Available online at www.sciencedirect.com

SCIENCE DIRECT®

Gene 340 (2004) 227–235

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Lateral gene transfer and the complex distribution of insertions in eukaryotic enolase

James T. Harper, Patrick J. Keeling*

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

Received 28 January 2004; received in revised form 7 June 2004; accepted 29 June 2004
Available online 25 August 2004
Received by A. Roger

Abstract

Insertions and deletions in protein-coding genes are relatively rare events compared with sequence substitutions because they are more likely to alter the tertiary structure of the protein. For this reason, insertions and deletions which are clearly homologous are considered to be stable characteristics of the proteins where they are found, and their presence and absence has been used extensively to infer large-scale evolutionary relationships and events. Recently, however, it has been shown that the pattern of highly conserved, clearly homologous insertions at positions with no other detectable homoplasy can be incongruent with the phylogeny of the genes or organisms in which they are found. One case where this has been reported is in the enolase genes of apicomplexan parasites and ciliates, which share homologous insertions in a highly conserved region of the gene with the apparently distantly related enolases of plants. Here we explore the distribution of this character in enolase genes from the third major alveolate group, the dinoflagellates, as well as two groups considered to be closely related to alveolates, haptophytes and heterokonts. With these data, all major groups of the chromalveolates are represented, and the distribution of these insertions is shown to be far more complicated than previously believed. The incongruence between this pattern, the known evolutionary relationships between the organisms, and enolase phylogeny itself cannot be explained by any single event or type of event. Instead, the distribution of enolase insertions is more likely the product of several forces that may have included lateral gene transfer, paralogy, and/or recombination. Of these, lateral gene transfer is the easiest to detect and some well-supported cases of eukaryote-to-eukaryote lateral transfer are evident from the phylogeny.

Keywords: Phylogeny; Paralogy; Lateral transfer; Recombination

1. Introduction

Insertions and deletions in protein-coding genes are relatively uncommon events, and when they do take place, it is most often in external loops of a protein where they are better tolerated. In many cases, such areas of a protein are highly prone to both insertions and deletions, and accumulate change rapidly, sometimes differing between closely related organisms. On the other hand, insertions and deletions can also be retained throughout long periods of evolutionary time with very little change, and such events are often seen as stable characteristics of the protein in a certain lineage of organisms. While the impact of insertion and deletion events on protein function are seldom known and the dynamics of their origin and loss are unclear, insertions or deletions which are obviously homologous are often used as markers for evolutionary events or major lineages in the tree of life (e.g., Baldauf and Palmer, 1993; Archibald et al., 2002).

Enolase (2-phospho-D-glycerate hydrolase, EC 4.2.1.11) is a ubiquitous enzyme responsible for catalyzing the interconversion of 2-phospho-D-glycerate and phosphoenolpyruvate. Enolase is highly conserved at the sequence level, but contains a large number of insertions and

Abbreviations: EST, expressed sequence tag; ML, maximum likelihood.

* Corresponding author. Tel.: +1 604 822 4906; fax: +1 604 822 6089.
E-mail address: pkeeling@interchange.ubc.ca (P.J. Keeling).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.06.048

deletions, sometimes in otherwise highly conserved regions of the gene. Some of the enolase insertions and deletions have been shown to be highly plastic (Bapteste and Philippe, 2002), while others have been found to be lineage-specific in their distribution among eukaryotes, and these have been used to infer a number of evolutionary relationships (Baldauf and Palmer, 1993; Read et al., 1994; Dzierszinski et al., 1999; Keeling and Palmer, 2000, 2001).

Two such insertions were noted as distinguishing features of enolases from both plants and the malaria parasite *Plasmodium falciparum* (Van der Straeten et al., 1991; Read et al., 1994; Dzierszinski et al., 1999). These insertions are flanked by highly conserved regions of sequence predicted to be part of a coil region at the outer face of enolase dimer interactions. Together with other features of the *Plasmodium* enolase, it was concluded that the insertions revealed a common ancestry of apicomplexans and plants, or that the enolase of apicomplexans is derived from their plastid endosymbiont (Read et al., 1994; Hannaert et al., 2000). However, neither insertion was found in the green alga *Chlamydomonas*, a close relative of plants, suggesting that the *Plasmodium* insertions were unlikely to be explained so simply. Indeed, surveying the distribution of these inserts in some of the relatives of both apicomplexans and plants showed that the insertions were restricted to plants and the charophytes, their closest green algal relatives (collectively the streptophytes), as well as apicomplexans and their close relatives the ciliates (together with dinoflagellates making up the alveolates). Furthermore, enolases possessing both insertions were not closely related in enolase phylogenies, so the phylogeny and distribution of insertions were not congruent (Keeling and Palmer, 2001).

Such incongruence is potentially very significant for the interpretation of insertions in proteins, since homologous insertions are frequently considered to be very stable through time. Several possible explanations for the incongruence in enolase insertions have been offered. Assuming the phylogeny is correct in separating the lineages that contain them, it is possible that the insertions were lost independently in several intervening lineages (at the very least red algae and chlorophyte algae—which are known to be more closely related to plants than are alveolates). Alternatively, the insertions may have been moved between lineages by subgenic lateral transfer, or lateral transfer followed by recombination, leading to homologous insertions in the otherwise distantly related streptophyte and alveolate enolases (Keeling and Palmer, 2001). It was also suggested that insertion-containing genes represent ancient paralogues of eukaryotic enolase, and that differential gene losses led to the present distribution of insertions (Bapteste and Philippe, 2002). This explanation, however, would suggest the groups sharing insertions should be related in the phylogeny, which has not yet been demonstrated.

We have characterized enolase genes from the third major group of alveolates (dinoflagellates) and organisms considered to be close relatives of alveolates (heterokonts and haptophytes; Cavalier-Smith, 1999; Fast et al., 2001; Harper and Keeling, 2003), so that enolase sequences are now known from representatives of all major chromalveolate groups. The distribution of the streptophyte–alveolate insertions in these groups is even more complex and even more incongruent with enolase and organismal phylogeny than originally thought. No simple explanation, such as a single gene duplication, lateral transfer, recombination event, or unresolved phylogeny can account for the phylogenetic distribution of enolase insertions, which appear to track a complex history that could include several or all of these events. This complexity may also have a significance that extends beyond enolase, since the insertions are markers for unusual evolutionary patterns that might easily be ignored if the phylogeny alone was examined.

2. Materials and methods

2.1. Algal strains, DNA isolation, and enolase amplification

Axenic cultures of the haptophytes *Isochrysis galbana* (strain CCMP 1323), *Pavlova lutheri* (strain CCMP 1325), *Prymnesium parvum* (strain CCMP 1926) and the heterokont *Phaeodactylum tricornutum* (strain CCMP 1327) were obtained from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton and grown in 100–300 ml of f/2-Si medium at 16 °C (12:12 light–dark cycle). Genomic DNAs from the oomycete heterokonts *Apodachlya brachynema* (strain CBS 557.69), *Phytophthora palmivora* (strain CBS 236.30), and *Thraustotheca clavata* (strain CBS 343.33) were kindly donated by A. W. DeCock, and genomic DNA from the raphidophyte heterokont *Heterosigma akashiwo* was kindly donated by K. Ishida. Algal cultures were harvested by centrifugation and cell pellets were lysed by grinding in a Knotes Duall 20 tissue homogeniser. Genomic DNAs (gDNAs) were extracted from *I. galbana*, *P. lutheri* and *P. parvum* lysates using the DNeasy Plant Mini Kit (Qiagen).

Enolase genes were PCR-amplified from gDNAs using either 5′ primers AGCGGCAACCCGACNGTNGAR GTNGA or CCGGTCGACCGGNATHTAYGARGC with primer 3′ GCGCTCGCGRCANGGNGCNCCNGTYTT. PCR was completed under the following conditions: 95 °C for 2 min; 40 cycles of 92 °C for 45 s, 48 °C for 45 s, and 72 °C for 1 min and 30 s; and 72 °C for 5 min. PCR products were gel-purified and cloned into the TOPO-TA vector pCR2.1 (Invitrogen), and multiple clones of each were sequenced on both strands with ABI BigDye terminator chemistry (Applied Biosystems). Expressed sequence tags (ESTs) for the three forms of enolase from the dinoflagellate *Heterocapsa triquetra* were recovered from an ongoing EST project, and the clones completely sequenced.

2.2. Phylogenetic analyses

New enolase sequences were deposited in GenBank (accession numbers AY430415-AY430424) and added to an existing amino acid alignment (Keeling and Palmer, 2001). Distance and maximum-likelihood (ML) analyses were performed on an alignment that included representative enolase sequences from a variety of eukaryotic groups. ML distances were calculated using TREE-PUZZLE 5.0 (Strimmer and von Haeseler, 1996), using the WAG substitution matrix with the frequency of amino acid usage calculated from the data. Rate-across-site variation was modeled on a discrete gamma distribution with eight variable rate categories, estimating invariable sites and the shape parameter alpha from the data. Distance trees were constructed with weighted neighbor-joining using WEIGHBOR 1.0.1a (Bruno et al., 2000) and Fitch-Margoliash using FITCH 3.6a (Felsenstein et al., 1993). Fitch-Margoliash trees were inferred using the global rearrangements option and 10 input order jumbles. Weighted neighbor-joining and Fitch-Margoliash bootstrap trees were constructed (without global rearrangements and implementing two input order jumbles in Fitch-Margoliash) from 100 resampled data sets with gamma-corrected distances (with the rate category parameters above) using PUZZLEBOOT 1.0.3 (by M. Holder and A. Roger: http://www.tree-puzzle.de).

Protein maximum-likelihood analyses were performed using PhyML (Guindon and Gascuel, 2003). PhyML was performed using an input tree generated by BIONJ, the JTT model of amino acids substitution, proportion of variable rates estimated from the data, and nine categories of substitution rates (eight variable and one invariable; parameters estimated by TREE-PUZZLE). PhyML bootstrap trees were constructed using the same parameters as the individual ML trees.

Topology tests were carried out by calculating site-likelihoods using PAML 3.12 (Yang, 1997) for the 100 ML bootstrap trees, the ML tree, and four alternative topologies where the insertion-containing genes were made monophyletic and moved to the position of each insertion-containing group in the ML tree (i.e., on the branches leading to *Tetrahymena* and *Colpods*, apcomplexa and *Paramecium*, haptophytes, and *Bigelowiella* and streptophytes). Approximately unbiased (AU) tests were then conducted on the site-likelihoods using CONSEL 0.1d (Shimodaira and Hasegawa, 2001). The main analyses excluded sequences with significant missing data, so additional trees were inferred using *Heterocapsa* enolase 3 and individual EST data from the dinoflagellates *Amphidinium* and *Alexandrium* using the ML method outlined above. The Heterocapsa EST is truncated at the 5′ end, so only ML trees were inferred. *Amphidinium* and *Alexandrium* ESTs were more substantially truncated, so these trees were only confirmation of their sister relationship to *Heterocapsa* sequences, and are not shown.

3. Results and discussion

3.1. Expanded distribution of insertions in eukaryotic enolase

Dinoflagellates are alveolates, while heterokonts, haptophytes and cryptomonads are hypothesized to be related to alveolates as part of a larger group, the chromalveolates (Cavalier-Smith, 1986; Fast et al., 2001; Harper and Keeling, 2003). To determine whether any other chromalveolates share the characteristic insertions found in apicomplexan and ciliate homologues, enolase genes were characterized from a dinoflagellate, three diverse haptophytes, and five diverse heterokonts.

The distribution of insertions in the newly determined sequences significantly complicates pattern of insertions in other eukaryotic enolases (Fig. 1). The first insertion in question consists of a single amino acid (at position 96 of the *Oryza sativa* 1 gene), and will not be considered in detail because its short length makes it difficult to determine whether insertions are homologous or parallel. We will only note that it is always present when the second insertion is present, but is also found in the *Rhodomonas* enolase and one copy of gene from the chlorophyte *Pycnococcus* (not shown). The second insertion (corresponding to positions 104–108 in the *Oryza* 1 sequence) is a highly conserved pentapeptide containing two rare tryptophan residues. This high degree of conservation makes it clear that homologous insertions are found in the enolases of apicomplexans, the chlorarachniophyte *Bigelowiella*, ciliates, haptophytes, some heterokont sequences, one dinoflagellate sequence, and streptophytes. Smaller insertions with no sequence similarity are found in some red algae and in all diplomonads. These smaller insertions confirm that this region can tolerate length heterogeneity relatively easily, as has been suggested (Keeling and Palmer, 2001; Bapteste and Philippe, 2002), but there is no evidence that these are homologous. Of the heterokonts, the oomycetes appear to possess two enolases, one with the insertions and one without, since an insertion-containing copy was found in *Phytophthora*, an insertion-lacking copy was found in *Thraustrotheca*, and both types were found in *Apodachlya*. No insertion-containing enolase was found in the non-oomycete heterokonts (the raphidophyte *Heterosigma* and the diatom *Phaeodactylum*), and both insertion-containing and insertion-lacking genes from heterokonts formed strongly supported groups in enolase phylogeny (see below), suggesting the presence of both types within the heterokonts may be characteristic of oomycetes specifically. Enolases were also examined from the nearly complete genome of the diatom *Thalassiosira pseudonana* (genome.jgi-psf.org/; estimated to be 95% complete) and, in agreement with this distribution, only insertion-lacking copies were detectable.

Interestingly, three distinct enolase genes were found in the dinoflagellate *H. triquetra*, and two lacked the insertions

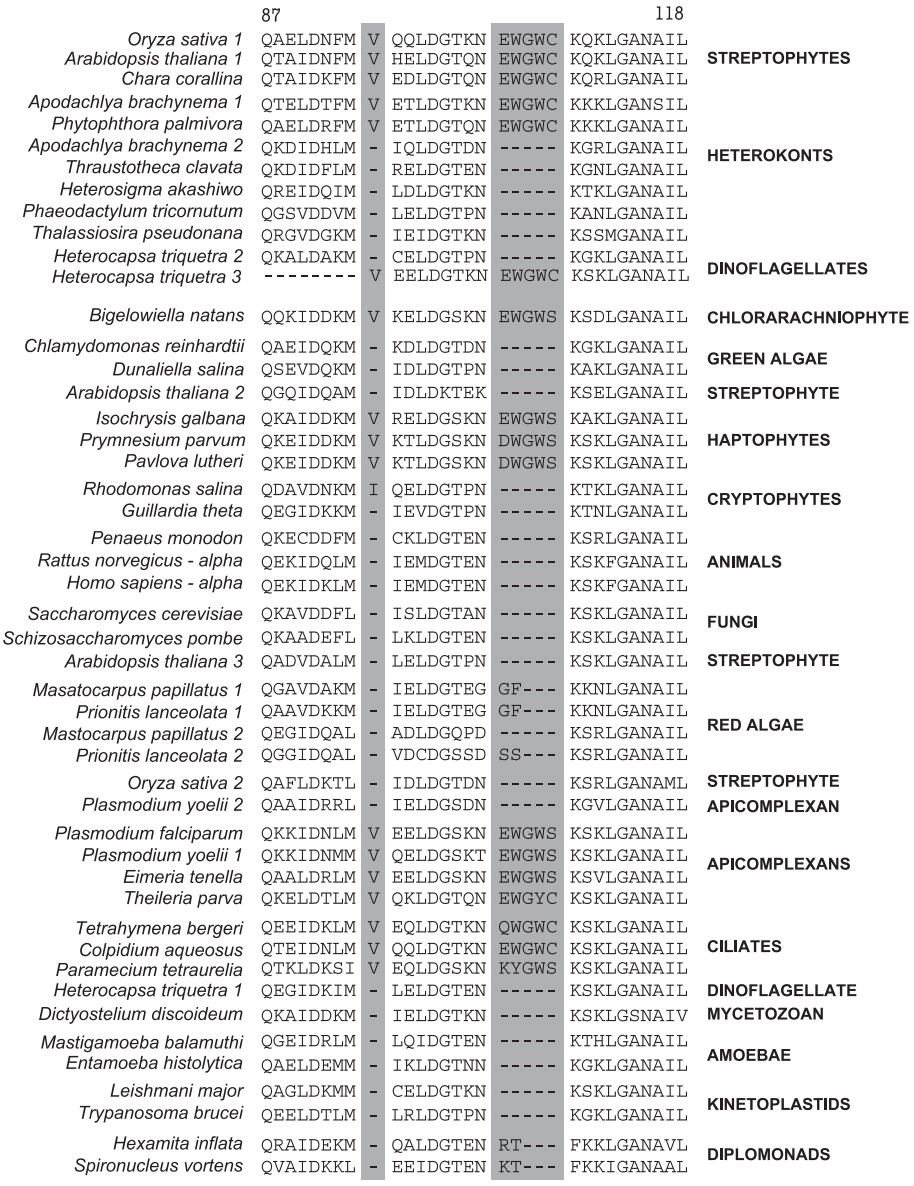

Fig. 1. Enolase amino acid sequence surrounding the two insertions from selected eukaryotes. The regions corresponding to the two insertions are each enclosed by a shaded box. Amino acid positions correspond to the *O. sativa* 1 sequence.

despite the fact that dinoflagellates are known to branch within the alveolates as sisters to apicomplexans (e.g., Fast et al., 2002) and virtually all sampled enolases from both apicomplexa and ciliates contain the insertions. The apicomplexan *Plasmodium yoelii* and plant *O. sativa* were also found to encode both insertion-containing and insertion-lacking paralogues and *Arabidopsis* has previously been shown to possess multiple forms (Keeling and Palmer, 2001).

To summarize, with the exceptions noted above, these insertions are now known to be present in all known enolases from apicomplexans, chlorarachniophytes, ciliates, haptophytes, streptophytes and in one of two classes of enolase from oomycete heterokonts. This distribution is not consistent with the phylogeny of eukaryotes as we know it. Indeed, plotting insertion-containing enolases onto a schematic of eukaryotic phylogeny (Fig. 2) reveals a relatively punctate distribution of insertion-containing genes among organisms with insertion-lacking or both types of enolase.

3.2. Incongruence between enolase phylogeny and insertion

The phylogeny of enolase including these new sequences (except the partial *Heterocapsa* enolase 3 which is described below) is shown in Fig. 3. This phylogeny resembles other analyses of enolase in that several major eukaryotic groups are resolved, but overall higher level relationships between groups are equivocal (Hannaert et al., 2000; Keeling and Palmer, 2001; Bapteste and Philippe, 2002). All apicomplexan enolases, which possess the pentapeptide insertion, group together in all analyses. The ciliates, on the other hand, form a monophyletic clade in some analyses, but not

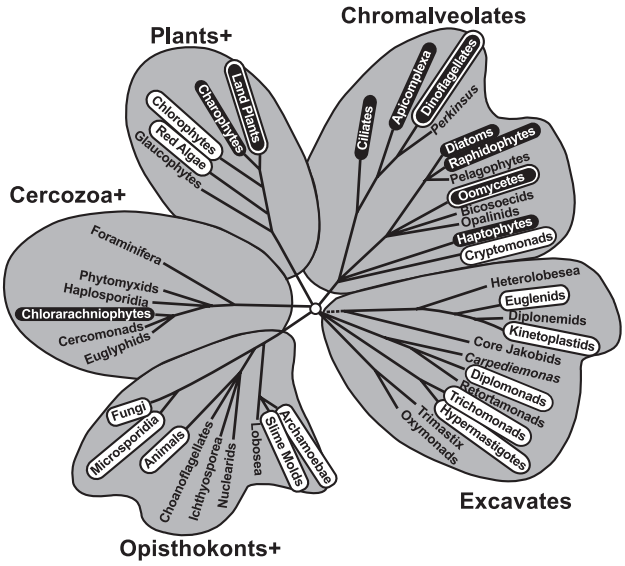

Fig. 2. Schematic hypothesis of eukaryotic relationships based on other data, plotting the distribution of enolase insertions. There are five hypothetical eukaryotic supergroups named here in informal terms: excavates, chromalveolates, plants, cercozoa and opisthokonts (the + denotes that this is named for the best known large group within the supergroup). Major lineages are shown in black text if no enolase data are known, in black text on a white oval where the pentapeptide insertion has not been found, in white text on a black oval if only insertion-containing enolases have been found, and where a group contains both insertion-containing and insertion-lacking copies, the name is shown within concentric white and black ovals. The distribution of this character is not congruent with our current understanding of eukaryotic relationships.

in that shown in Fig. 3 where peniculid ciliates group solidly together as do the tetrahymenid ciliates. Interestingly, a specific relationship between two insertion-lacking enolases from the apicomplexan *P. yoelii* and the plant *O. sativa* was found to be strongly supported in all analyses, which is unusual since complete genome sequences are known for relatives of both (*P. falciparum* and *A. thaliana*, respectively), and these lack such a paralogue, suggesting a recent origin of one or perhaps both of these genes in these organisms.

All other taxa with enolase genes containing the insertions (streptophytes, heterokonts, a chlorarachniophyte, and haptophytes) fall in a large and poorly supported clade with other insertion-lacking enolases from chlorophytes, heterokonts, and one other dinoflagellate gene. Within this cluster, the heterokonts lacking the insertions (diatoms, raphidophytes, and oomycetes) consistently form a well-supported group (73–84%). This group also includes the second dinoflagellate homologue, which branches specifically at the base of the diatoms with high support (93–100%). The oomycete insertion-containing paralogues branch with the streptophytes (Fig. 3), while the insertion-containing haptophytes and chlorarachniophyte genes branch in various positions with low support.

Two other dinoflagellate EST projects have recently been conducted, and enolase genes were found in both. A single gene was identified from each of *Amphidinium* and *Alexandrium*, and these branched with *Heterocapsa* enolase 1 and 2, respectively (not shown). This confirms that both of these paralogues are relatively ancient within dinoflagellate even though neither branch in the position expected of a dinoflagellate enolase. A third dinoflagellate paralogue was also identified in *Heterocapsa*, and this gene did contain the insertions, but did not branch with other alveolates. Instead, the *Heterocapsa* enolase 3 branched with the insertion-containing oomycete genes, albeit with no support (Fig. 4).

3.3. Complex models of evolution for eukaryotic enolase

Multiple gains of the pentapeptide insertion can be ruled out as highly unlikely because of the high degree of sequence conservation in the insertion. Multiple loss is more difficult to rule out, but is becoming increasingly unlikely with more sampling since the number of independent losses required to explain the distribution is growing. Previously, independent losses in red and green algae were necessary to achieve this distribution, but now additional losses in apicomplexans, streptophytes, dinoflagellates, heterokonts, and cryptomonads would have to be considered. The fact that multiple insertion-containing and insertion-lacking paralogues are also emerging from whole-genome sequencing (e.g., *Oryza*, *Arabidopsis* and *Plasmodium*) and EST analysis (e.g., *Heterocapsa*) indicates that this pattern is likely to become further complicated as such data become available from a broader variety of eukaryotes.

For all other simple models of evolution, one would expect the genes that contain them, and/or the organisms where these genes are found to be related. Yet, while the presence of insertions is generally confined to well-supported subgroups, insertion-containing and insertion-lacking genes are found in apparently distantly related genes scattered among some members of three distinct and divergent supergroups of eukaryotes (Fig. 2). Plants are known to be related to charophyte and chlorophyte green algae, and more distantly related to red algae (Moreira et al., 2000). Similarly, heterokonts, haptophytes and cryptomonads are currently thought to form a group (Yoon et al., 2002), which is related to alveolates (Baldauf and Palmer, 1993; Van de Peer et al., 1996; Fast et al., 2001; Harper and Keeling, 2003), collectively called chromalveolates (Fig. 2). Chlorarachniophytes are members of the Cercozoa (Archibald et al., 2002), and are not closely related to either of these two supergroups (Fig. 2). Typically, when phylogenies and characters such as insertions are in direct conflict with established relationships, they are explained by one of several factors, such as lack of phylogenetic resolution, ancient paralogy, lateral transfer, or interspecific recombination, each of which we will examine in turn.

If the insertion-containing and insertion-lacking genes represent ancient paralogues, then each should form a discrete group, especially if the gene duplication was an ancient event as would have to be the case here. Clearly enolase is present in multiple copies in many genomes, and

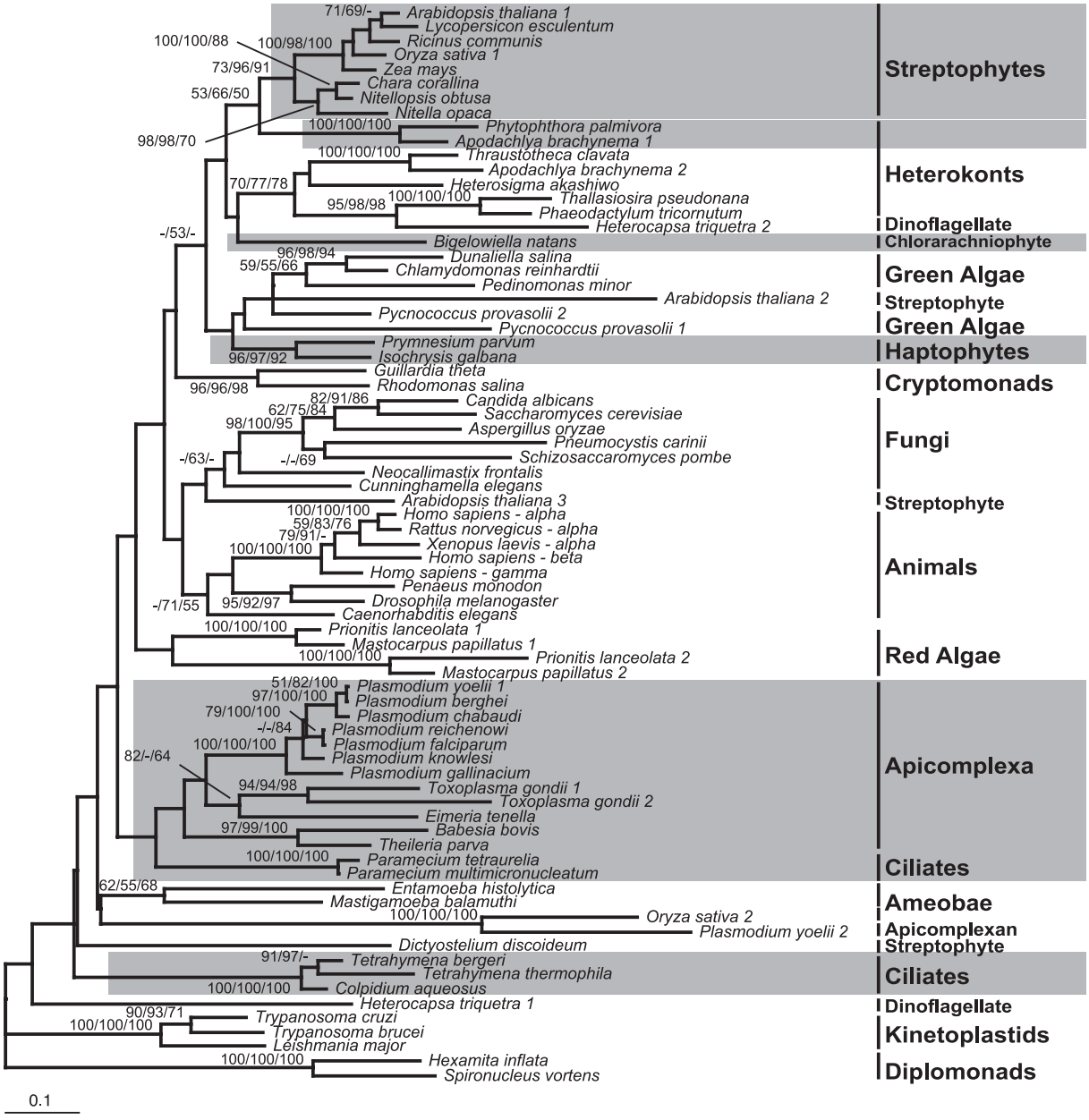

Fig. 3. Protein maximum likelihood phylogeny (PhyML) corrected for site-to-site rate variation. Bootstrap values are shown for major nodes with support over 50% from most methods, and are (left to right) weighted neighbor-joining, Fitch-Margoliash, and PhyML (dashes represent support lower than 50%). Enolase genes possessing the insertions are highlighted with shaded boxes and major groups are bracketed and labeled to the right.

the phylogeny is not strong and it is entirely possible that more of the insertion-containing subgroups may be related than the phylogeny shows (i.e., lack of phylogenetic resolution probably does play a role in this distribution). The possibility that the insertion-containing genes form a group was examined using AU-tests to compare the ML tree with four trees consisting of monophyletic insertion-containing clades positioned at each node where any insertion-containing gene is found in the ML tree. The topologies with this clade positioned where apicomplexa and ciliates fall, or where the remainder of the ciliates fall were both rejected at the 1% level. Interestingly, the other two positions (sister to green algae and sister to heterokonts) were not rejected. While the phylogeny clearly does not exclude this possibility, a paralogy-only model would demand parallel loss of one enolase in virtually all of the groups where enolase is known.

The distribution of insertions may also be due to widespread lateral transfer, which would also demand that the insertion-containing genes be related to one another. While this seems to be an oversimplification of the data, there are significant exceptions. First, all three genes from *H. triquetra* are unusual for a number of reasons suggesting lateral transfer. Dinoflagellates are known to branch within

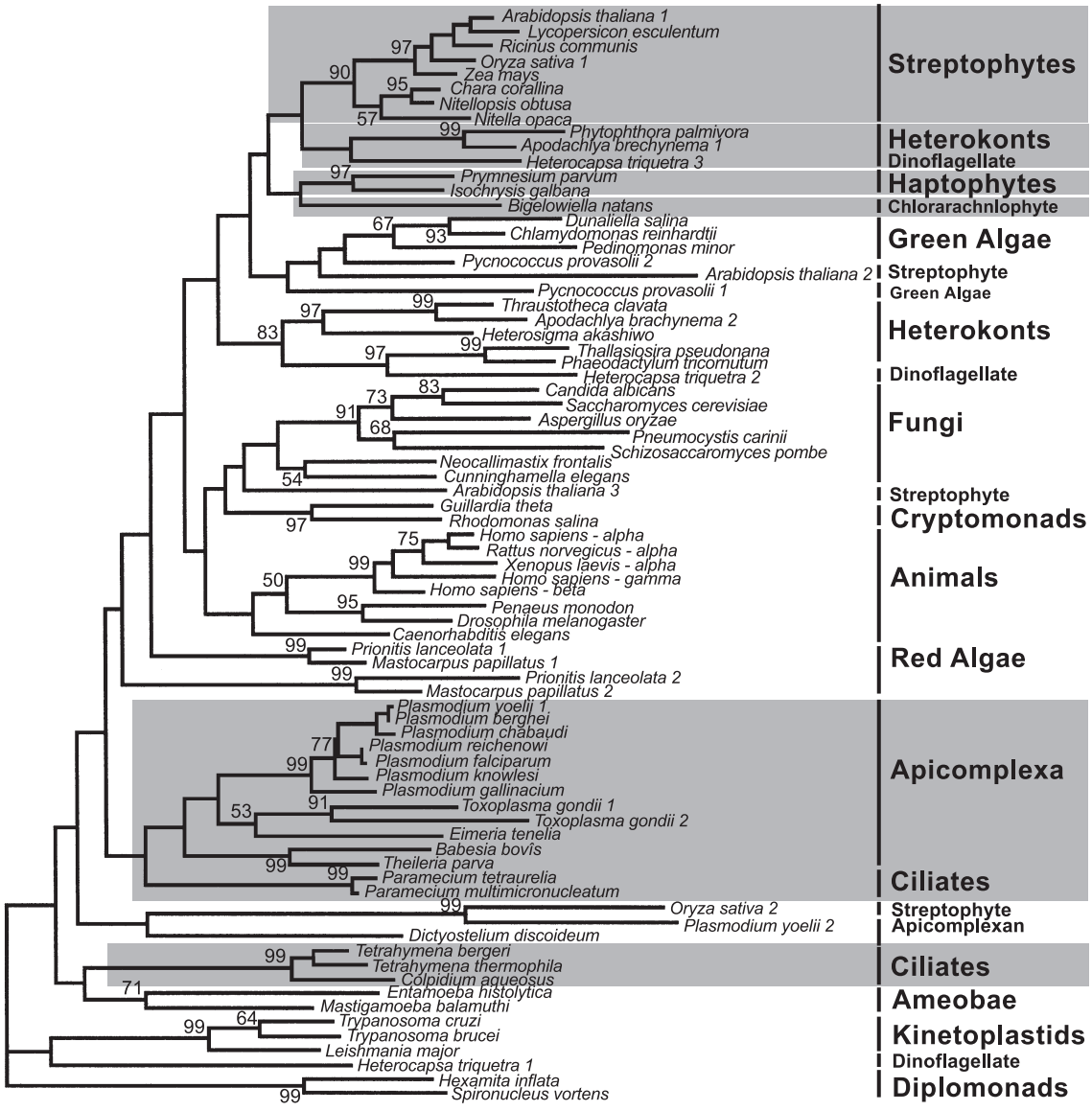

Fig. 4. Protein maximum likelihood phylogeny of enolase including the truncated insertion-containing enolase 3 from *H. triquetra*. Bootstrap values are shown for major nodes with support over 50% from PhyML. Other notations are as in Fig. 3.

the alveolates as sister to the apicomplexa (Fig. 2), but two dinoflagellate paralogues lack the insertions despite the presence of insertion-containing genes in all other alveolates examined so far. More specifically, dinoflagellate enolase 2 (found in *Heterocapsa* and *Amphidinium*) falls in a strongly supported group with the diatoms. This close relationship with the diatoms (which fall in the expected phylogenetic context of heterokonts) is exactly what one would expect if this dinoflagellate enolase was derived by lateral gene transfer from a diatom or a close relative of diatoms. The third dinoflagellate paralogue contains the insertions and so could be expected to be more straightforward, but in fact it is more closely related to the insertion-containing oomycete genes than to other alveolates (Fig. 4), suggesting further possible lateral transfer to the dinoflagellate. The insertion-lacking forms of enolase from *P. yoelii* and *O. sativa* are also probably indicative of lateral transfer rather than insertion loss since they are clearly related to one another and do not branch with the insertion-containing apicomplexan enolases or the two forms of enolase from streptophytes. These examples are all of interest since cases of lateral transfer between two eukaryotes are only just beginning to emerge (e.g., Andersson et al., 2002; Archibald et al., 2003; Bergthorsson et al., 2003), and the prevalence in enolase might be taken to show that this process is more active than presently thought.

Lastly, it has been suggested that insertions can “move” between genes by recombination (Archibald and Roger, 2002) resulting in two distantly related genes sharing a homologous insertion not found in their close relatives. This

was evoked to explain the presence of this insertion in alveolates and streptophytes (Keeling and Palmer, 2001), but with the addition of these new data, it is clear that this explanation is too simple since it does not explain the presence of insertion-containing and lacking genes in oomycetes and does not explain either of the insertion-lacking dinoflagellate genes. At present, there is no single plausible explanation for the evolution of eukaryotic enolases. Instead, it appears that the insertions in enolase track a complex series of many kinds of events, perhaps including all of those listed above.

3.4. Remaining questions

As a model for molecular evolution, enolase is a remarkable gene because it is highly conserved at the sequence level, but has a large number of insertions and deletions, a relatively rare combination. These features allow a complex evolutionary history that might otherwise have gone unrecognized to stand out: for example, the absence of a close relationship between the two oomycete paralogues or the dinoflagellate enolase 1 and other alveolates might be put down to insufficient phylogenetic resolution were it not for the more glaring inconsistencies in the presence and absence of insertions. While the insertions are useful tools to demonstrate these unusual trends, there is no evidence that whatever lies at the root of these observations is restricted to enolase or is caused by the insertions. By extension, more proteins may have similarly complex histories, but these would be difficult to discern without a well-resolved phylogeny or other markers. The nature of the five amino acid insertion itself is growing more interesting as well, as it is highly conserved at the sequence level, suggesting strong selection, but at the same time highly polymorphic in presence versus absence, suggesting the opposite. It may be that the insertion is well tuned to its sequence environment in the genes that have it, which would also raise interesting questions if it is being transmitted between distantly related genes by recombination.

Acknowledgements

This work was supported by a grant (MOP-42517) from the Canadian Institutes for Health Research (CIHR). *H. triquetra* EST sequencing is part of the Protist EST Program supported by Genome Canada via Genome Atlantic. We thank A.W. de Cock for providing gDNAs from oomycetes, K. Ishida for providing gDNAs from *H. akashiwo*, and Ross Waller, John Archibald, Naomi Fast, and Audrey de Koning for critical reading of the manuscript. PJK is a scholar of the Canadian Institute of Advanced Research and a New Investigator of the Michael Smith Foundation for Health Research and the CIHR.

References

Andersson, J.O., Sjogren, A.M., Davis, L.A., Embley, T.M., Roger, A.J., 2002. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr. Biol. 13, 94–104.

Archibald, J.M., Roger, A.J., 2002. 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, J.M., Longet, D., Pawlowski, J., Keeling, P.J., 2002. A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. Mol. Biol. Evol. 20, 62–66.

Archibald, J.M., Rogers, M.B., Toop, M., Ishida, K., Keeling, P.J., 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc. Natl. Acad. Sci. U. S. A. 100, 7678–7683.

Baldauf, S.L., Palmer, J.D., 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc. Natl. Acad. Sci. U. S. A. 90, 11558–11562.

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.

Bergthorsson, U., Adams, K.L., Thomason, B., Palmer, J.D., 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424, 197–201.

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.

Cavalier-Smith, T., 1986. The kingdom Chromista: origin and systematics. Progr. Phycol. Res. 4, 309–347.

Cavalier-Smith, T., 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J. Eukaryot. Microbiol. 46, 347–366.

Dzierszinski, F., Popescu, O., Toursel, C., Slomianny, C., Yahiaoui, B., Tomavo, S., 1999. The protozoan parasite *Toxoplasma gondii* expresses two functional plant-like glycolytic enzymes. Implications for evolutionary origin of apicomplexans. J. Biol. Chem. 274, 24888–24895.

Fast, N.M., Kissinger, J.C., Roos, D.S., Keeling, P.J., 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol. Biol. Evol. 18, 418–426.

Fast, N.M., Xue, L., Bingham, S., Keeling, P.J., 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. J. Eukaryot. Microbiol. 49, 30–37.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package). University of Washington, Seattle.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J.A., Albert, M.A., Sensen, C.W., 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.

Harper, J.T., Keeling, P.J., 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol. Biol. Evol. 20, 1730–1735.

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

Keeling, P.J., Palmer, J.D., 2001. Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. Proc. Natl. Acad. Sci. U. S. A. 98, 10745–10750.

Moreira, D., Le Guyader, H., Phillippe, H., 2000. The origin of red algae and the evolution of chloroplasts. Nature 405, 69–72.

Read, M., Hicks, K.E., Sims, P.F., Hyde, J.E., 1994. Molecular characterisation of the enolase gene from the human malaria parasite *Plasmodium falciparum*. Evidence for ancestry within a photosynthetic lineage. Eur. J. Biochem. 220, 513–520.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

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

Van de Peer, Y., Van der Auwera, G., De Wachter, R., 1996. The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. J. Mol. Evol. 42, 201–210.

Van der Straeten, D., Rodrigues-Pousada, R.A., Goodman, H.M., Van Montagu, M., 1991. Plant enolase: gene structure, expression, and evolution. Plant Cell 3, 719–735.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yoon, H.S., Hackett, J.D., Pinto, G., Bhattacharya, D., 2002. A single, ancient origin of the plastid in the Chromista. Proc. Natl. Acad. Sci. U. S. A. 99, 15507–15512.