# Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase

**Patrick J. Keeling* and Jeffrey D. Palmer**

Department of Biology, Indiana University, Bloomington, IN 47405

Contributed by Jeffrey D. Palmer, July 2, 2001

**Enolase genes from land plants and apicomplexa (intracellular parasites, including the malarial parasite, *Plasmodium*) share two short insertions. This observation has led to the suggestion that the apicomplexan enolase is the product of a lateral transfer event involving the algal endosymbiont from which the apicomplexan plastid is derived. We have examined enolases from a wide variety of algae, as well as ciliates (close relatives of apicomplexa), to determine whether lateral transfer can account for the origin of the apicomplexan enolase. We find that lateral gene transfer, likely occurring intracellularly between endosymbiont and host nucleus, does account for the evolution of cryptomonad and chlorarachniophyte algal enolases but fails to explain the apicomplexan enolase. This failure is because the phylogenetic distribution of the insertions—which we find in apicomplexa, ciliates, land plants, and charophyte green algae—directly conflicts with the phylogeny of the gene itself. Protein insertions have traditionally been treated as reliable markers of evolutionary events; however, these enolase insertions do not seem to reflect accurately the evolutionary history of the molecule. The lack of congruence between insertions and phylogeny could be because of the parallel loss of both insertions in two or more lineages, or what is more likely, because the insertions were transmitted between distantly related genes by lateral transfer and fine-scale recombination, resulting in a mosaic gene. This latter process would be difficult to detect without such insertions to act as markers, and such mosaic genes could blur the ''tree of life'' beyond the extent to which whole-gene lateral transfer is already known to confound evolutionary reconstruction.**

phylogeny | protein insertions | apicomplexa | recombination

Our concept of a universal tree of life rests on the assumption that most or all of the components of an organism share a common evolutionary heritage. However, there are growing signs that this assumption is often violated at the molecular level by the process of lateral gene transfer, or the movement of genetic information between genomes. This process first was recognized to occur at high frequencies with the acceptance that mitochondria and plastids are of eubacterial, endosymbiotic origin. Most organellar proteins are encoded by nuclear genes, and many of these genes can be shown to have been transferred from the bacterially derived endosymbiont genomes to the nucleus (1). This special case was extended by the discovery that many genes whose protein products are *not* associated with modern organelles are nonetheless also derived from mitochondrial and plastid endosymbionts, a process dubbed endosymbiotic gene replacement (2). In these two cases of *intracellular* gene transfer, the transfer of genetic information is associated with longstanding endosymbiotic partnerships.

A more general process of interorganismal lateral transfer also occurs without the benefit of such a close association. There is evidence that, in the bacterial world, significant quantities of genetic information move between closely related species or strains (3). Individual cases of gene transfer involving more distantly related organisms also are being characterized with an ever-increasing frequency; many of these cases involve genes moving between eubacteria and archaebacteria, or between eubacteria and eukaryotes. Most recently, the potential importance of such evolutionarily distant lateral transfers has taken on a completely new dimension as whole-genome analyses have suggested that a considerable fraction of the genes in many microbial genomes have phylogenetic histories that do not match our expectations (for review, see ref. 3). Often, these cases are documented only poorly, and processes other than lateral transfer certainly are also involved (4), but the vast numbers of such genes have led to the belief that lateral gene transfer may be an unexpectedly powerful force in genome evolution, challenging the very notion of a single ''tree of life'' (5, 6).

Here, we have investigated a potential case of eukaryote–eukaryote lateral transfer involving enolase genes of apicomplexa. Apicomplexa are obligate intracellular parasites, many of which cause a number of destructive diseases in humans and other animals, including malaria, toxoplasmosis, and cryptosporidiosis. Enolase proteins from *Plasmodium* and *Toxoplasma* contain two insertions that otherwise are found only in land plants (7–10). Apicomplexa recently have been found to harbor a relic plastid homologous to the chloroplast of plants and algae (11, 12), and this plastid is now known to have originated through the secondary endosymbiotic incorporation of a photosynthetic eukaryote by an ancestor of apicomplexa (13–15). Several putatively plant-like enzymes now have been described in apicomplexa, and many of these enzymes do seem to be derived from the plastid (14–16). In light of these findings, the shared insertions in apicomplexan and plant enolases have been interpreted as revealing some relationship between apicomplexan and plant enolases (7–9), hinting that perhaps a lateral transfer between the algal endosymbiont and the apicomplexan host took place (10).

We have surveyed the distribution of these insertions in enolases from several key eukaryotic groups and compared this distribution with enolase phylogeny to determine whether apicomplexan enolases bear any relationship with those of plants or algae. We find that lateral transfer has been an important force in the evolution of eukaryotic enolases, being responsible for the origin of enolases in cryptomonads, *Chlorarachnion*, and *Arabidopsis*, but that the apicomplexan case is unusual. We find that the distribution of insertions in enolase is in direct conflict with the phylogeny of the molecule, suggesting that the insertions do not simply reflect the evolutionary history of the gene. Instead, it seems that the enolase insertions are positively misleading either because of multiple loss events or because of lateral transfer followed by recombination, thereby generating a mosaic enolase gene. Either explanation suggests that insertion and deletion data must be interpreted with caution.

EVOLUTION

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF348914–AF348938).

*To whom reprint requests should be sent at the present address: Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4. E-mail: pkeeling@interchange.ubc.ca.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

## Materials and Methods

**Cloning and Sequencing of Enolase Genes.** Sequences were amplified from five ciliates (*Paramecium tetraurelia* 51.s, *Paramecium multimicronucleatum*, *Tetrahymena thermophila* B4, *Tetrahymena bergeri*, and *Colpidium aqueous*), two red algae (*Prionitis lanceolata* and *Mastocarpus papillatus*), five chlorophyte green algae (*Chlamydomonas reinhardtii*, *Scenedesmus rubescens* CCAP232/1, *Nephroselmis olivacea* NIES-484, *Pycnococcus provasolii* CCMP 1203, and *Pedinomonas minor* UTEX LB 1350), three charophyte green algae (*Chara corallina* X-656, *Nitella opaca* X-867, and *Nitellopsis obtusa* X-854), one chlorarachniophyte (*Chlorarachnion* sp. CCMP 621), and two cryptophytes (*Guillardia theta* CCMP 327 and *Rhodomonas salina*). In most cases, 80–90% of the coding sequence was amplified by using either 5′ primers AGCGGCAACCCGACNGTNGARGTNGA or CCGGTCGACCGGNATHTAYGARGC with 3′ primer GCGCTCGCGRCANGGNGCNCCNGTYTT. Exceptions to this treatment were *Scenedesmus, Nephroselmis*, and one paralog from each of *Prionitis* and *Mastocarpus*, where only a small portion of the coding sequence surrounding the inserts could be obtained by using the same 5′ primer and the 3′ primer GGCCGGCAGDATCATRAAYTCYTG. The same portion of the enolase gene was amplified from genomic DNAs from *Chlamydomonas* (to confirm and extend a published, partial cDNA sequence) and from *Chlorarachnion.* The remainder of the *Chlorarachnion* gene was then amplified from a cDNA library, and this transcript was found to match the exonic sequences of the genomic fragment exactly. All amplifications were carried out with a denaturation of 2 min at 94°C followed by 35 cycles of denaturation of 30 sec at 92°C, a 30-sec annealing at 45°C, and a 60-sec extension at 72°C, all followed by a 2-min extension at 72°C. Amplification products were cloned with pCR2.1 (Invitrogen), and multiple copies were sequenced on both strands. In several instances, multiple distinct copies of the gene were found; these are distinguished in the figures by numbers following the organism names.

**Phylogenetic Analysis.** Enolase phylogeny was inferred from 372 clearly alignable amino acid positions (the alignment is available upon request of P.J.K.). Maximum likelihood distances were calculated with PUZZLE Version 4.0.1 (17) using the Jones–Taylor–Thornton substitution matrix and amino acid frequency estimated from the data. Site–site rate variation was modeled on a gamma distribution with eight rate categories plus invariant sites, and the shape parameter was estimated from the data. Bootstrap distances were calculated with PUZZLE and PUZZLEBOOT Version 1.03 (by M. Holder and A. Roger, available at http://www.tree-puzzle.de/#puzzleboot) by using the settings described above and the $\alpha$-parameter estimated from the original alignment. Trees were constructed with BioNJ (18), Fitch–Margoliash (19), and WEIGHBOR (20). Maximum likelihood trees were inferred by using quartet puzzling with 10,000 puzzling steps and the PUZZLE settings described above (this tree did not differ significantly from the distance trees and is not shown). Kishino–Hasegawa tests (21) were performed on constrained trees by using PUZZLE with the same settings as those used for tree construction. Data sets both including and excluding the insertions were tested and found to vary only insignificantly. For constrained trees, all groups labeled in Fig. 2 were constrained to be monophyletic and left unresolved, and the overall topology of these groups also was left unresolved. The alveolates were moved to each major branch of this test tree, and were left among the unresolved comb as well (tests also were done without including this unresolved tree, but no difference was observed).

Recombination sites were sought with two methods by using an alignment that excluded the insertions themselves. First, phylogenetic profiles of different regions of the gene were compared by using PHYLPRO (22), which revealed no detectable incoherence between different regions of the alignment when tested with a variety of settings. This finding is not surprising because the event must have taken place in the very ancient past to achieve its present widespread distribution throughout these two lineages, and because even recent gene conversion events often are very difficult to detect (23). Second, the tree topologies analyzed by Kishino–Hasegawa tests (see above) were compared by using a 20-aa sliding window over 10-aa increments. In most cases, the best tree placed alveolates with other protozoa, but in the two windows surrounding the site of the insertions (not including the insertions), the best tree placed the alveolates with the streptophytes. No tree ever was favored by a significant margin in the Kishino–Hasegawa test, which is hardly surprising considering the short window used.

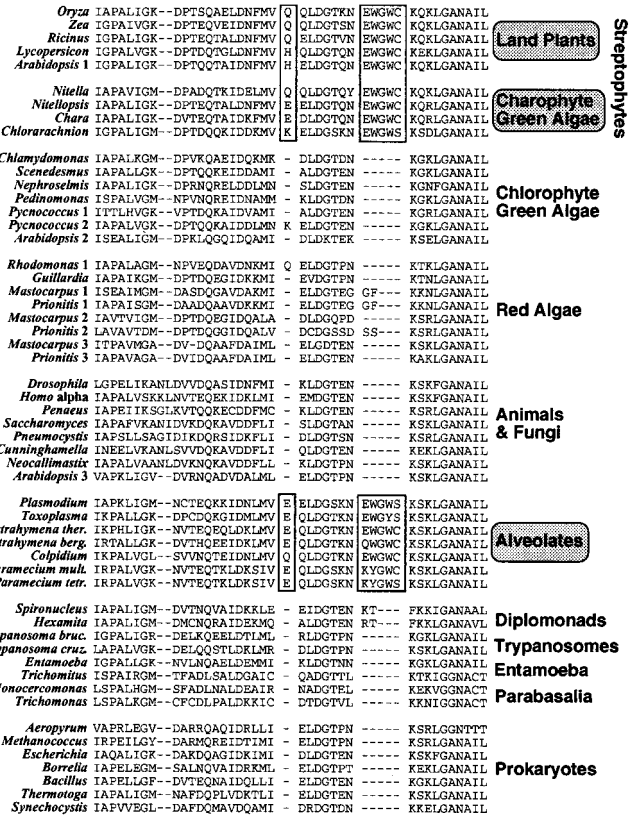


**Fig. 1.** Alignment of the enolase region corresponding to amino acid residues 73–118 of the *Oryza* gene. The two insertions shared by land plants, charophytes, and alveolates are boxed. *Arabidopsis* genes 2 and 3 are shown grouped with chlorophytes and with animals and fungi, respectively, because these genes occupy these positions in all phylogenetic analyses (see text for discussion).

## Results and Discussion

**Distribution of Insertions and Enolase Phylogeny.** Enolase genes were characterized from multiple representatives of several groups of eukaryotes specifically related to either plants or apicomplexa: charophyte green algae, chlorophyte green algae, red algae (all related to land plants), and ciliates (related to apicomplexa). Enolase genes also were characterized from two other eukaryotic lineages that contain plastids of secondary endosymbiotic origin, *Chlorarachnion* and cryptomonads (24). A partial alignment of these and other enolases (Fig. 1) reveals that the insertions in question are highly restricted in distribution: they are found only in streptophytes (land plants and charophyte

green algae) and alveolates (apicomplexa and ciliates), but not in chlorophytes, red algae, or any other eukaryote (with the exception of *Chlorarachnion*, which is discussed below). Mapping the insertion region to the crystal structure of yeast enolase (25–26) shows that the insertions are located in a loop, and not on the surface where enolase monomers interact in forming homodimers (25–26), suggesting that this region is amenable to alterations in size and shape. Indeed, insertions are found at the same locations in diplomonads, in some genes from red algae, and in one gene from a green alga, further implying that these regions, like others in enolase (27), are prone to insertions and deletions. These latter insertions are distributed only sporadically within the lineages in which they are found, and, in the case of the second insertion region, display no sequence similarity between distantly related genes, suggesting that they are the result of independent events in isolated lineages. Conversely, the two streptophyte/alveolate insertions are present throughout both lineages. The first insertion is not compelling on its own, as it is only 1 aa in size, is not universally conserved, and is also found in *Guillardia* and in one of the two *Pycnococcus* genes. However, the 5-aa insertion is quite striking, because not only is it highly conserved across both streptophytes and alveolates but also it usually contains two tryptophans—the rarest amino acid (Fig. 1). The likelihood of such a conserved insertion arising at the same position in the same protein independently is very low; so, even with the greatly expanded sampling of eukaryotic enolases, the original conclusion that at least the 5-aa insertion (and probably both insertions) are homologous is almost certainly true (7–9).

Normally, the shared presence of highly conserved and presumably homologous insertions such as these should reveal some relationship between the genes that bear them. However, when the phylogeny of the enolase molecule is examined, there does not seem to be any such relationship (Fig. 2). Here, neither land plants nor charophytes are specifically related to alveolates, but instead, both are closely related to chlorophytes and red algae, in agreement with their evolutionary position based on other data (28). Alveolates, on the other hand, branch with other protozoa, but their exact position is not well resolved. It has been suggested that this phylogeny is poorly resolved, and apicomplexa actually should branch with plants (10); however, these conclusions were made before data were available from ciliates, charophytes, or red algae, and with only *Chlamydomonas* to represent the chlorophytes. To confirm that the relationship between alveolate and streptophyte enolases is indeed very distant, alveolates were moved to 18 alternative positions in a partially unresolved tree (the positions are shown as filled dots on nodes of Fig. 2; however, it should be remembered that the test trees were partially unresolved constructions, as described in the *Materials and Methods*), and these positions were compared statistically by using the Kishino–Hasegawa test (21). Alternatives that placed the alveolates with any of the protozoan groups, with animals and fungi, as sister to the clade composed of red algae, green algae, and land plants, or as sister to the green algae and land plants, could not be rejected at the 5% confidence level. Conversely, topologies that placed alveolates with *Arabidopsis* enolase 3, with animals or fungi specifically, or with red algae or green algae were rejected. Most importantly, all topologies in which alveolates were specifically related to streptophytes as a whole (including the *Chlorarachnion* sequence), or to any subgrouping of streptophytes, were rejected: those topologies that placed alveolates as sister to charophytes or land plants were rejected at the 1% level, and the topologies where alveolates were sister-group to streptophytes as a whole, as sister-group to charophytes and land plants, or with the *Chlorarachnion* sequence alone, were rejected at the 5% level (in fact, all three topologies were rejected at a confidence level of 2%). In short, although the exact position of alveolates is not well resolved in enolase phylogeny, they do not seem to be related to the only other lineages that contain the same two insertions.

**Origin of Alveolate Enolases.** The conflict between the distribution of enolase insertions and the phylogeny of enolase sequences cannot be explained simply by lateral transfer from a streptophyte (i.e., as an endosymbiont) to an alveolate host, or from an alveolate to a streptophyte. In either case, alveolates and streptophytes would form a clade in enolase phylogeny, but this topology is never recovered by phylogenetic analysis, either in this study or elsewhere (9, 10), and is rejected by Kishino–Hasegawa tests. Moreover, the topology of red algae, green algae, and land plants that is recovered in enolase phylogeny is the topology that one would expect based on other data (e.g., ref. 28). The likelihood of recovering this expected topology by chance alone is remote, so it is much simpler to conclude that enolase is recovering the correct topology of at least these organisms. Similarly, postulating the presence of paralogous enolase families (one containing insertions and the other lacking insertions) fails to explain this conflict for the same reason (i.e., insert-containing paralogs should be related in the phylogeny, but they are not). We have dismissed already as very unlikely the possibility that the insertions arose independently in alveolates and streptophytes (see preceding section). This leaves two explanations for the incongruence between insertions and phylogeny: either the insertions have been deleted at exactly the same positions at least twice independently, or they have been transmitted between distantly related enolases.

For the former explanation to be the case, the insertions would have originated in a common ancestor of alveolates and streptophytes. Then, both insertions would have been lost independently in red algae, chlorophytes, and also potentially in other protist lineages, unless one assumes that alveolates are more closely related to red and green algae than any other lineage for which enolase is known (as well as any unexamined protistan lineage whose enolase lacks the two insertions). Multiple independent loss of one such insertion, especially the single amino acid insertion, is relatively easy to imagine; however, multiple independent loss of two insertions in the same lineages is highly coincidental. Moreover, the loss of the larger insertion would require the deletion of exactly the same 15 nucleotide positions in multiple lineages (because the sequence of the flanking regions is highly conserved), and, again, this event would have to be coincident with the independent loss of the single amino acid insertion.

Alternatively, and in our view the most likely scenario, the conflict between the distribution of the enolase insertions and the phylogeny of the gene in which they reside could be explained by the insertions having been horizontally transmitted between these lineages independently of the rest of the gene by an ancient recombination event between two distantly related enolase genes. Potential traces of such an event can still be detected in the alignment; for example, the conserved valine immediately upstream of the single amino acid insertion, or the less conserved glutamic acid or glutamine immediately downstream. Moreover, when a sliding window of the protein alignment was compared among the 18 tree topologies tested by Kishino–Hasegawa tests (see above), the region around the insertions was found to favor a streptophyte-alveolate relationship, whereas all other regions of the protein favored a relationship between alveolates and one of the other protist groups. Although both observations suggest some weak sequence similarity between streptophytes and alveolates in the insertion region, it should be pointed out that no actual sites of recombination can be detected with common techniques (see *Materials and Methods*). This result is hardly surprising, because such an event must have happened before the divergence both of land plants and charophytes and of apicomplexa and ciliates, both of which are thought to be ancient

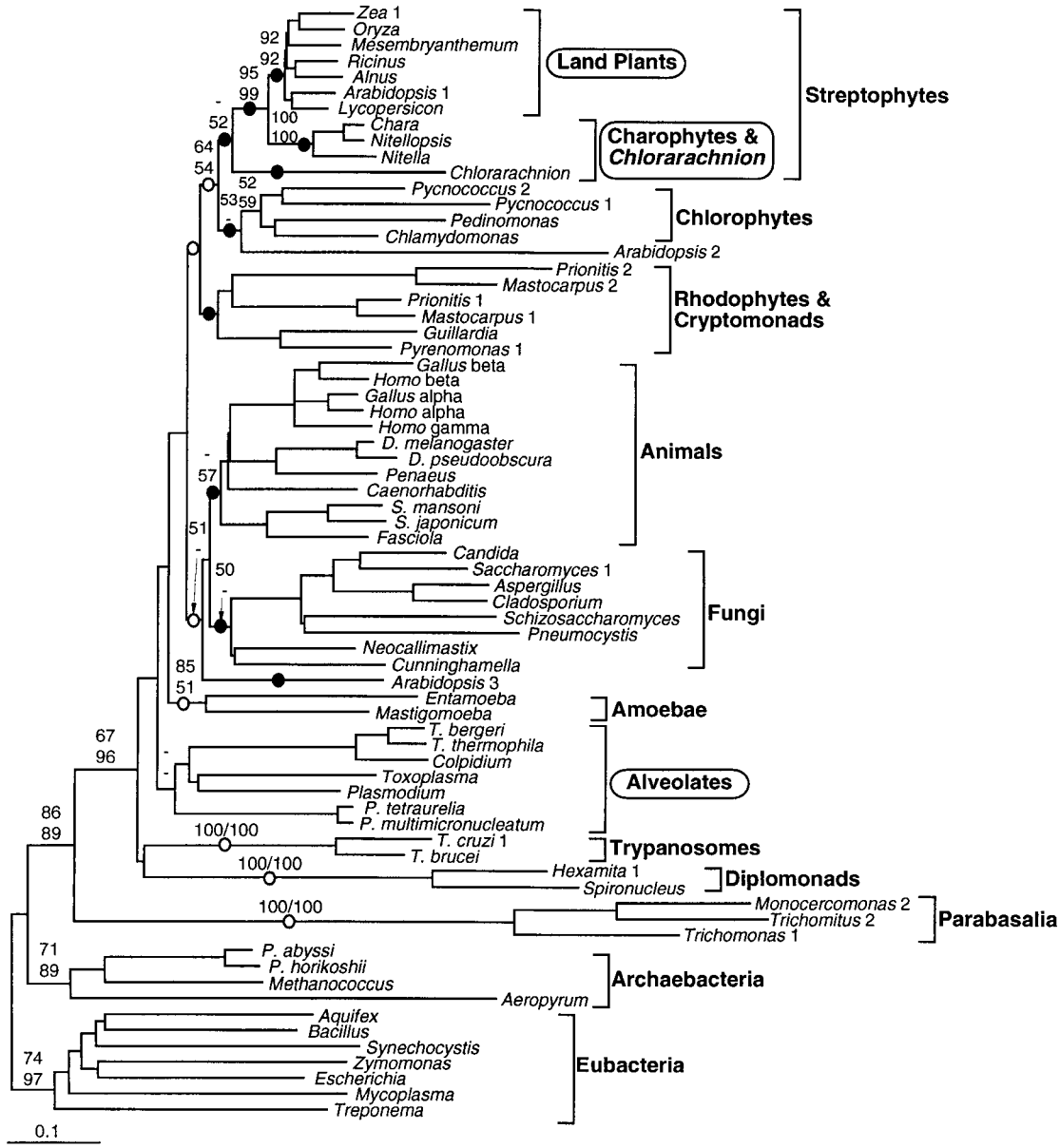


**Fig. 2.** Phylogeny of enolase. Numbers at nodes correspond to bootstrap support over 50% for major nodes from neighbor-joining (upper number) and Fitch–Margoliash (lower number) analyses. Groups of related organisms are named on the right, and the names of those with the two insertions highlighted in Fig. 1 are circled. Nodes where alveolates were positioned in partially unresolved trees and were tested by Kishino–Hasegawa tests are indicated as follows: ●, positions rejected at the 5% level (many were rejected at the 1% level—see text); ○, positions not rejected. The position of alveolates is not well supported by bootstrap analysis (and within alveolates, ciliates are not monophyletic), but their separation from land plants and charophytes is supported strongly by Kishino–Hasegawa tests (see text).

divergences. Such a span of time would almost inevitably erase any weak signal resulting from local sequence similarity (23), so that the only detectable signal that might remain would be the comparatively highly conserved presence or absence characteristic of these protein insertions.

Although we favor the scenario of mosaicism caused by subgenic lateral transfer, we cannot conclusively differentiate between this explanation and parallel loss of both insertions, because, in either case, the events took place a very long time ago, and because both scenarios are consistent with the distribution of the insertions and with enolase phylogeny. On one hand, it is impossible to guess at the likelihood that both insertions would be lost in two lineages independently, although the fact that these two losses would also have to have taken place in two related lineages (red and green algae) is highly suspect. On the other hand, the frequency of recombination between such distantly related sequences is certainly quite low. There are numerous studies on recombination barriers between bacterial species that show an exponential decrease in recombination frequency as sequence diversity grows (e.g., ref. 29). However, these studies also show that recombination frequencies are affected by the presence of blocks of similarity rather than overall similarity (29) and are dramatically altered by the activity or inactivity of cellular processes such as mismatch repair and SOS (30), so it is impossible to rule out recombination between distantly related sequences. Moreover, even when recombination frequencies are very low, rare events are expected to happen at some frequency. Indeed, it has recently been proposed that

small subunit rRNA genes in two genera of Gram-positive bacteria have recombined, resulting in mosaic rRNA genes (31) similar to those we propose for enolase.

The case for enolase mosaicism will become weaker if additional seemingly unrelated lineages are found to contain these two insertions, or if alveolates turn out to be particularly closely related to red algae, green algae, and plants. In contrast, parallel deletion will become increasingly unlikely if and when other lineages of eukaryotes that are specific relatives of streptophytes or alveolates (to the exclusion of the other group) are found to lack both enolase insertions. There are a multitude of protistan lineages whose enolase genes have not yet been sequenced; however, for very few of these lineages is their position relative to streptophytes and alveolates yet established. The best starting point is the heterokonts, which increasingly seem to be the sister group to alveolates (32–34).

In any event, the insertions in alveolate enolases probably have nothing to do with the presence of the algal endosymbiont that gave rise to the apicomplexan plastid, as has been assumed previously (7–10). If the insertions have been lost in red algae and chlorophytes, then their presence in alveolates merely suggests a relatively close relationship between the alveolate host lineage and the red/green algal clade. Alternatively, in the case of lateral transfer and recombination, a link between enolase and the apicomplexan plastid is possible only if the donor was the plastid endosymbiont. However, several lines of evidence favor a red algal origin (34–37) for the apicomplexan plastid, not a green algal origin (13). A more plausible explanation is a variation on the recent suggestion that eukaryotic genomes could laterally acquire DNA at substantial rates by a gene transfer ratchet, whereby the continual uptake of DNA from food sources leads to occasional but inevitable gene replacement events (38). In this case, an ancestor of the alveolates would have ingested an ancestor of the charophyte green algae, and the enolase gene from this food source, instead of replacing entirely the predator's own enolase, would have recombined with it, resulting in a mosaic gene. We favor a transfer from charophyte to alveolate because the alveolate ancestor was very likely to be heterotrophic, whereas the charophyte ancestor almost certainly was not.

**Lateral Transfer of Enolase in *Chlorarachnion*, Cryptomonads, and *Arabidopsis*.** A counterpoint to the conflict in alveolate enolases is found in the phylogenetic positions of *Chlorarachnion* and cryptomonad genes. Like apicomplexa, these two groups contain plastids of secondary endosymbiotic origin: the *Chlorarachnion* plastid is derived from a green alga and the cryptomonad plastid is derived from a red alga, but unlike apicomplexa, these two endosymbionts retain vestigial nuclei, or nucleomorphs (24). The characteristics of the *Chlorarachnion* and cryptomonad enolases are all consistent with their being encoded by the (host) nuclear genome in both cases. For instance, cryptomonad and *Chlorarachnion* enolases have a moderate G+C content and contain numerous, large spliceosomal introns, exactly as one would expect of host nuclear genes (39, 40). Conversely, nucleomorph genes from both groups typically are highly A+T-rich, cryptomonad nucleomorph genes contain few introns, and *Chlorarachnion* nucleomorph genes contain distinctive 18-, 19-, or 20-bp introns (24). Yet, in enolase phylogeny, *Chlorarachnion* and cryptomonad genes branch with the green and red algae, respectively, exactly where one would expect genes from their endosymbionts to branch.

The phylogenetic positions of cryptomonad and *Chlorarachnion* enolases suggest that these genes are derived from the genomes of their endosymbionts, as was previously suggested of apicomplexa. Although it is possible that the products of *Chlorarachnion* and cryptomonad enolase genes function in the endosymbiont, there is no evidence for primary carbon metabolism in the remnant cytosol of either *Chlorarachnion* or cryptomonad endosymbionts, so it is more likely that these enolases have assumed a function in the host cytosol [the complete sequence of the *Chlorarachnion* enolase cDNA also lacks an apparent targeting peptide that could direct the protein to the symbiont (P.J.K., unpublished data)]. Also, it is possible that these enolases were derived from some source other than the endosymbionts (perhaps a food source), but the presence of green and red algal genes in the host genomes of *Chlorarachnion* and cryptomonads, respectively, would be highly coincidental if they were not simply derived from their endosymbionts. These results demonstrate that the process of endosymbiotic gene replacement observed with bacterial symbionts (2) likely holds for eukaryotic endosymbionts as well. Both transient and permanent eukaryotic endosymbionts are quite abundant in nature; however, only a handful of genes have been sampled from the genomes of the hosts of such partnerships, and until more data are available, the importance of this process will remain unclear. If enolase is at all typical, then the host genomes in these endosymbiotic systems may have been widely invaded by endosymbiont genes.

If these genes are derived from the endosymbionts, then the position of the *Chlorarachnion* enolase has particular significance, because this gene branches specifically in a position expected of a charophyte (Fig. 2), contains insertions homologous to those of charophytes and land plants (Fig. 1), and shares two conserved intron positions with plants (the three charophytes sampled all lack introns, and so are not comparable). In sum, the *Chlorarachnion* enolase clearly seems to be derived from a charophyte, which is potentially very important evidence for the nature of the *Chlorarachnion* secondary endosymbiont. Presently there is strong biochemical (photosynthetic pigment) and molecular evidence that the *Chlorarachnion* plastid endosymbiont was a green alga, but it is unclear exactly what kind of green alga because the relic genome of the endosymbiont is highly divergent, resulting in uncertain phylogenies. To date, endosymbiont and plastid gene phylogenies have suggested relationships to practically every possible subgroup of chlorophyte (for a review and an example see ref. 41), but the *Chlorarachnion* enolase is unambiguously derived from a charophyte. If the *Chlorarachnion* enolase is derived from the plastid endosymbiont, then the endosymbiont must have been a charophyte.

Last, the unusual phylogenetic positions of *Arabidopsis* enolase genes 2 and 3 need to be addressed, because these genes do not branch with other land plant enolases (Fig. 2), and they also lack the streptophyte insertions (Fig. 1). *Arabidopsis* 2 branches with chlorophyte green algae, and *Arabidopsis* 3 branches consistently, but with little statistical support, at the base of the animals and fungi (note also from Fig. 1 that this gene lacks the animal–fungal insertions). In addition, both these genes encode N-terminal leaders. The presence of this leader in enolase 2 led to the suggestion that this enolase is plastid-targeted in *Arabidopsis* (10). Certainly, this protein may presently function in the plastid; however, its origin remains mysterious, as does the presence of enolase 3 and its leader. In addition, we have examined the *Chlamydomonas* expressed sequence tag database and found only transcripts from a single enolase corresponding to the gene sequenced here, and this gene also encodes a short leader (data not shown). Determining clearly the cellular location of the products of these genes is critical before attempting to make any assertions as to their roles in the cell. Nevertheless, from the phylogeny, we can make a few general conclusions about their evolutionary origin. Most importantly, both the *Arabidopsis* enolase 2 and 3 genes most likely are derived from lateral gene transfer, one from a chlorophyte and the other from some ancestor of animals and fungi. These constitute two of the

relatively few even moderately clear instances of eukaryote–eukaryote gene transfer.

**Concluding Remarks.** The shared presence of homologous insertions in the enolases of apicomplexan parasites and land plants has little or no relationship to the presence of a plastid in apicomplexa. Rather, the distribution of insertions in enolase and the phylogeny of enolase are at odds with one another, a phenomenon with more far-reaching implications on how we perceive gene and genome evolution and molecular phylogeny. First, insertions and deletions are not unconditionally reliable markers of evolutionary relationships; even if free of homoplasy, they can, indeed, be transmitted between lineages more or less independently of the genes in which they are situated. This observation does not negate the use of insertions or deletions to reconstruct ancient events, but does emphasize the need to weigh the merit of such characters in conjunction with the phylogeny of the gene in which they are found (27, 42–45) and with other external information (46). Second, if alveolate enolases are products of recombination, it is doubtful that this recombination happened *because* these insertions were present. Rather, it is likely that it was detected only because the insertions act as a flag, drawing attention to the incongruent evolutionary histories between different parts of the gene. It follows that subgene-level transfer may have taken place in other genes as well, but without any flags to mark these events, they go undetected.

One of the prevailing notions about early genome evolution in recent years has been the importance of lateral gene transfer. Some have argued that ancient lateral transfer between genomes has led to a tree of life that does not branch like an ordinary tree, but rather weaves a web of diverging and intersecting branches that has smudged any crisp phylogenetic definition of the genome (5, 6). If the insertions were indeed transmitted horizontally between two distantly related genes, then evolutionarily distant lateral transfer can be extended to the subgene level, which, in turn, suggests that individual gene trees also may be reticulate in nature at the most ancient levels. As lateral gene transfer raised questions as to the phylogenetic definition of the genome, enolase raises new questions as to the phylogenetic definition of the gene itself.

We thank F. Lang, J. Preer, M. Strueder-Kypke, D. Lynn, V. Huss, S. J. Surzycki, R. L. Chapman, and S. Meiers for generously providing DNA, G. Fast for providing live material from red algae, K. Adams for management, and D. Roos for sharing unpublished data. This work was supported by National Institutes of Health Grant RO1 GM35087 (to J.D.P.) and a Medical Research Council of Canada Fellowship (to P.J.K.). P.J.K. is a scholar of the Canadian Institute for Advanced Research.

1. Gray, M. W. (1993) *Curr. Opin. Genet. Dev.* **3,** 884–890.
2. Martin, W. & Schnarrenberger, C. (1997) *Curr. Genet.* **32,** 1–18.
3. Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000) *Nature (London)* **405,** 299–304.
4. Eisen, J. A. (2000) *Curr. Opin. Genet. Dev.* **10,** 606–611.
5. Doolittle, W. F. (1999) *Trends Cell Biol.* **9,** 5–8.
6. Doolittle, W. F. (1999) *Science* **284,** 2124–2129.
7. Read, M., Hicks, K. E., Sims, P. F. & Hyde, J. E. (1994) *Eur. J. Biochem.* **220,** 513–520.
8. Hyde, J. E., Sims, P. F. G. & Read, M. (1994) *Parasitol. Today* **10,** 25.
9. Dzierszinski, F., Popescu, O., Toursel, C., Slomianny, C., Yahiaoui, B. & Tomavo, S. (1999) *J. Biol. Chem.* **274,** 24888–24895.
10. Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J. A., Albert, M. A., Sensen, C. W., Gaasterland, T., Muller, M., Michels, P. & Martin, W. (2000) *Mol. Biol. Evol.* **17,** 989–1000.
11. Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W. & Williamson, D. H. (1996) *J. Mol. Biol.* **261,** 155–172.
12. McFadden, G. I., Reith, M. E., Munholland, J. & Lang-Unnasch, N. (1996) *Nature (London)* **381,** 482 (lett.).
13. Köhler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D. & Roos, D. S. (1997) *Science* **275,** 1485–1489.
14. Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S. & McFadden, G. I. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 12352–12357.
15. McFadden, G. I. & Roos, D. S. (1999) *Trends Microbiol.* **7,** 328–333.
16. Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., *et al.* (1999) *Science* **285,** 1573–1576.
17. Strimmer, K. & Von Haeseler, A. (1996) *Mol. Biol. Evol.* **13,** 964–969.
18. Gascuel, O. (1997) *Mol. Biol. Evol.* **14,** 685–695.
19. Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Univ. of Washington, Seattle).
20. Bruno, W. J., Socci, N. D. & Halpern, A. L. (2000) *Mol. Biol. Evol.* **17,** 189–197.
21. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29,** 170–179.
22. Weiller, G. F. (1998) *Mol. Biol. Evol.* **15,** 326–335.
23. Drouin, G., Prat, F., Ell, M. & Clarke, G. D. (1999) *Mol. Biol. Evol.* **16,** 1369–1390.
24. McFadden, G. I., Gilson, P. R., Douglas, S. E., Cavalier-Smith, T., Hofmann, C. J. & Maier, U. G. (1997) *Trends Genet.* **13,** 46–49.
25. Zhang, E., Brewer, J. M., Minor, W., Carreira, L. A. & Lebioda, L. (1997) *Biochemistry* **36,** 12526–12534.
26. Larsen, T. M., Wedekind, J. E., Rayment, I. & Reed, G. H. (1996) *Biochemistry* **35,** 4349–4358.
27. Keeling, P. J. & Palmer, J. D. (2000) *Nature (London)* **405,** 635–637.
28. Moreira, D., Le Guyader, H. & Philippe, H. (2000) *Nature (London)* **405,** 69–72.
29. Vulic, M., Dionisio, F., Taddei, F. & Radman, M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9763–9767.
30. Majewski, J. & Cohan, F. M. (1999) *Genetics* **153,** 1525–1533.
31. Wang, Y. & Zhang, Z. (2000) *Microbiology* **146,** 2845–2854.
32. Van de Peer, Y., Baldauf, S. L., Doolittle, W. F. & Meyer, A. (2000) *J. Mol. Evol.* **51,** 565–576.
33. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000) *Science* **290,** 972–977.
34. Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001) *Mol. Biol. Evol.* **18,** 418–426.
35. McFadden, G. I. & Waller, R. F. (1997) *BioEssays* **19,** 1033–1040.
36. Blanchard, J. L. & Hicks, J. S. (1999) *J. Eukaryotic Microbiol.* **46,** 367–375.
37. Zhang, Z., Green, B. R. & Cavalier-Smith, T. (1999) *Nature (London)* **400,** 155–159.
38. Doolittle, W. F. (1998) *Trends Genet.* **14,** 307–311.
39. Keeling, P. J., Deane, J. A. & McFadden, G. I. (1998) *J. Eukaryotic Microbiol.* **45,** 561–570.
40. Deane, J. A., Fraunholz, M., Su, V., Maier, U.-G., Martin, W., Durnford, D. G. & McFadden, G. (2000) *Protist* **151,** 239–252.
41. Ishida, K., Cao, Y., Hasegawa, M., Okada, N. & Hara, Y. (1997) *J. Mol. Evol.* **45,** 682–687.
42. Rivera, M. C. & Lake, J. A. (1992) *Science* **257,** 74–76.
43. Baldauf, S. L. & Palmer, J. D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11558–11562.
44. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7749–7754.
45. Gupta, R. S. & Johari, V. (1998) *J. Mol. Evol.* **46,** 716–720.
46. Keeling, P. J. & McFadden, G. I. (1998) *Trends Microbiol.* **6,** 19–23.