-GVGAAIV----PSGASTGSKE	IKKAGYENRVKIAL--DVA	
Escherichia coli	VHLEG-GVGMAAA----PSGASTGSRE	VKAAGYELGKDITLAMDCA	
Agrobacterium tumefaciens	VYLED-GMGRAAV----PSGASTGAHE	IEKAGYRPGEDMYVGLDCA	
Treponema pallidum	VSLSD-GFGRACV----PSGASTGEFE	IAKAGLAPRKDVCIALDCA	
Bacillus subtilis	VYTET-GFGRALV----PSGASTGEYE	IEKAGFKPGEEVKLAMDAA	
Mycobacterium tuberculosis	VALID-GFARAAV----PSGASTGEHE	IESAGLRPGADVALALDAA	
Deinococcus radiodurans	VHLDS-GSGRAIV----PSGASTGSHE	IQQAGYEPGKDICIALDPA	
Thermotoga maritima	VVLED-GMGRAIV----PSGASTGKFE	IEKAGYKPGEEVFIALDCA	
Aquifex aeolicus	VELES-GLGRAIV----PSGASTGERE	IEKAGYKPGEDILLALDVA	
Synechocystis sp. PCC 6803	VRLES-GHGIAQV----PSGASTGSFE	IEQAGYKPGSQIALAMDIA	
Chloroflexus aurantiacus	VRLES-GVGRAIV----PSGASTGAHE	IEKAGYRPGEQIVIALDPA	

Fig. 1. Insertions and deletions in two regions of enolase numbered according to the *Zea* sequence. **A** An example of two insertions that support the monophyly of parabasalian enolases examined here. **B** Two single-amino acid deletions formerly known from all eukaryotes except parabasalia. Five of the new *Zootermopsis* symbiont sequences also lack these deletions (two of those lacking the deletions are very similar to those shown), but four resemble other eukaryotes. Other deletions in the same area are seen in some prokaroytes. Note that the number of these is skewed by the overrepresentation of deletion-containing prokaryotic sequences in order to show their state: The vast majority of prokaryotic enolases does not have deletions or insertions in this region.

2000) and Fitch–Margoliash using FITCH 3.6a (Felsenstein 1993). An alignment restricted to parabasalian sequences with 23 taxa and 319 characters was also analyzed as above and with protein maximum likelihood (ML). Protein ML phylogeny and bootstraps of parabasalian enolases were inferred using ProML 3.6a (Felsenstein 1993) with 10 random addition replicates and global rearrangements. Site-to-site rate variation was modeled on a gamma curve using the –R option with invariable sites and eight categories of variable sites estimated by TREE-PUZZLE.

Because recombination has been suggested for other data with similar characteristics (Archibald and Roger 2002b; Keeling and Palmer 2001), evidence for phylogenetic incongruence was sought in parabasalian sequences using a version of LIKEWIND (Archibald and Roger 2002a) modified for protein data (courtesy of A.J. Roger). This program detects phylogenetic incongruence between subalignments in a sliding window across a molecule. To specifically examine potential incongruence involving parabasalia, 25 four-taxon data sets were analyzed, each including a random example of one parabasalian sequence with the insertion, one without, one other eukaryote, and one prokaryote. Confidence intervals for patches of incongruence were estimated using SIMPLOCKPRO, a protein sequence version of the program SIMBLOCK (Archibald and Roger 2002a) (SIMPLOCKPRO was developed in collaboration with M. Field and is distributed with LIKEWIND: hades.biochem.dal.ca/Rogerlab/Software/software.html).

Results and Discussion

Characterization of Parabasalian Enolases by Semienvironmental 3′-RACE

To significantly extend the known diversity of parabasalian enolases, total RNA from the hindgut contents of the western damp-wood termite *Zootermopsis angusticolis* was used as a template for 3′-RACE with an enolase-specific primer known to recognize parabasalian genes (Keeling and Palmer 2000). A single product was recovered, and sequencing of 14 individual clones revealed 9 unique sequences, each terminating in a short 3′ UTR. All new sequences were most similar to previously characterized parabasalian homologues, and all contained most of the four insertions that are characteristic of parabasalian genes (e.g., see Fig. 1). One group of sequences lacked an insertion common to other parabasalia (not shown), but this insertion is poorly conserved compared with other eukaryotes, and therefore such a polymorphism is not surprising. Unexpectedly, however, some of the

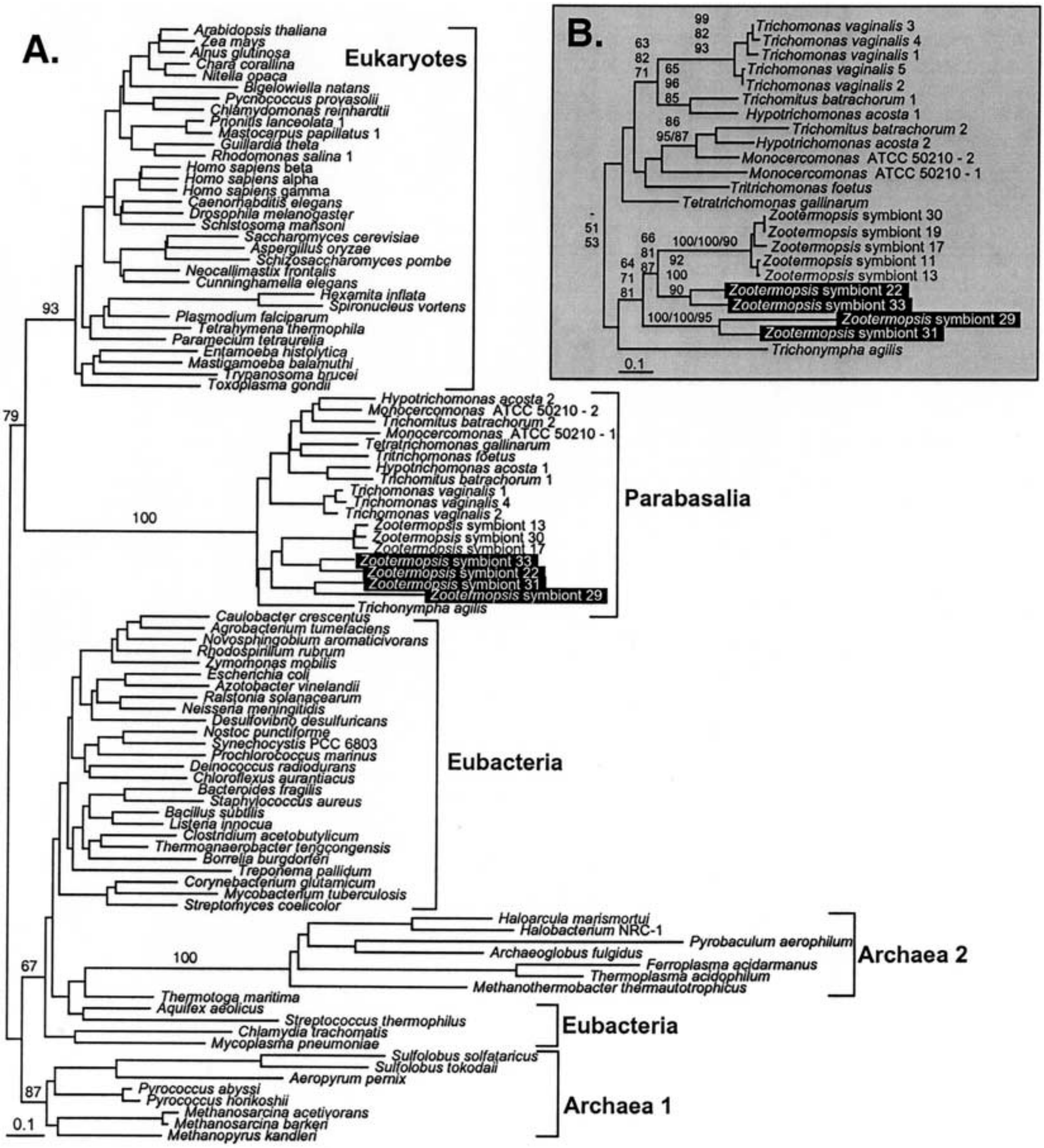

Fig. 2. Phylogeny of parabasalian enolase, with deletion-containing sequences shaded black. **A** Global enolase phylogeny showing that all *Zootermopsis* symbiont sequences are members of the strongly supported and divergent parabasalian clade. **B** Protein maximum likelihood phylogeny of parabasalian enolases showing that the deletion-containing sequences (black) are closely related to other *Zootermopsis* symbiont sequences and the hypermastigote *Trichonympha agilis*.

new sequences did not contain the distinctive and otherwise highly conserved "prokaryotic" character previously used to argue for an ancient origin of parabasalia (Keeling and Palmer 2000). Of the nine new sequences, five resembled other parabasalia in lacking the two single-amino acid deletions common to all other eukaryotes, while the other four sequences included the eukaryotic deletions (Fig. 1).

Parabasalian Enolase Phylogeny and the Distribution of Insertions and Deletions

The polymorphic distribution of this character in parabasalian enolases seemingly undermines the strength of this character for inferring the evolutionary history of the group, but without knowing the pattern of the character in parabasalian phylogeny, it is difficult to interpret the distribution. A simple explanation may be that these new deletion-containing genes are the deepest-branching parabasalia and represent the ancestral state of parabasalian enolases. However, a phylogeny of enolase (Fig. 2) does not support this interpretation.

A universal enolase phylogeny shows the general characteristics of enolase phylogenies shown previously (Hannaert et al. 2000; Keeling and Palmer 2000, 2001). The eukaryotes form a well-supported group to the exclusion of parabasalia, and the para-

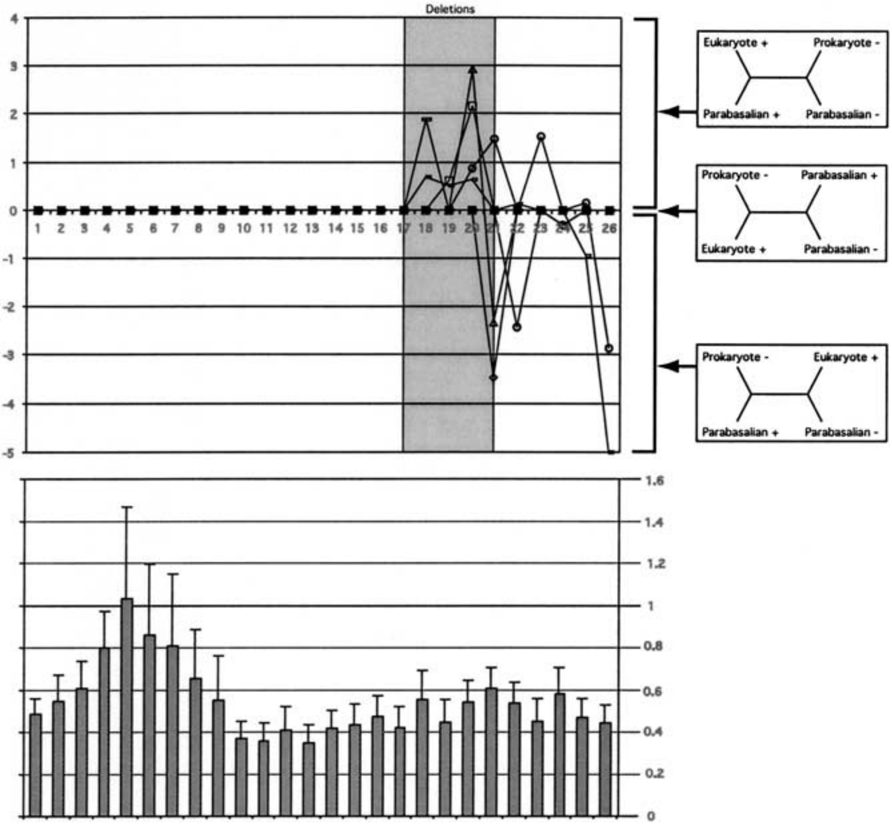

Fig. 3. Phylogenetic incongruence profile of parabasalian enolases. **A** Δln*L* values of 50 amino acid windows sliding 10 amino acid increments for 25 four taxon trees (taxon selection is described in text). Windows yielding the expected tree (shown at right, middle) have a Δln*L* of zero. Windows yielding an alternate tree are given positive or negative values depending on the tree (shown at right, above and below the expected tree) so the profile can more easily be read. The deletions and the two amino acids between them are not included in the analysis. The region where they are found falls in windows 17–21 (shaded), in the cases of windows 17 and 21, at the extreme ends. **B** Pairwise distances between all windows and all parabasalia in tests shown in A.

basalia branch at the base of other eukaryotes with good support. The archaebacteria fall in two positions: Archaea 1, between eukaryotes and other prokaryotes: and the divergent Archaea 2, within the eubacteria (in some analyses, the archaeal genes formed a single group at the position of Archaea 1; not shown). All new *Zootermopsis* symbiont sequences branch within the very strongly supported clade of parabasalian enolases (Fig. 2A), confirming the parabasalian nature of these genes also suggested by most of the insertions and deletions noted above. However, the *Zootermopsis* symbiont sequences that possess the distinctive eukaryotic deletions (black shading) do not fall as a sister group to all other parabasalia but, instead, are nested within the group and closely related to the non-deletion-containing genes from *Zootermopsis*. This relationship between parabasalian enolases was analyzed in more detail using more comprehensive ML and distance methods (Fig. 2B), and in all cases the deletion-lacking sequences were paraphyletic, the *Zootermopsis* symbiont sequences branched together, and all *Zootermopsis* sequences were weakly associated with the hypermastigote, *Trichonympha* (Gerbod et al. 2004). It is likely that the *Zootermopsis* sequences are also from a hypermastigote, as this termite contains three species of *Trichonympha* in addition to several monocercomonads and a trichomonad (Yamin 1979), the latter of which are represented elsewhere in the tree (*Monocercomonas* and *Trichomonas*, respectively). Regardless of the exact source of these sequences, however, the distribution of deletion-containing and deletion-lacking parabasalia is not consistent with a single simple explanation.

Patterns of Phylogenetic Incongruence in Parabasalian Enolase

The region around the deletions in the deletion-containing parabasalian sequences shares a number of weak similarities specifically with other, nonparabasalian eukaryotic sequences (e.g., V247 and K248 in Fig. 1B). In two other cases where the distribution of apparently homologous insertions conflicted with the phylogeny of the genes where they were found, recombination was suggested as a mechanism to move insertions between distantly related genes (Archibald and Roger 2002b; Keeling and Palmer 2001). To determine if the deletion-lacking or deletion-containing parabasalian sequences may have been involved in a recent recombination event in the area of the deletions, phyiogenetic incongruence was tested between different regions of the enolase genes. To focus as specifically as possible on the these four sequences, 25 sliding window tests were carried out (as described under Materials and Methods), each including a random selection of one eukaryote, one prokaryote, one deletion-lacking parabasalian, and one deletion-containing parabasalian. The incongruence profiles across the protein are plotted in Fig. 3A.

In nearly all windows in all four-taxon tests the expected topology was obtained (right, middle). The exceptions are plotted such that the difference in log likelihoods was assigned a positive value if the alternate tree is that which places the taxa with the deletions together (top right) and a negative value if they and the parabasalia are separated (bottom right). The profile is striking in that no deviations are observed in the first 17 windows, but the last third of the protein exhibits variants in 8 of the 25 tests. Most of the variants that place the deletion-containing taxa together occur in the region surrounding the deletions (shaded; the deletions themselves and the two amino acids between them were not included in these analyses). The significance of the discrepancies in the eight sets where incongruence was observed was tested by parametric boostrapping (Archibald and Roger 2002a), and the 95% confidence level of all eight fell between 0.001 to 0.026.

The absence of incongruence at the amino terminus of the protein suggests that there is something unusual about the carboxy terminus. To examine the possibility that the carboxy terminus is simply more divergent in parabasalian enolases, and therefore more prone to noise, the pairwise distances between all parabasalia selected in the 25 tests were profiled such that each pair contains one deletion-containing sequence and one deletion-lacking sequence. This profile is shown in Fig. 3B, where it can be seen that the region around the deletions is neither unusually divergent nor conserved, but similar to the upstream region where no incongruence was detected.

While these tests are interesting, they do not suggest an obvious explanation. First, incongruence was not detected in all 25 tests, so the taxa chosen to represent eukaryotes and prokaryotes are likely very important (there is no obvious correlation in the taxa chosen and whether incongruence was detected). If the incongruence does result from recombination, then it could be that choosing a close relative of the donor is critical. Second, a number of tests revealed incongruence, supporting the opposite alternative topology—that in which neither parabasalia nor possession of deletions was monophyletic. In some cases different windows in the same test supported opposite alternatives, for which there is no obvious or simple explanation.

Interpreting Insertions and Deletions in Macroevolutionary Time Scales

In recent years, the utility of analyzing insertions and deletions has been questioned on a number of fronts. In the case of reconstruction large-scale eukaryotic relationships, it has been argued that such characters are too prone to polymorphism to be particularly useful (Bapteste and Philippe 2002). However, a number of important relationships are supported by insertion or deletion characters with no conflicting data, including animals and fungi (Baldauf and Palmer 1993; Baldauf et al. 2000) and cercozoa and foraminifera (Archibald et al. 2002, 2004). These relationships are also supported by multiple molecular phylogenies (Baldauf and Doolittle 1997; Baldauf et al. 2000; Keeling 2001; Longet et al. 2003), indicating that such characters can remain highly conserved over long evolutionary timescales. At the same time, it has also been shown that recombination can transmit insertions between distantly related sequences (e.g., Archibald and Roger 2002b), providing additional reason to interpret insertion and deletion data with caution. Therefore, while many sites of insertion and deletion clearly do evolve too quickly to be useful in resolving large-scale phylogenetic questions, and all insertion and deletion data are best interpreted in the context of other supporting phylogenetic information, this does not mean that all such characters are useless in the same way that all phylogeny is not discredited by the general understanding that phylogenies can be actively misleading.

The polymorphic nature of enolase deletions in parabasalia, on the other hand, shows the importance of adequate taxon sampling at all levels: it is important to sample broadly among different groups of eukaryotes, but it is also important to sample deeply within key groups to reveal if the conservation of characters differs at different levels. In this case, although the deletions appear to be a generally highly conserved characters across the diversity of eukaryotes, the parabasalia are themselves polymorphic. While the conservation of this character across eukaryotes in general remains unchallenged, the polymorphism among parabasalia discredits the utility of this character for inferring the position of this group within eukaryotes and, by extension, any evidence that parabasalia are a particularly ancient group.

Acknowledgments. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (227301-00). The author thanks A.J. Roger for LIKEWINDPRO and for advice on the development of SIMBLOCKPRO, M. Field for developing SIMBLOCKPRO software, S. Moriya for providing the *Trichonympha agilis* sequence prior to publication, and J.M. Archibald and N.M. Fast for helpful discussions and critical reading of the manuscript. The author is a Scholar of the Canadian Institute for Advanced Research and a New Investigator of the Michael Smith Foundation of Health Research and the Canadian Institutes for Health Research.

References

Archibald JM, Roger AJ (2002a) Gene conversion and the evolution of euryarchaeal chaperonins: A maximum likelihood-based method for detecting conflicting phylogenetic signals. J Mol Evol 55:232–245

Archibald JM, Roger AJ (2002b) Gene duplication and gene conversion shape the evolution of archaeal chaperonins. J Mol Biol 316:1041–1050

Archibald JM, Longet D, Pawlowski J, Keeling PJ (2002) A novel polyubiquitin structure in Cercozoa and Foraminifera: Evidence for a new eukaryotic supergroup. Mol Biol Evol 20:62–66

Archibald JM, Longet D, Pawlowski J, Keeling PJ (2004) Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. J Eukaryot Microbiol 51:113–118

Baldauf SL (2003) The deep roots of eukaryotes. Science 300:1703–1706

Baldauf SL, Doolittle WF (1997) Origin and evolution of the slime molds (Mycetozoa). Proc Natl Acad Sci USA 94:12007–12012

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

Baldauf SL, Palmer JD, Doolittle WF (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc Natl Acad Sci USA 93:7749–7754

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–977

Bapteste E, Philippe H (2002) The potential value of indels as phylogenetic markers: Position of trichomonads as a case study. Mol Biol Evol 19:972–977

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Bruno WJ, Socci ND, Halpern AL (2000) Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction. Mol Biol Evol 17:189–197

Embley TM, Hirt RP (1998) Early branching eukaryotes? Curr Opin Genet Dev 8:624–629

Felsenstein J (1993) PHYLIP (phylogeny inference package). University of Washington, Seattle

Gerbod D, Sanders E, Moriya S, Noël C, Takasu H, Fast NM, Delgado-Viscogliosi P, Ohkuma M, Judo T, Capron M, Palmer JD, Keeling PJ, Viscogliosi E (2004) Molecular phylogenies of parabasalia inferred from four protein genes and comparison with rRNA trees. Mol Phylogenet Evol (in press)

Hannaert V, Brinkmann H, Nowitzki U, Lee JA, Albert MA, Sensen CW, Gaasterland T, Muller M, Michels P, Martin W (2000) Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: Pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. Mol Biol Evol 17:989–1000

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

Kamaishi T, Hashimoto T, Nakamura Y, Nakamura F, Murata S, Okada N, Okamoto K-I, Shimzu M, Hasegawa M (1996) Protein phylogeny of translation elongation factor EF-1α suggests Microsporidians are extremely ancient eukaryotes. J Mol Evol 42:257–263

Keeling PJ (2001) Foraminifera and Cercozoa are related in actin phylogeny: Two orphans find a home? Mol Biol Evol 18:1551–1557

Keeling PJ, Fast NM (2002) Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 56:93–116

Keeling PJ, Leander BS (2003) Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. J Mol Biol 326:1337–1349

Keeling PJ, Palmer JD (2000) Parabasalian flagellates are ancient eukaryotes. Nature 405:635–637

Keeling PJ, Palmer JD (2001) Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. Proc Natl Acad Sci USA 98:10745–10750

Longet D, Archibald JM, Keeling PJ, Pawlowski J (2003) Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. Int J Syst Evol Microbiol (in press)

Philippe H, Adoutte A (1998) The molecular phylogeny of eukaryota: solid facts and uncertainties. In: Coombs GH, Vickerman K, Sleigh MA, Warren A (eds) Evolutionary relationships among protozoa. Chapman & Hall, London, pp 25–56

Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257:74–76

Simpson AG, Roger AJ (2002) Eukaryotic evolution: Getting to the root of the problem. Curr Biol 12:R691–R693

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

Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. Science 297:89–91

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

Trager W (1934) The cultivation of a cellulose-digesting flagellate, Trichomonas termopsidis, and of certain other termite protozoa. Biol Bull 66:182–190

Yamin MA (1979) Flagellates of the orders Trichomonadida Kirby, Oxymonadida Grassé, and Hypermastigida Grassi & Foà reported from lower termites (Isoptera Falilies Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, and Serritermididae) and from the wood-feeding roach *Cryptocercus* (Dictyoptera:Ceyptocercidae). Sociobiology 4:1–120