Protist, Vol. ▮, ▮▮▮—▮▮▮, ▮▮ ▮▮▮▮
http://www.elsevier.de/protis
Published online date 14 August 2006

Protist

ORIGINAL PAPER

EFL GTPase in Cryptomonads and the Distribution of EFL and EF-1α in Chromalveolates

Gillian H. Gile, Nicola J. Patron, and Patrick J. Keeling[1]

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, V6 T 1Z4 Canada

Submitted March 16, 2005; Accepted June 25, 2006
Monitoring Editor: Hervé Philippe

EFL (EF-like protein) is a member of the GTPase superfamily that includes several translation factors. Because it has only been found in a few eukaryotic lineages and its presence correlates with the absence of the related core translation factor EF-1α, its distribution is hypothesized to be the result of lateral gene transfer and replacement of EF-1α. In one supergroup of eukaryotes, the chromalveolates, two major lineages were found to contain EFL (dinoflagellates and haptophytes), while the others encode EF-1α (apicomplexans, ciliates, heterokonts and cryptomonads). For each of these groups, this distribution was deduced from whole genome sequence or expressed sequence tag (EST) data from several species, with the exception of cryptomonads from which only a single EF-1α PCR product from one species was known. By sequencing ESTs from two cryptomonads, *Guillardia theta* and *Rhodomonas salina*, and searching for all GTPase translation factors, we revealed that EFL is present in both species, but, contrary to expectations, we found EF-1α in neither. On balance, we suggest the previously reported EF-1α from *Rhodomonas salina* is likely an artefact of contamination. We also identified EFL in EST data from two members of the dinoflagellate lineage, *Karlodinium micrum* and *Oxyrrhis marina*, and from an ongoing genomic sequence project from a third, *Perkinsus marinus. Karlodinium micrum* is a symbiotic pairing of two lineages that would have both had EFL (a dinoflagellate and a haptophyte), but only the dinoflagellate gene remains. *Oxyrrhis marina* and *Perkinsus marinus* are early diverging sister-groups to dinoflagellates, and together show that EFL originated early in this lineage. Phylogenetic analysis confirmed that these genes are all EFL homologues, and showed that cryptomonad genes are not detectably related to EFL from other chromalveolates, which collectively form several distinct groups. The known distribution of EFL now includes a third group of chromalveolates, cryptomonads. Of the six major subgroups of chromalveolates, EFL is found in half and EF-1α in the other half, and none as yet unambiguously possess both genes. Phylogenetic analysis indicates EFL likely arose early within each subgroup where it is found, but suggests it may have originated multiple times within chromalveolates as a whole.

Key words: Chromalveolates; cryptomonads; dinoflagellates; lateral gene transfer; translation factor.

Introduction

Translation elongation factor-1α (EF-1α) plays an integral role in cellular information flow by bringing charged tRNAs to the ribosome during peptide

[1]Corresponding author;
fax +1 604 822 6089
e-mail pkeeling@interchange.ubc.ca (P.J. Keeling)

doi:10.1016/j.protis.2006.06.002

elongation. It is a highly conserved protein found across the three domains of life (called EF-Tu in bacteria). Because of its core role in translation and many interactions with other proteins, it is considered essential and unlikely to be moved from genome to genome by lateral gene transfer, and it has been used in many analyses of phylogeny and molecular evolution (Baldauf and Doolittle 1997; Baldauf and Palmer 1993; Gaucher et al. 2001; Inagaki et al. 2004). However, a recent investigation found that several eukaryotic genomes lack any evidence of an EF-1α gene, and instead encode a distantly related paralogue called EF-like, or EFL (Keeling and Inagaki 2004). EFL is only found in a few lineages scattered across the tree of eukaryotes, and nearly all of these have close relatives that encode EF-1α but not EFL. These relationships suggest that EFL has spread by eukaryote-to-eukaryote lateral gene transfer, functionally replacing EF-1α several times independently despite its crucial role in translation (Keeling and Inagaki 2004).

One lineage where EFL has been found is the hypothetical 'supergroup', chromalveolates. The chromalveolate hypothesis states that the chromists (cryptomonads, haptophytes, and heterokonts) and alveolates (ciliates, dinoflagellates and apicomplexans) share a common ancestor and that this ancestor acquired a secondary red algal plastid (Cavalier-Smith 1999). Although no single gene unites all chromalveolates at once, several host and endosymbiont-derived genes support this hypothesis (Fast et al. 2001; Harper and Keeling 2003; Harper et al. 2005; Patron et al. 2004; Yoon et al. 2002). Within the chromalveolates, two major lineages were found to contain EFL (dinoflagellates and haptophytes), while the other four were found to contain EF-1α (apicomplexans, ciliates, heterokonts and cryptomonads) (Keeling and Inagaki 2004). In the cases of dinoflagellates and haptophytes, this was based on multiple expressed sequence tag (EST) sequencing projects, from which no EF-1α was evident. Similarly, whole genomes and EST projects from apicomplexans, ciliates, and heterokonts bore no evidence of EFL. Only cryptomonads lack data from genome-wide surveys, and here the evidence for EF-1α came from a single gene amplified by PCR (Harper et al. 2005).

We have used EST sequence data to clarify our understanding of EFL's distribution in chromalveolates. Previous sampling from haptophytes represents the entire range of known diversity (because it includes the earliest known lineage, *Pavlova*), but the distribution of EFL in dinoflagellates was less clear because it included no early branching lineages. We accordingly sought EFL and EF-1α in two of the most ancient groups in the dinoflagellate lineage, the parasite *Perkinsus marinus* and the predator *Oxyrrhis marina* (Goggin and Barker 1993; Leander and Keeling 2004; Reece et al. 1997; Saldarriaga et al. 2003). In addition, we have sampled a dinoflagellate that has a haptophyte endosymbiont, *Karlodinium micrum* (Patron et al. 2006; Tengs et al. 2000), to see which EFL was retained from a partnership that involved two EFL-containing organisms. Most importantly, we used EST sequences from two cryptomonads, *Guillardia theta* and *Rhodomonas salina*, to reassess the presence and absence of EFL and EF-1α in this group. In both taxa, EFL was found but EF-1α was not present in our sampling. This refines our understanding of several aspects of the distribution of EFL in chromalveolates: in groups where EFL is found, it appears to be common to all members of that group; of the six major lineages of chromalveolates, half have EFL and half have EF-1α; and within this supergroup the lineages with EFL are not related to one another to the exclusion of those lineages with EF-1α. Phylogenetic analyses suggest that the ancestor of all chromalveolates had EF-1α, but the phylogeny of EFL is not consistent with a common origin of EFL in chromalveolates. At face value this suggests multiple origins of EFL within the supergroup.

Results and Discussion

An Expressed Gene for EFL in Cryptomonads

Members of the translation factor GTPase family were sought from ongoing cryptomonad EST projects using known EFL and EF-1α sequences to search 14,080 *G. theta* sequences comprising 6267 clusters and 2848 *R. salina* sequences comprising 1773 clusters. In both cases sequences corresponding to EFL were found, but EF-1α was not found in our sampling from either species. The *R. salina* EFL was represented by three non-overlapping clusters of ESTs: one with 9 ESTs spanning the 3′ end of the gene, one single EST at the 5′ end, and one single EST in the middle of the gene. A single, truncated EST spanning the 3′ end of the gene represented EFL from *G. theta*. The level of representation seen in *R. salina* is characteristic of EFL from other EST samples (Keeling and Inagaki 2004), but the single EST from 14,080 sequences in *G. theta* is unusual.

Representation does not necessarily relate to expression levels, so we suspect the single sequence is most likely an indication of under-representation in the library. However, the sequence did not contain sites for the restriction enzyme used in library construction (*Not*I), so there is no obvious reason for its under-representation. In neither case did the EST clusters cover the entire gene sequence, so a large fragment of the gene was amplified by RT-PCR, and 14 and 12 individual clones were sequenced from *R. salina* and *G. theta* respectively. In *G. theta,* only 6 synonymous variations were found among all sequences, and the RT-PCR fragments correspond exactly to the EST fragment in the region of overlap. In *R. salina,* two slightly different copies of the gene were found several times each in both RT-PCR and EST sequences (one copy was found in all three EST clusters and the second only in the 3′ cluster). The sequences varied only at synonymous positions, but they shared no 3′ UTR sequence similarity, confirming they are different loci. One full-length *R. salina* EFL had a 13 bp 5′ UTR and 38 bp 3′ UTR, while the second copy lacked sequence for the extreme 5′ end and had a 53 bp 3′ UTR. The *G. theta* sequence is slightly truncated at the 5′ end (approximately the first 8 codons are missing) and there is a 45 bp 3′ UTR. We compared these sequences to an independent collection of *G. theta* ESTs recently released to public databases (Gould et al. 2006) and found two short, identical fragments (see accession AM183813).

Rhodomonas salina is the only cryptomonad previously reported to contain EF-1α (Harper et al. 2005). We specifically searched for this sequence in our databases of *R. salina* and *G. theta* ESTs, and found no evidence for its presence in either collection. The identity of this sequence is therefore open to question. Cryptomonads retain the genome of their red algal endosymbiont (Douglas et al. 2001), so it is possible this gene is derived from that endosymbiont, but since this gene was not demonstrably related to red algal EF-1α (Harper et al. 2005), this seems unlikely. If *R. salina* contains both cytosolic EFL and cytosolic EF-1α, it is of great importance, since the genes are nearly always mutually exclusive (one possible case in the fungus *Basidiobolus* has been reported but not yet confirmed: DQ282610 and DQ275340). However, the acquisition of *R. salina* EF-1α by RT-PCR has not been repeated, and it was part of a large survey that attempted to sequence EF-1α from many cryptomonads and failed. PCR from genomic DNA also failed to recover this gene (Harper et al. 2005). Altogether, we believe this sequence is most likely an artefact of contamination, but if further evidence confirms it does exist in *R. salina*, its origin is of great interest.

EFL from *Karlodinium*, *Oxyrrhis* and *Perkinsus*

EFL has previously been reported from only a few dinoflagellates, namely a full-length mRNA from *Heterocapsa triquetra* and fragments from *Amphidinium carterae* and *Lingulodinium polyedrum* (Bachvaroff et al. 2004; Hackett et al. 2004; Keeling and Inagaki 2004). To determine whether EFL originated with the lineage or predated the dinoflagellate radiation, we identified EFL in two deep-branching lineages that are sister-groups to true dinoflagellates, *O. marina* and *Perkinsus marinus*. Both are non photosynthetic [*O. marina* is a predator and *P. marinus* is a parasite (Azevedo 1989; Droop 1953)] and molecular phylogenetic data show that both diverged early in dinoflagellate evolution, with *P. marinus* branching prior to *O. marina* (Goggin and Barker 1993; Leander and Keeling 2004; Reece et al. 1997; Saldarriaga et al. 2003). The genome-sequencing project of *P. marinus* (http://www.tigr.org/tdb/e2k1/pmg/) and EST data from *O. marina* (40 individual ESTs from a total of 18,012) both contained multiple copies of EFL but no copy of EF-1α. This suggests that EFL originated before the radiation of extant dinoflagellate lineages.

In contrast to *P. marinus* and *O. marina*, *K. micrum* diverged relatively recently within dinoflagellates, but it is of interest because it, and the closely related genus *Karenia*, has lost its original dinoflagellate plastid and replaced it with an endosymbiotic haptophyte (Tengs et al. 2000). It is therefore a symbiotic partnership between two organisms, both of which would have encoded EFL. The nucleus of the haptophyte has since been lost, but the *K. micrum* nuclear genome contains many genes for plastid-targeted proteins derived from this genome (Ishida and Green 2002; Patron et al. 2006; Yoon et al. 2005). Several distinct EFL genes were identified from 47 ESTs from *K. micrum*, but all ESTs were extremely similar to one another and to homologues from other dinoflagellates (supported by phylogeny—see below). We found no evidence of an EFL gene of haptophyte origin in this genome.

Phylogeny of EFL

Phylogenetic trees of all full-length EFL sequences were inferred using a variety of methods, all of

which yielded a topology similar to that shown in Figure 1. This unrooted maximum likelihood (ML) tree has *Bigelowiella natans* as the outgroup because this sequence is always the earliest branch of EFL when analysed with related GTPases (Keeling and Inagaki 2004). The *B. natans* sequence is highly divergent, however, so the root of the EFL tree must be taken with extreme caution, and all analyses were repeated excluding this sequence, which had no major effect on the tree or support levels (not shown). Most nodes are relatively strongly supported by ML and distance bootstrap methods, as well as by Bayesian posterior probabilities (which were all close to 1.0 except the node uniting chytrid fungi and cryptomonads which was 0.725, not shown). Most irrefutably supported lineages are recovered (i.e., green algae, haptophytes, chytrid fungi, dinoflagellates, and cryptomonads), with the exception of the clade uniting dinoflagellates and *Perkinsus* (see below). *Oxyrrhis marina* branches with the dinoflagellates and is basal to *H. triquetra* and *K. micrum*, as expected given its position in phylogenies inferred from other proteins (Leander and Keeling 2004; Saldarriaga et al. 2003). The *K. micrum* EFL branches with dinoflagellates and not haptophytes, confirming its host origin.

Evolution of EFL in Chromalveolates

Two connected features of EFL evolution specifically relating to chromalveolates stand out as unusual. First, why do so many chromalveolates contain EFL rather than EF-1α, and conversely, why are so many of the EFL-containing lineages

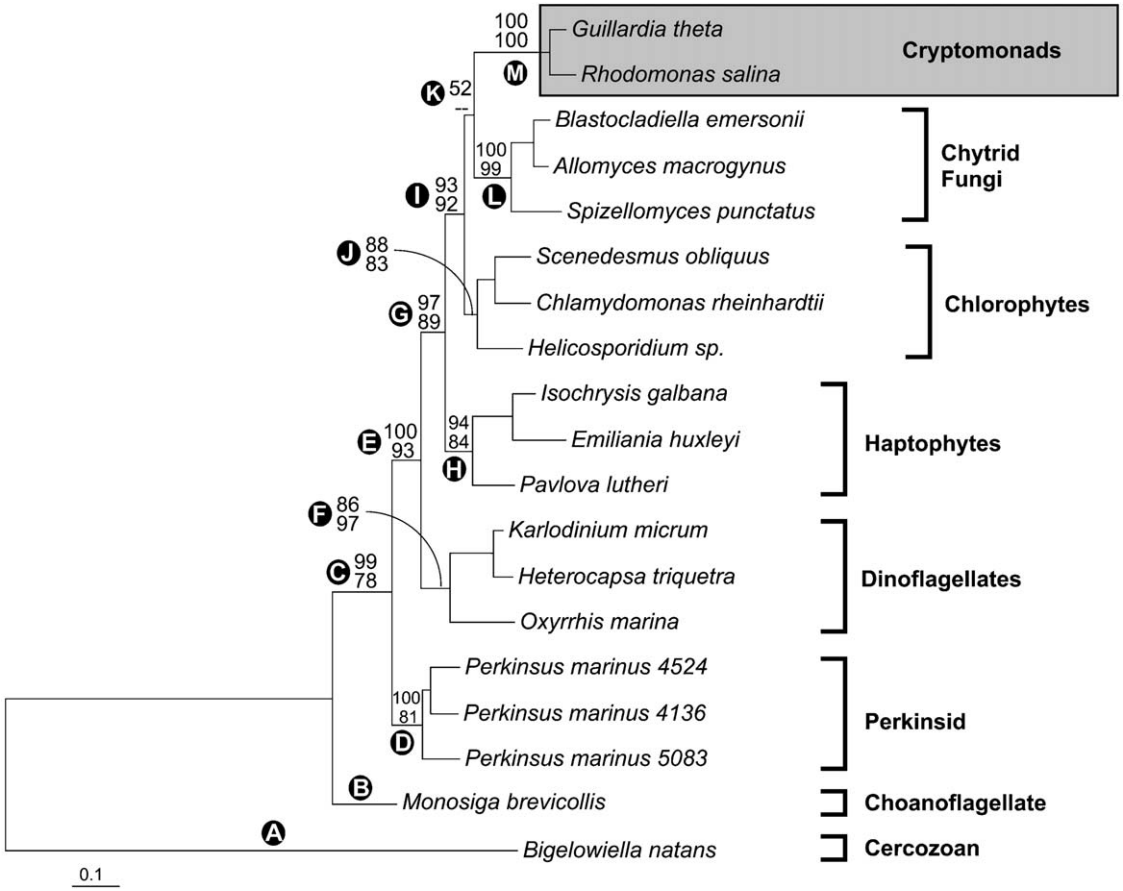

Figure 1. Protein maximum likelihood phylogeny of full-length EFL proteins. Major groups are bracketed and named to the right. New cryptomonad sequences are indicated by a box. Numbers at nodes correspond to bootstrap support from ML (top) and distance (bottom). Letters at nodes correspond to positions where alternative topologies were tested, the results of which are shown in Table 1. All analyses were repeated excluding the divergent *Bigelowiella natans* sequence, but no major differences in either the tree topology or support were observed (data not shown).

chromalveolates? Second, why do the chromalveolate EFL genes not form a single clade?

EFL is very rare in eukaryotes: it has been described in only seven lineages to date, and now nearly half of these are chromalveolates. On the other side of the same coin, half the major lineages of chromalveolates contain EFL, meaning it is more abundant in this supergroup than in any other (Fig. 2). Unlike many other protist groups, there is relatively deep sampling of molecular data from a broad diversity of chromalveolates. Noting that EFL has almost exclusively been found through genome-wide analyses (genome sequences or ESTs), it is possible that the high frequency of EFL in chromalveolates is simply due to the fact that this level of sampling is not widely available in protists. This suggests that improved sampling of protists as a whole may reveal many more lineages with EFL. Alternatively, EFL-containing chromalveolates may simply be more common than other eukaryotes, raising the question, did the EFL-containing chromalveolate lineages acquire EFL several times independently, or was it present in their common ancestor? To distinguish between these possibilities we need to consider the known evolutionary relationships among the chromalveolate lineages as well as the phylogenies of EFL and EF-1α.

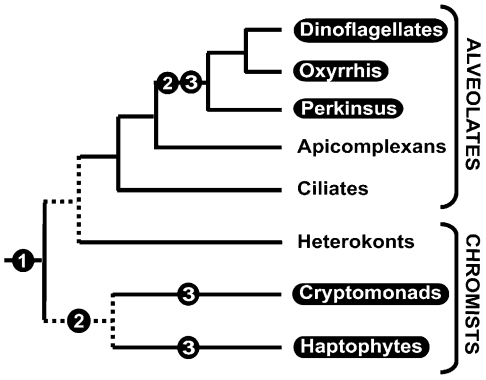

Figure 2. Schematic of relationships between the six chromalveolate groups based on a variety of molecular and morphological data. Groups with names in black text possess EF-1α while groups with names in white text on black backgrounds possess EFL. Highly-supported relationships are shown as solid lines while hypothetical ones are shown as dotted lines. The alveolates (ciliates, dinoflagellates and apicomplexans) are strongly supported by virtually all known molecular and morphological data. There are also molecular data supporting a relationship between alveolates and heterokonts (see text for references), whereas the positions of haptophytes and cryptomonads are not well understood. Numbers indicate three possible scenarios to explain the current distribution of EFL and EF-1α. (1) There was a single origin of EFL with several losses of either EFL or EF-1α. (2) There were two independent origins of EFL if haptophytes and cryptomonads are sister groups. (3) Lastly, three origins of EFL are possible if all known lineages acquired EFL independently.

In chromalveolate phylogeny (Fig. 2), the monophyly of alveolates and branching order between them (ciliates first, then apicomplexans and dinoflagellates) are well established (Fast et al. 2002; Gajadhar et al. 1991; Van de Peer et al. 1996; Wolters 1991). The branching order among chromists, whether they are holophyletic or paraphyletic, and whether all three chromist groups are actually related to alveolates are all less clear, although many genes support a sister relationship between heterokonts and alveolates (Baldauf et al. 2000; Harper et al. 2005; Van de Peer et al. 1996). Regardless of those aspects of the phylogeny we do not yet know, there is no simple explanation for the distribution of EFL in chromalveolates. Dinoflagellates are certainly more closely related to other alveolates with EF-1α (ciliates and apicomplexans), and probably also more closely related to the EF-1α-containing heterokonts, than they are to EFL-containing haptophytes and cryptomonads. To arrive at the present distribution, therefore, EFL must have either been acquired by chromalveolates more than once, or co-existed with EF-1α for a long period of time, with different lineages subsequently losing one gene or the other. Even if EFL replaced the core translation role of EF-1α, EF-1α has several other functions in the cell and it is therefore likely that a complete loss of EF-1α would take more than just the appearance of EFL. Accordingly, we expect that the co-existence of both genes would be essential for some period of time, perhaps indefinitely under certain circumstances. It is therefore conceivable that EF-1α could ‘recapture’ its role in translation, making an early origin with subsequent lineage sorting a viable explanation (Scheme 1 in Fig. 2). On the other hand, EFL appears to have been acquired by several eukaryotic groups independently (Keeling and Inagaki 2004), so it may have originated in all three chromalveolate lineages independently (scheme 3 in Fig. 2). Moreover, if we consider the possibility that cryptomonads and haptophytes are sister groups, then only two independent origins in chromalveolates would be needed to explain the distribution (Scheme 2 in Fig. 2).

If EFL originated once in chromalveolates, however, then chromalveolates should be monophyletic in phylogenies of both EF-1α and EFL. EF-1α phylogeny has been studied extensively [e.g. (Baldauf 1999; Baldauf et al. 2000; Harper et al. 2005; Inagaki et al. 2002, 2004; Saldarriaga et al. 2003)] and it has been shown that ciliate EF-1α genes are subject to a different mode of evolution and are therefore difficult to interpret (Moreira et al. 2002), but the apicomplexan and heterokont homologues are related to one another with modest support in most analyses (Baldauf et al. 2000; Harper et al. 2005; Steenkamp et al. 2006). This suggests that EF-1α was present in the last common ancestor of at least heterokonts and alveolates.

In EFL phylogeny, on the other hand, the chromalveolates do not form a clade. In fact, despite their unusually frequent occurrence, no two EFL-encoding chromalveolate lineages cluster together: *Perkinsus*, dinoflagellates, and haptophytes emerge in a paraphyletic way and cryptomonads branch separately. This is not consistent with a single origin of EFL, and it is not simply due to a poorly resolved tree since most nodes are well supported. We specifically tested three of the more unusual aspects of this tree using Approximately Unbiased (AU) tests. First, the cryptomonads are never observed to branch with any other chromalveolate group, so we tested alternative trees where cryptomonads are moved to all internal nodes (A—K in Fig. 1). With the exception of node J, where cryptomonads are sister to green algae, and I, where cryptomonads are sister to a clade of chytrid fungi plus green algae (the topology found in the distance tree), all these alternatives are rejected at the 5% level, including all positions with other chromalveolates (Table 1). It is noteworthy, however, that the tree placing cryptomonads with haptophytes is only rejected at 0.049, very close to the 5% level. Second, the position of *P. marinus* is unexpected because it is known to be a close relative of dinoflagellates (Goggin and Barker 1993; Leander and Keeling 2004; Reece et al. 1997; Saldarriaga et al. 2003), so we also tested all alternative positions of the three *P. marinus* sequences. In this case, all of the alternatives were rejected, including the expected position as sister to dinoflagellates (Table 1). Lastly, we forced all chromalveolates to be monophyletic. All four of these topologies were rejected regardless of the position of chromalveolates (Table 1). The same topologies were rejected at similar levels when *B. natans* was excluded from the analysis (data not shown).

These results appear to reject the conclusion that chromalveolate EFL genes are monophyletic, but it is important to note that this tree is unrooted: one could interpret it as a clade of chromalveolates with other EFL-encoding groups deriving from within (e.g., fungal and green algal genes coming from a cryptomonad or related source). In addition, we find the rejection of the monophyly of

Table 1. Summary of AU tests comparing alternative phylogenetic positions of cryptomonad, *Perkinsus* and chromalveolate EFL genes.

Position	Cryptomonads	*Perkinsus*	Chromalveolates
A	0.001	0.045	0
B	0.001	0.044	0
C	0	**0.982**	0
D	0.003	NA	NA
E	0.022	NA	NA
F	0.004	0.002	NA
G	0.014	0.018	NA
H	0.049	0.001	NA
I	**0.154**	0.001	NA
J	**0.651**	0	0
K	**0.653**	0	NA
L	NA	0	0
M	NA	0	NA

Position corresponds to the label on Figure 1. At each position, the group being tested (cryptomonads column 1, *Perkinsus* column 2 and all chromalveolates constrained as a group in column 3) was grafted and numbers are *P*-values from AU tests for that topology. NA indicates the position is identical to one of the other positions in that test.

two closely related groups such as *P. marinus* and dinoflagellates highly suspicious, and therefore interpret the tree with caution regardless of the statistical support. The phylogeny may indicate multiple origins of chromalveolate EFL genes, but independent transfers to closely related groups like *P. marinus* and dinoflagellates would require exceptionally strong evidence, and the paraphyletic relationship found here is not very compelling. Continued sampling of EFL diversity may well show that the tree is not as well supported as it appears with the current sampling.

Regardless of how many times the chromalveolates acquired EFL, its distribution in this group raises many interesting questions about its evolution. If it did arise more than once in chromalveolates or if it transferred from chromalveolates to other eukaryotes, this underscores the apparent mobility of this gene. If, on the other hand, it was acquired once in an ancestral chromalveolate, then several lineages must have subsequently lost it (at least ciliates and apicomplexans and probably also heterokonts), raising interesting questions about its functional relationship to EF-1α.

Methods

Identification and characterization of cryptomonad and dinoflagellate EFL genes: Homologues of EFL were identified in expressed sequence tag (EST) projects from two cryptomonads, *Guillardia theta* (CCMP 327) and *Rhodomonas salina* (CCMP 1319), the dinoflagellate *Karlodinium micrum* (CCMP 415) and the non-photosynthetic sister to dinoflagellates, *Oxyrrhis marina* (CCMP 1788). Databases containing these EST sequencing projects (http://www.bch.umontreal.ca/pepdb/pep.html) were searched using tBLASTx for orthologues of both EF-1α and EFL. In some cases multiple copies of the gene were found, but in all such cases they were identical or nearly identical at the amino acid level, so one full-length EST was chosen to represent them and the clone was completely sequenced on both strands. *Perkinsus marinus* sequences were identified using tBLASTx searches from the genome-sequencing project (http://www.tigr.org/tdb/e2k1/pmg/). Three EFL sequences, named 1099751674524, 1099751674136, and 1099751675083 at the time of writing, were conceptually translated and added to the alignment for phylogenetic analyses. For *G. theta* and *R. salina,* EFL was also amplified from total RNA using a degenerate EFL primer designed for the 5′ end of the gene (CTGTCGATCGTCATHTGYGGN-CAYGTNGA) and a species-specific primer (CTTCTTAGCACCACCATCATCGCGAGCAAC for *G. theta* and CGCTTGTGGTGCATCTCCACGGT-GAAGATC for *R. salina*) for the 3′ end, using the Superscript III RT-PCR kit (Invitrogen). Products of the expected size were cloned using TOPO-TA cloning (Invitrogen) and several clones were sequenced on both strands. *K. micrum* ESTs were described in Patron et al. (2006), and *G. theta*, *R. salina*, and *O. marina* EST projects are currently onging. All new EFL sequences were deposited in GenBank under accession numbers DQ659242—DQ659245 and DQ666284.

Phylogenetic analysis: New sequences were added to an existing amino acid alignment of EFL and EF-1α (Keeling and Inagaki 2004) and phylogenetic trees were inferred using maximum likelihood (ML), distance, and Bayesian methods. Trees were inferred using both genes, which confirmed the new sequences were EFL (not shown). All other analyses were restricted to full-length or near full-length EFL sequences alone, from which 438 unambiguously aligned positions from 19 taxa were analysed. ML trees were inferred using PhyML 2.4.4 (Guindon and Gascuel 2003) with input trees generated by BIONJ, the JTT model of amino acids substitution, the proportion of variable rates estimated from the data, and 8 variable categories of substitution rates and invariable sites. One thousand bootstrap trees were inferred with PhyML using the same parameters from the original tree. For distance analyses, gamma corrected distances were calculated by TREE-PUZZLE 5.2 (Strimmer and von Haeseler 1996) using the WAG substitution matrix with 8 variable rate categories and invariable sites. Trees were inferred by weighted neighbour-joining using WEIGHBOR 1.0.1a (Bruno et al. 2000). One thousand bootstrap resampling replicates were performed in batches of 250 using PUZZLEBOOT (shell script by A. Roger and M. Holder, http://www.tree-puzzle.de) with rates and frequencies estimated using TREE-PUZZLE 5.2. MRBAYES 3.0 (Ronquist and Huelsenbeck 2003) was used to perform Bayesian analysis using the JTT substitution model with rates assigned by four equally probable categories approximating a gamma distribution. One cold and three heated chains were run for one million generations, sampling one tree every thousand generations. After 4000 generations, log likelihood values stabilized, and subsequent trees were used to compute the 50% majority-rule consensus tree.

Approximately Unbiased (AU) tests (Shimodaira and Hasegawa 2001) were carried out to examine alternate positions of cryptomonads and *Perkinsus marinus* and the monophyly of chromalveolates. For cryptomonads, ML trees excluding *G. theta* and *R. salina* were optimized as described above, which gave the same topology of the remaining taxa as found in the trees with cryptomonads included. Cryptomonads were added to this optimized tree as sister to all major groups and at all other inter-group nodes, resulting in the ML tree and 10 alternatives. Site-likelihoods for these trees and 100 bootstrap trees were calculated by TREE-PUZZLE 5.1 using the—wsl option with the parameters used for the ML tree, and AU tests were performed using CONSEL 1.19 (Shimodaira 2002). The position of *Perkinsus marinus* and the monophyly of chromalveolates were tested using the same procedure. The entire analysis was repeated with the highly divergent sequence of *Bigelowiella natans* removed from the alignment.

Acknowledgements

This work was supported by a grant (227301) from the Natural Sciences and Engineering Research Council of Canada. *Guillardia theta, R. salina, O. marina* and *K. micrum* EST sequencing was supported by the Protist EST Program of Genome Canada/Genome Atlantic. Sequencing of the *P. marinus* genome is supported by the NSF/USDA-CSREES Microbial Sequencing Program (Award NSF/USDA ♯ 0333240). We thank Uwe Maier for sharing sequences from an independent *G. theta* EST project prior to publication, C. Slamovits for help with *O. marina* EST data, and G. Noble for technical assistance. PJK is a Fellow of the Canadian Institute for Advanced Research and a New Investigator of the Canadian Institutes for Health Research and the Michael Smith Foundation for Health Research.

References

Azevedo C (1989) Fine structure of *Perkinsus atlanticus* n. sp. (Apicomplexa, Perkinsea) parasite of the clam *Ruditapes decussatus* from Portugal. J Parasitol **75**: 627—635

Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF (2004) Dinoflagellate expressed sequence tags data indicate massive transfer of chloroplast genes to the nuclear genome. Protist **155**: 65—78

Baldauf SL (1999) A search for the origins of animals and chytrid fungi: comparing and combining molecular data. Am Nat **154**: S178—S188

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

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

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

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

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

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG (2001) The highly reduced genome of an enslaved algal nucleus. Nature **410**: 1091—1096

Droop MR (1953) Phagotrophy in *Oxyrrhis marina* Dujardin. Nature **172**: 250—251

Fast NM, Kissinger JC, Roos DS, Keeling PJ (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol Biol Evol **18**: 418—426

Fast NM, Xue L, Bingham S, Keeling PJ (2002) Re-examining alveolate evolution using multiple protein molecular phylogenies. J Eukaryot Microbiol **49**: 30—37

Gajadhar AA, Marquardt WC, Hall R, Gunderson J, Ariztia-Carmona EV, Sogin ML (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasitol **45**: 147—154

Gaucher EA, Miyamoto MM, Benner SA (2001) Function-structure analysis of proteins using covarion-based evolutionary approaches: elongation factors. Proc Natl Acad Sci USA **98**: 548—552

Goggin CL, Barker SC (1993) Phylogenetic position of the genus *Perkinsus* (Protista, Apicomplexa) based on small subunit ribosomal RNA. Mol Biochem Parasitol **60**: 65—70

Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier U-G (2006) Protein targeting into the complex plastid of cryptophytes. J Mol Evol **62**: 674—681

Guindon S, Gascuel O (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—218

Harper JT, Keeling PJ (2003) Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol Biol Evol **20**: 1730—1735

Harper JT, Waanders E, Keeling PJ (2005) On the monophyly of the chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol **55**: 487—496

Inagaki Y, Doolittle WF, Baldauf SL, Roger AJ (2002) Lateral transfer of an EF-1alpha gene: origin and evolution of the large subunit of ATP sulfurylase in eubacteria. Curr Biol **12**: 772—776

Inagaki Y, Susko E, Fast NM, Roger AJ (2004) Covarion shifts cause a long-branch attraction artifact that unites microsporidia and archaebacteria in EF-1 alpha phylogenies. Mol Biol Evol **21**: 1340—1349

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—9299

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—15385

Leander BS, Keeling PJ (2004) Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from HSP90 and actin phylogenies. J Phycol **40**: 341—350

Moreira D, Kervestin S, Jean-Jean O, Philippe H (2002) Evolution of eukaryotic translation elongation and termination factors: variations of evolutionary rate and genetic code deviations. Mol Biol Evol **19**: 189—200

Patron NJ, Rogers MB, Keeling PJ (2004) Gene replacement of fructose-1,6-bisphosphate aldolase (FBA) supports a single photosynthetic ancestor of chromalveolates. Eukaryot Cell **3**: 1169—1175

Patron NJ, Waller RF, Keeling PJ (2006) A tertiary plastid uses genes from two endosymbionts. J Mol Biol **357**: 1373—1382

Reece KS, Siddall ME, Burreson EM, Graves JE (1997) Phylogenetic analysis of *Perkinsus* based on actin gene sequences. J Parasitol **83**: 417—423

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**: 1572—1574

Saldarriaga JF, McEwan ML, Fast NM, Taylor FJR, Keeling PJ (2003) Multiple protein phylogenies show that *Oxyhrris marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. Int J Sys Evol Microbiol **53**: 355—365

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol **51**: 492—508

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

Steenkamp ET, Wright J, Baldauf SL (2006) The protistan origins of animals and chytrid fungi. Mol Biol Evol **23**: 93—106

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

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—729

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

Wolters J (1991) The troublesome parasites: molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-ciliate clade. Biosystems **25**: 75—84

Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) A single, ancient origin of the plastid in the Chromista. Proc Natl Acad Sci USA **99**: 15507—15512

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