

Lateral Gene Transfer of a Multigene Region from Cyanobacteria to Dinoflagellates Resulting in a Novel Plastid-Targeted Fusion Protein

Ross F. Waller,¹ Claudio H. Slamovits, and Patrick J. Keeling

Canadian Institute for Advanced Research, Botany Department, University of British Columbia, Vancouver, British Columbia, Canada

The number of cases of lateral or horizontal gene transfer in eukaryotic genomes is growing steadily, but in most cases, neither the donor nor the recipient is known, and the biological implications of the transfer are not clear. We describe a relatively well-defined case of transfer from a cyanobacterial source to an ancestor of dinoflagellates that diverged before *Oxyrrhis* but after *Perkinsus*. This case is also exceptional in that 2 adjacent genes, a paralogue of the shikimate biosynthetic enzyme AroB and an O-methyltransferase (OMT) were transferred together and formed a fusion protein that was subsequently targeted to the dinoflagellate plastid. Moreover, this fusion subsequently reverted to 2 individual genes in the genus *Karlodinium*, but both proteins maintained plastid localization with the OMT moiety acquiring its own plastid-targeting peptide. The presence of shikimate biosynthetic enzymes in the plastid is not unprecedented as this is a plastid-based pathway in many eukaryotes, but this species of OMT has not been associated with the plastid previously. It appears that the OMT activity was drawn into the plastid simply by virtue of its attachment to the AroB paralogue resulting from their cotransfer and once in the plastid performed some essential function so that it remained plastid targeted after it separated from AroB. Gene fusion events are considered rare and likely stable, and such an event has recently been used to argue for a root of the eukaryotic tree. Our data, however, show that exact reversals of fusion events do take place, and hence gene fusion data are difficult to interpret without knowledge of the phylogeny of the organisms—therefore their use as phylogenetic markers must be considered carefully.

Introduction

Lateral gene transfer (LGT) or horizontal gene transfer is the movement of genetic material between distantly related genomes resulting in the sudden appearance of novel genetic diversity in a lineage or the replacement of an existing gene by a distantly related homologue. The impact of this process has been debated extensively in prokaryotes because comparative genomics revealed apparently large rates of interspecies genetic exchange in many systems. However, its impact on eukaryotes has been less clear, in part due to the relative dearth of genomic data from diverse eukaryotes as compared with prokaryotes. There have been many individual cases of LGT reported from a variety of eukaryotes (recently summarized in Andersson 2005) and a few cases where genome-wide analyses have indicated a widespread introduction of new genes in a genome (Archibald et al. 2003; Bergthorsson et al. 2003; Loftus et al. 2005). In most reported cases, a prokaryotic gene has moved into the eukaryotic genome, and these are also the most easily detected events because the phylogenetic position of the eukaryotic gene stands out. Nevertheless, a growing number of eukaryote-to-eukaryote LGT cases are also emerging (Bergthorsson et al. 2003; Keeling and Inagaki 2004; Andersson 2005), but because these are even more heavily dependent on a good representation of eukaryotic diversity in order to be detected, they are still relatively few in number.

In the majority of cases, the source of a transferred gene is only known in broad terms because transferred genes rarely show a close relationship with a well-defined lineage that may represent the possible donor. Because of this, little can be said about the nature of the event that led to the transfer or its impact on the recipient cell because the

event took place too long ago to reconstruct it with any detail, its distribution in eukaryotes is not clear, and its potential functional implications impossible to guess, in particular, in cases where the gene replaces an existing homologue.

Here we describe a case of LGT from bacteria to eukaryotes where the nature of both the donor and the recipient are clear. The transfer involves 2 genes, a paralogue of the shikimate biosynthetic enzyme 3-dehydroquinate synthase (AroB) and an O-methyltransferase (OMT) sharing greatest similarity to caffeoyl CoA 3-O-methyltransferases. The genes were cotransferred to a well-defined subgroup of the dinoflagellate lineage where they formed a fusion protein. This protein was targeted to the plastid, potentially introducing novel biochemical functionalities to the organelle. Intriguingly, the OMT was later cleaved from the fusion partner in the genus *Karlodinium*, but still targeted to the plastid, indicating that it had acquired an essential function as part of the fusion protein. Overall, the evolution of this protein in dinoflagellates represents one of the best resolved prokaryote-to-eukaryote LGT events presently known and also sheds light on the fusion and fission of genes in the nuclear genome and the acquisition of novel functions by the plastid, all thought to be very rare events in evolution.

Materials and Methods

Characterization of AroB Genes in Dinoflagellates

All new sequences were identified by searching ongoing or complete expressed sequence tag (EST) projects on *Heterocapsa triquetra* strain CCMP 449 (Patron et al. 2005), *Karlodinium micrum* CCMP 415 (Patron et al. 2006), and *Oxyrrhis marina* CCMP 1788 (EST sequencing in progress). AroB genes were detected by similarity searches using Blast algorithms and existing plastid-targeted or cytosolic AroB sequences from other eukaryotes or bacteria. All AroB-encoding ESTs were completely sequenced from both strands. These sequences were in turn used to search public databases for further similar sequences in all available taxa. OMT sequences were sought in the same way, and in the case of *Karlodinium*, independent

¹ Present address: School of Botany, University of Melbourne, Melbourne, Australia.

Key words: lateral gene transfer, dinoflagellates, gene fusion, gene fission.

E-mail: r.waller@unimelb.edu.au.

Mol. Biol. Evol. 23(7):1437–1443. 2006
doi:10.1093/molbev/msl008
Advance Access publication May 4, 2006

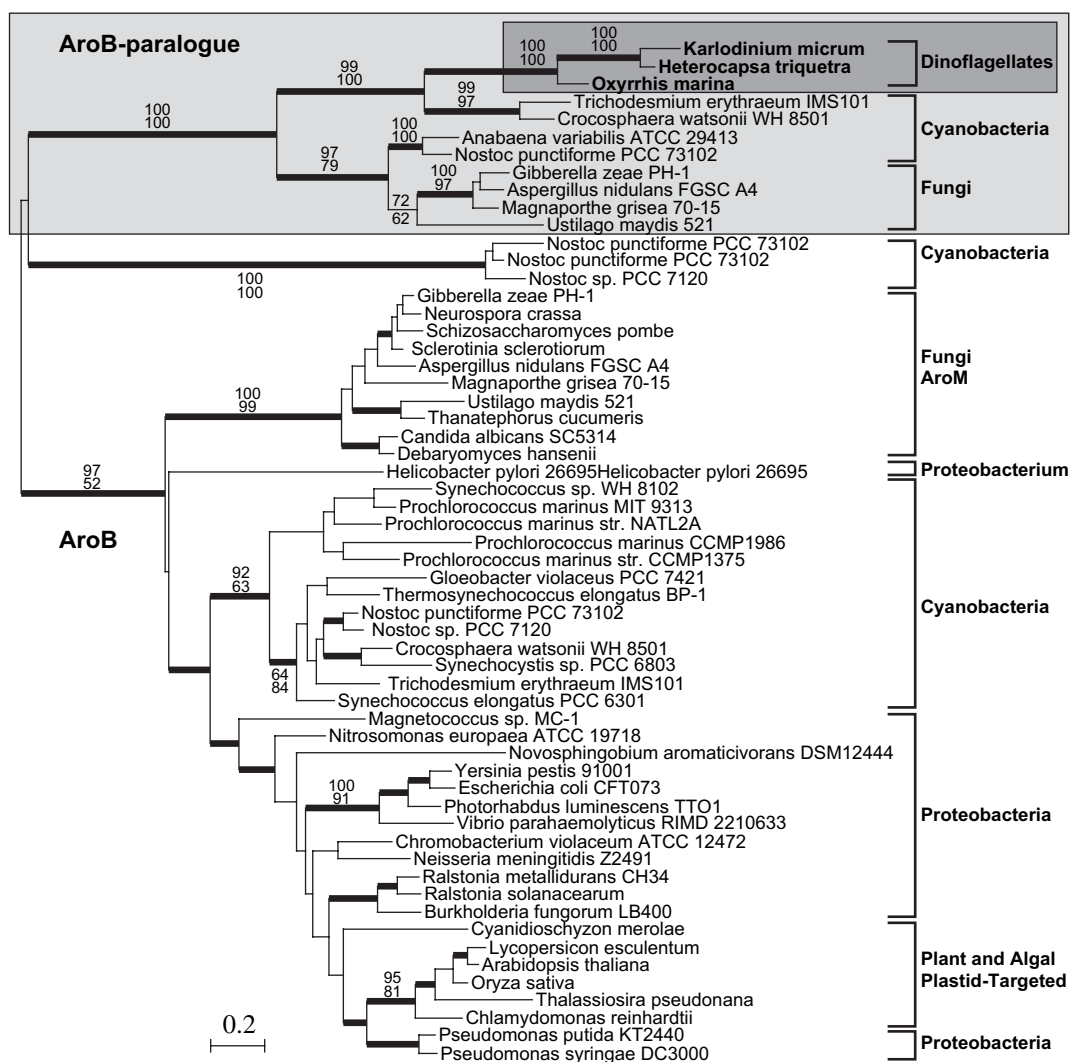


FIG. 1.—Protein maximum likelihood phylogeny of AroB and AroB paralogues. Numbers at nodes correspond to bootstrap support for major nodes over 50% (ML above, NJ below), and major groups are named to the right. Thickened nodes indicate Bayesian posterior probabilities of 100%. AroB paralogues are shown at the top of the tree, where a specific relationship between *Oxyrrhis* and the dinoflagellates received 100% bootstrap support, as did the relationship of this group with sequences from the cyanobacteria *Crocosphaera watsonii* and *Trichodesmium erythraeum* (Gene identifier numbers for all taxa are given in Supplementary data, Supplementary Material online).

OMT-encoding genes were discovered and completely sequenced. Accession numbers for new sequences are DQ517901–4. N-terminal regions of dinoflagellate genes were analyzed for the presence of plastid-targeting sequences as described previously for both *Heterocapsa* and *Karlodinium* (Patron et al. 2005, 2006). Briefly, signal peptides were inferred using SignalP (Bendtsen et al. 2004) and transit peptides predicted based on the properties known from plastid-targeted proteins of known plastid function (Patron et al. 2005, 2006). Cyanobacteria and fungal genomic data were recovered from existing public databases, and the region around the AroB paralogue was drawn to scale with the aid of Artemis (Rutherford et al. 2000).

Phylogenetic Analyses

AroB and OMT amino acid sequence alignments were constructed using ClustalX (Thompson et al. 1997) and

sequences available in public databases. In the case of AroB paralogues (fig. 1), all homologues related to those of dinoflagellates were gathered by extensive comparisons with public databases resulting in sequences only from cyanobacteria and fungi. Representative canonical AroB proteins from diverse taxa were included, and the fungal cytoplasmic AroB domain of the multifunctional AroM protein was extracted for several fungi and added to the alignment. Removal of ambiguously aligned sites resulted in an alignment of 234 characters (see also Supplementary data, Supplementary Material online). OMT are a very large and complex family of proteins, so a large preliminary alignment was constructed with 251 sequences. The subgroup including the dinoflagellate sequences was identified from preliminary analyses of this alignment, and these sequences were further analyzed in greater detail including a larger number of alignable sites (159). Because this protein is neither large nor highly conserved, these analyses are not conclusive. Alignments are available on request.

Phylogenetic analyses on these alignments were carried out using maximum likelihood, distance, and Bayesian analysis. Maximum likelihood trees were inferred using PhyML 2.4.4 (Guindon and Gascuel 2003). In both cases, site-to-site rate variation was modeled on a gamma distribution with 8 rate categories and invariable sites. The shape parameter alpha and proportion of invariable sites were estimated by PhyML using the WAG substitution matrix. Distances were calculated by Tree-Puzzle 5.2 (Strimmer and von Haeseler 1996) with the WAG substitution matrix, with 8 rate categories and invariable sites all estimated from the data. Bootstraps were calculated using puzzleboot (shell script by A. Roger and M. Holder, <http://www.tree-puzzle.de>) and the same conditions. Trees were inferred using WEIGHBOR 1.0.1a (Bruno et al. 2000). Alpha and Γ parameters for AroB were 1.15 and 0.08, respectively and for OMT were 1.82 and 0.02, respectively. Bayesian analysis was performed using MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) implementing the covarion model and using the WAG + Γ model. Metropolis-coupled Markov chain Monte Carlo was run from a random starting tree for 1 000 000 generations sampling every 1000 cycles. Four chains were run simultaneously, 3 heated and 1 cold, with the first 200 000 cycles (200 trees) discarded as the burn-in. Node posterior probabilities were determined from the consensus of the remaining 800 trees.

Results and Discussion

AroB and AroB-OMT Fusion Genes in Dinoflagellates

The origin and location of shikimate biosynthesis in apicomplexa has been a subject of some debate (Roberts et al. 1998; Keeling et al. 1999; Fitzpatrick et al. 2001), prompting us to search for homologous proteins in their sister group, the dinoflagellates. EST libraries from 3 phylogenetically distinct members of the dinoflagellate lineage, the peridinin-containing *H. triquetra* (6804 sequences), the fucoxanthin-containing *K. micrum* (16 548 sequences), and the nonphotosynthetic basal heterotroph *O. marina* (18 012 sequences) were searched for genes encoding proteins involved in shikimate biosynthesis. In all 3 species, genes with high similarity (60–63%) to known 3-dehydroquinate synthases, or AroBs, were identified.

To investigate whether these proteins are likely targeted to the plastid, as are homologues in some other algae, we examined the dinoflagellate genes for evidence of N-terminal leaders encoding plastid-targeting peptides. The characteristics of plastid-targeting peptides from both *Heterocapsa* and *Karlodinium* have been examined in some detail, both types contain novel and distinctive features (Patron et al. 2005, 2006) and both putative AroB proteins were found to be preceded by an extension matching these features. The *Heterocapsa* protein initiates with a hydrophobic domain consistent with the presence of a signal peptide, although not actually predicted to encode one according to the parameters of SignalP (Bendtsen et al. 2004). Following this, however, is a FVQP that fulfills the criteria of the FVAP motif shown to be common to most *Heterocapsa* transit peptides; an acid residue-poor and glutamine-rich region; and lastly a predicted transmembrane domain followed by basic residues, as expected for a stop-transfer

membrane anchor. These properties fulfill all the criteria of Class I dinoflagellate transit peptides (Patron et al. 2005) (Supplementary Figure 1a, Supplementary Material online). In *Karlodinium*, a plastid-replacement event has occurred, and this lineage now contains a tertiary plastid derived from an endosymbiosed haptophyte (Tengs et al. 2000; Yoon et al. 2005; Patron et al. 2006). Coincident with this event, the character of the plastid-targeting transit peptides has been significantly modified (Patron et al. 2006). The *Karlodinium* AroB protein that we describe here begins with a region predicted to be a signal peptide (Bendtsen et al. 2004) that is followed by a region consistent with the distinctive features of *Karlodinium* transit peptides (a tendency for elevated hydroxylated residues serine/threonine, glutamine, and arginine, but no pronounced decrease in acid residues: Supplementary Figure 1a, Supplementary Material online) (Patron et al. 2006), consistent with plastid targeting as seen for *Heterocapsa*. The N-terminus of *Oxyrrhis* AroB protein does not encode an apparent leader and lacks properties consistent with targeting to a plastid, in keeping with the current absence of evidence for a plastid in the nonphotosynthetic *Oxyrrhis*. When aligned with other AroB proteins, all 3 sequences also contain an additional domain at the N-terminus, in the case of *Heterocapsa* and *Karlodinium* following the plastid-targeting peptide. This region is approximately 180 residues in length and shows sequence conservation between the dinoflagellates but does not share similarity with any known protein (Supplementary Figure 1a, Supplementary Material online).

Most interestingly, the *Oxyrrhis* and *Heterocapsa* genes also have C-terminal extensions, and these sequences share a high similarity with caffeoyl CoA 3-O-methyltransferases found throughout the tree of life. The *Karlodinium* sequence lacks a homologous extension, so we sought a stand-alone copy of the gene in the EST data and found 2 similar copies of such an OMT with a high degree of identity. One *Karlodinium* OMT cDNA was truncated, but the other was apparently full length and very interestingly also encoded an N-terminal leader bearing the characters of a plastid-targeting leader (signal peptide and *Karlodinium*-type transit peptide: Supplementary Figure 1b, Supplementary Material online) (Patron et al. 2006). The *Karlodinium* AroB-coding sequence terminates at the expected location and bears no residue of the fusion moiety.

Lastly, we searched the existing genomic sequence data from *Perkinsus marinus* (<http://www.tigr.org/tdb/e2k1/pmg>), the sister group to dinoflagellates, for closely related homologues of the dinoflagellate AroB and OMT sequences. Although sequence matches were found, none showed high levels of conservation with the other dinoflagellate sequences, and no putative AroB and OMT sequences were linked. Further, the conserved 180 residue N-extension of the dinoflagellates returned no matches from the *Perkinsus* data.

These AroB data from dinoflagellates are quite distinct from AroB from apicomplexans where AroB exists as part of a cytosolic pentafunctional protein consisting of 5 proteins that participate in shikimate synthesis (Campbell et al. 2004). An OMT is not associated with this protein, and no plastid-targeted AroB has been identified. The arrangement of this fusion protein is very similar to that known

from fungi, implying the common origin of this cytosolic pathway.

LGT and Fusion of AroB and OMT

Although the AroB protein homologues of *Heterocapsa* and *Karlodinium* are apparently targeted to the plastid, phylogenetic analysis of a broad diversity of AroB sequences (fig. 1) showed that they are not closely related to plastid-targeted AroB from plants and other algae. Instead, they show a strong phylogenetic relationship to a divergent AroB paralogue otherwise known only from a handful of cyanobacteria and fungi. Each of these later taxa also encode a canonical AroB sequences (fig. 1), and the precise role of the paralogues is not known. The dinoflagellate genes form a strongly supported clade (100/100% bootstrap support), with *Oxyrrhis* at the base, as expected given previous analysis suggesting it to be a deeply diverging member of the dinoflagellate lineage (Saldarriaga et al. 2003; Leander and Keeling 2004). The dinoflagellate proteins in turn are specifically sisters with AroB paralogues from 2 cyanobacteria, *Crocospaera watsonii* WH 8501 and *Trichodesmium erythraeum* IMS101, again with very strong support (99/100%). This relationship is not consistent with the dinoflagellate proteins originating with other eukaryotic AroBs and suggests instead that they moved between cyanobacteria and dinoflagellates.

This case of LGT is strongly corroborated by the genomic context of the cyanobacterial genes. Figure 2 shows a schematic of the AroB paralogue phylogeny (from fig. 1) with the structure and known genomic context of AroB paralogues plotted on the phylogeny. In all 4 cyanobacterial genera currently known to encode this AroB paralogue, the AroB gene is immediately followed by an OMT gene with a high degree of similarity to the C-terminal OMT moiety of the dinoflagellate fusion protein. The AroB and OMT are generally very closely spaced (in *C. watsonii*, they are annotated as being overlapping genes in different coding frames, but reanalysis of the N-terminus of OMT suggests that this is not the case, and they are actually spaced by 10 bp). In other cyanobacterial genomes, AroB is not encoded upstream of an OMT, including the *Nostoc* genes that fall between AroB and AroB paralogue clades (fig. 1). In phylogenetic analysis of OMT (Supplementary Figure 2, Supplementary Material online), only *Karlodinium* formed a well-supported clade (98/97% bootstrap support) with the 4 cyanobacterial OMTs found downstream of AroB: the position of *Oxyrrhis* and *Heterocapsa* OMTs was not resolved with any support, although no strongly supported node separated them from the *Karlodinium*/cyanobacterial clade (Supplementary Figure 2, Supplementary Material online). The OMT protein is very short and not highly conserved, so the phylogeny cannot be conclusive; however, it is nonetheless consistent with the genomic organization and phylogeny of AroB that strongly suggest concurrent LGT of both of these genes. In addition, the OMT of *C. watsonii* and *T. erythraeum* and the dinoflagellates are all united by the presence of a short insertion of similar sequence in the C-terminal half (not shown), again consistent with their common origin. In the cyanobacteria, there is further gene order conservation with a carbamoyl phosphate synthetase

occurring immediately downstream of the OMT (fig. 2). However, no homologue of this gene was found in any of the dinoflagellate sequence data presently available.

The fungal AroB paralogues are also of interest because in AroB phylogeny (fig. 1), these genes are not closely related to canonical fungal AroB sequences (which are part of a large pentafunctional protein in fungi), but instead, form a strongly supported clade (97/79%) with 2 cyanobacterial AroB paralogues (*Anabaena variabilis* ATCC 29413 and *Nostoc punctiforme* PCC 73102). This phylogeny suggests an independent LGT event from different cyanobacteria into fungi, and the evolutionary distance between dinoflagellates and fungi, plus the absence of this AroB paralogue in other eukaryotes, corroborates that 2 separate LGT events likely occurred. It is interesting to note that the cyanobacterial sequences clustering with the fungi are from terrestrial taxa, whereas those that cluster with the dinoflagellates are marine, thus providing the physical association of lifestyles consistent with these incidences of LGT. Compared with the cyanobacteria, there is relatively little gene order conservation surrounding the coding sequence for the AroB paralogue in the fungi—however, we note 2 intriguing linkages. In *Gibberella zeae* and *Aspergillus nidulans*, a homologue of carbamoyl phosphate synthetase, that is downstream of OMT in cyanobacteria, is situated upstream of the AroB paralogue (fig. 2). Further, in *G. zeae* and *Magnaporthe grisea*, a class of OMT that is distinct from the caffeoyl CoA 3-O-methyltransferase was found in dinoflagellates, and these cyanobacteria also occur upstream of the AroB paralogue (fig. 2). This OMT belongs to a class of animal and fungal catechol-O-methyltransferases (designated COMT in fig. 2), that are homologous to other OMTs, but belong to a distant and very divergent clade (data not shown). It appears possible, therefore, that when these fungi acquired an AroB paralogue from cyanobacteria, a carbamoyl phosphate synthetase was also acquired. If the cyanobacterial OMT was also gained (as the intervening gene), this has apparently since been lost in these fungi, perhaps even replaced by the recruitment of a catechol-O-methyltransferase. Interpretation of the evolution of these genes in fungi is confounded by the propensity of fungi to cluster genes of common metabolic pathways (Keller and Hohn 1997), and so gene linkage may simply indicate a common function in this case.

Implications of AroB–OMT Evolution in Dinoflagellates

Reconstructing the history of this LGT event in some detail is aided by the evidence for several unusual aspects of this transfer. First, it appears that not just 1 but 2 adjacent genes were moved concomitantly from a cyanobacterial genome to an ancestor of dinoflagellates. Secondly, we can pinpoint both the donor and recipient with a reasonable degree of certainty because, in the case of the former (cyanobacterium), these 2 genes are only adjacent in a few genomes and the phylogeny is consistent with the genomic structure, and in the case of the latter (dinoflagellate), the fusion protein is absent from *Perkinsus* but intact in *Oxyrrhis*. From the phylogeny of several protein-coding genes and insertion–deletion data in HSP90, *Oxyrrhis* appears to have diverged after *Perkinsus* (Saldarriaga et al. 2003;

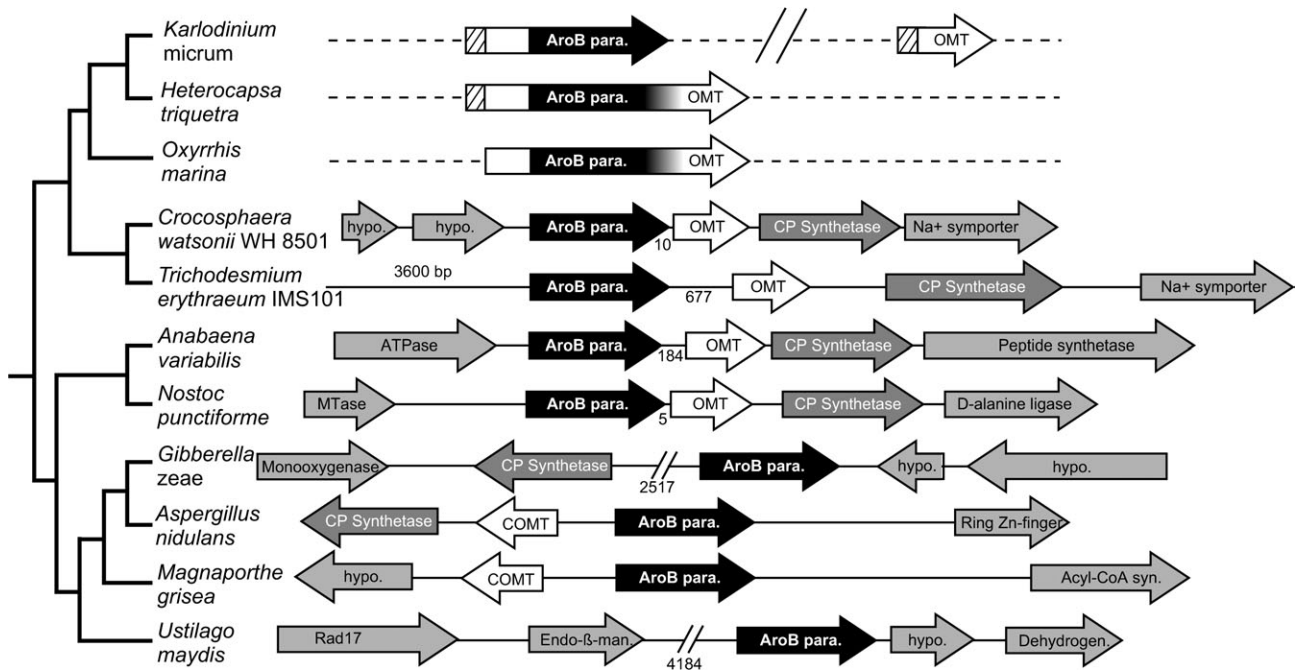


FIG. 2.—Conserved genomic position of the AroB paralogue and caffeoyl CoA 3-O-methyltransferase in cyanobacteria and dinoflagellates. The genomic context of AroB paralogue is shown plotted on the AroB phylogeny as inferred in figure 1. In all 4 cyanobacteria, AroB (black) is followed by OMT (white) in the same strand, which in turn is followed by carbamoyl phosphate synthetase (CP Synthetase, dark gray). For dinoflagellate AroB genes, open boxes represent an additional N-terminal conserved domain (~180 residues) with no detectable similarity to anything, and the diagonally hatched boxes represent domains sharing all the hallmarks of plastid-targeting leaders in the 2 dinoflagellates. *Oxyrrhis* and *Heterocapsa* AroB genes are fused to OMT, whereas in *Karlodinium*, the OMT is found in several copies as stand-alone genes with plastid-targeting leaders. In some of the fungi, a CP Synthetase is linked to AroB paralogue, although upstream on the reverse strand, as is an alternative and distantly related form of OMT (COMT, catechol-O-methyltransferase).

Leander and Keeling 2004) but before the majority of dinoflagellate groups; so we can infer that the transfer and fusion of AroB-OMT took place after the divergence of *Perkinsus* but before that of *Oxyrrhis*. The transfer and fusion must have taken place at nearly the same time because the 2 genes did not exist in the dinoflagellate genome for long enough for random rearrangements to separate them, which presumably happen relatively quickly. They were probably fused by the simple loss of a stop codon in AroB. It is, of course, also possible that the fusion had already occurred in the cyanobacterium donor, but no such condition has been observed in a cyanobacterium to date. At some early stage after the transfer, the protein also acquired an N-terminal extension of unknown function but that has been preserved in all dinoflagellates sampled to date. The absence of a plastid-targeting leader in *Oxyrrhis* and the lack of evidence for a plastid in that lineage further suggests that the fusion protein acquired its plastid-targeting peptide sometime later in the evolution of dinoflagellates, but we cannot rule out that it was immediately targeted to a plastid and this has been reversed in *Oxyrrhis*. Later still, the fusion was reversed in the lineage including *Karlodinium*, and the OMT moiety acquired its own plastid-targeting leader sequence and therefore apparently maintained its function in the plastid. Although we cannot rule out that further copies of AroB and OMT genes may be found in *Karlodinium*, perhaps even in the original fused state, it is nevertheless significant that in *Karlodinium* the fusion has separated in at least one case.

These events have several interesting implications for dinoflagellate plastid biochemistry and the role of transfers and gene fusions in it. The shikimate biosynthetic pathway is often found in the plastid (Herrmann and Weaver 1999), so it is perhaps easy to envision a role for a newly acquired AroB paralogue in the plastid (e.g., in a related side reaction to the shikimate pathway). OMT proteins of this type, however, are not known to target to plastids despite their broad function, so this protein may simply owe its plastid location to its fortuitous attachment to the AroB paralogue. On the other hand, given that AroB is an enzyme involved in the generation of aromatic compounds, and caffeoyl CoA 3-O-methyltransferases specifically methylate aromatic groups, it is plausible that these 2 enzymes are implicated in the same biochemical pathway. The linkage of these genes in cyanobacteria and the fungi (catechol-O-methyltransferase also methylates aromatics) further suggests that these genes are functionally connected. Regardless, the OMT domain clearly found a role in plastid biochemistry that ensures its retention because when AroB and the OMT were subsequently resplit in the *Karlodinium* lineage, the OMT independently acquired a new plastid-targeting leader of its own, suggesting that it was by then essential in the organelle. *Karlodinium* provides an additional fascinating insight into plastid evolution because this dinoflagellate has relatively recently replaced its original plastid with a tertiary endosymbiont derived from a haptophyte (Tengs et al. 2000; Yoon et al. 2005; Patron et al. 2006). Many of its nucleus-encoded plastid genes are derived from the haptophyte

nucleus through internal nucleus–nucleus transfer during endosymbiont establishment (Ishida and Green 2002; Yoon et al. 2005; Patron et al. 2006). However, we have recently described several examples of original dinoflagellate plastid genes whose proteins are now employed in this haptophyte-derived plastid resulting in a chimeric plastid proteome (Patron et al. 2006). AroB and OMT provide additional examples of this process. However, in this case, the haptophyte presumably lacked these novel plastid proteins, and *Karlodinium* has reinstated their plastid roles even during the process of plastid replacement.

These events are also interesting in the light they shed on processes generally considered to be extremely rare. LGT and targeting to organelles are both complex processes, but well-documented cases of both are common enough to not be surprising on their own. Similarly, targeting of nonorganellar proteins to mitochondria or plastids has been shown to happen, but examples generally involve host proteins and not genes from exogenous sources (with a few exceptions: Archibald et al. 2003; Hackett et al. 2004). Gene fusions and fissions, however, are generally thought to be exceedingly rare. Indeed, the root of eukaryotes has recently been inferred based solely on the strength of one such event (Stechmann and Cavalier-Smith 2002). In fact, the frequency of fusion and fission events is poorly known, and we have no evidence to suggest these are either frequent or rare, which must depend a great deal on the nature of the proteins involved. One could imagine that when 2 proteins that are able to function individually are fused, resplitting them may not be functionally deleterious but is still complex because a stop codon must evolve between them and the second protein must evolve expression signals for transcription and translation. This could be simplified by the event taking place in 2 stages involving multiple copies of the gene: in one copy, the upstream or downstream moiety could be eliminated by the origin of a start or stop signal within the gene, allowing the other copy to lose the other moiety at a later time. Other mechanisms are also possible, but in any case, this example serves to illustrate the potential reversibility of fusion events, which should be considered when using such characters to interpret macroevolutionary history.

Supplementary Material

Supplementary Figures 1 and 2 and other Supplementary data are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by a grant from the Canadian Institutes for Health Research (CIHR) to P.J.K. (MOP-42517). EST sequencing was supported by the Protist EST Program funded by Genome Canada/Genome Atlantic. We thank the PEPdb for maintaining EST databases and N. Patron for assistance with *Heterocapsa* and *Karlodinium* EST sequencing and analysis and examination of the *Karlodinium* transit peptides. R.F.W. was supported by a fellowship from the CIHR and Michael Smith Foundation for Health Research (MSFHR). P.J.K. is a Fellow of the

Canadian Institute for Advanced Research and a New Investigator of CIHR and MSFHR.

Literature Cited

- Andersson JO. 2005. Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* 62:1182–97.
- Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc Natl Acad Sci USA* 100:7678–83.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–95.
- Berghthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424:197–201.
- Bruno WJ, Socci ND, Halpern AL. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* 17:189–97.
- Campbell SA, Richards TA, Mui EJ, Samuel BU, Coggins JR, McLeod R, Roberts CW. 2004. A complete shikimate pathway in *Toxoplasma gondii*: an ancient eukaryotic innovation. *Int J Parasitol* 34:5–13.
- Fitzpatrick T, Ricken S, Lanzer M, Amrhein N, Macheroux P, Kappes B. 2001. Subcellular localization and characterization of chorismate synthase in the apicomplexan *Plasmodium falciparum*. *Mol Microbiol* 40:65–75.
- Guindon S, Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
- Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Nosenko T, Bhattacharya D. 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr Biol* 14:213–8.
- Herrmann KM, Weaver LM. 1999. The shikimate pathway. *Annu Rev Plant Physiol Plant Mol Biol* 50:473–503.
- Huelsenbeck JP, Ronquist V. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–5.
- Ishida K, Green BR. 2002. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci USA* 99:9294–9.
- Keeling PJ, Inagaki Y. 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc Natl Acad Sci USA* 101:15380–5.
- Keeling PJ, Palmer JD, Donald RG, Roos DS, Waller RF, McFadden GI. 1999. Shikimate pathway in apicomplexan parasites. *Nature* 397:219–20.
- Keller NP, Hohn TM. 1997. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genet Biol* 21:17–29.
- Leander BS, Keeling PJ. 2004. Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from HSP90 and actin phylogenies. *J Phycol* 40:341–50.
- Loftus B, Anderson I, Davies R et al. (54 co-authors). 2005. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433:865–8.
- Patron NJ, Waller RF, Archibald JM, Keeling PJ. 2005. Complex protein targeting to dinoflagellate plastids. *J Mol Biol* 348:1015–24.
- Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. *J Mol Biol* 357:1373–82.
- Roberts F, Roberts CW, Johnson JJ et al. (12 co-authors). 1998. Evidence for the shikimate pathway in apicomplexan parasites. *Nature* 393:801–5.

- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–5.
- Saldarriaga JF, McEwan ML, Fast NM, Taylor FJR, Keeling PJ. 2003. Multiple protein phylogenies show that *Oxyhris marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int J Syst Evol Microbiol* 53:355–65.
- 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* 13:964–9.
- Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS. 2000. Phylogenetic analyses indicate that the 19'hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* 17:718–29.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–82.
- Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D. 2005. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol Biol Evol* 22:1299–1308.

Martin Embley, Associate Editor

Accepted May 2, 2006