

Research article

Open Access

## Complex distribution of EFL and EF-1 $\alpha$ proteins in the green algal lineage

Geoffrey P Noble, Matthew B Rogers and Patrick J Keeling\*

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

Email: Geoffrey P Noble - [gnoble@interchange.ubc.ca](mailto:gnoble@interchange.ubc.ca); Matthew B Rogers - [mbrogers@interchange.ubc.ca](mailto:mbrogers@interchange.ubc.ca); Patrick J Keeling\* - [pkeeling@interchange.ubc.ca](mailto:pkeeling@interchange.ubc.ca)

\* Corresponding author

Published: 23 May 2007

Received: 5 February 2007

*BMC Evolutionary Biology* 2007, **7**:82 doi:10.1186/1471-2148-7-82

Accepted: 23 May 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/82>

© 2007 Noble et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** EFL (or elongation factor-like) is a member of the translation superfamily of GTPase proteins. It is restricted to eukaryotes, where it is found in a punctate distribution that is almost mutually exclusive with elongation factor-1 alpha (EF-1 $\alpha$ ). EF-1 $\alpha$  is a core translation factor previously thought to be essential in eukaryotes, so its relationship to EFL has prompted the suggestion that EFL has spread by horizontal or lateral gene transfer (HGT or LGT) and replaced EF-1 $\alpha$  multiple times. Among green algae, trebouxiophyceans and chlorophyceans have EFL, but the ulvophycean *Acetabularia* and the sister group to green algae, land plants, have EF-1 $\alpha$ . This distribution singles out green algae as a particularly promising group to understand the origin of EFL and the effects of its presence on EF-1 $\alpha$ .

**Results:** We have sampled all major lineages of green algae for both EFL and EF-1 $\alpha$ . EFL is unexpectedly broad in its distribution, being found in all green algal lineages (chlorophyceans, trebouxiophyceans, ulvophyceans, prasinophyceans, and mesostigmatophyceans), except charophyceans and the genus *Acetabularia*. The presence of EFL in the genus *Mesostigma* and EF-1 $\alpha$  in *Acetabularia* are of particular interest, since the opposite is true of all their closest relatives. The phylogeny of EFL is poorly resolved, but the *Acetabularia* EF-1 $\alpha$  is clearly related to homologues from land plants and charophyceans, demonstrating that EF-1 $\alpha$  was present in the common ancestor of the green lineage.

**Conclusion:** The distribution of EFL and EF-1 $\alpha$  in the green lineage is not consistent with the phylogeny of the organisms, indicating a complex history of both genes. Overall, we suggest that after the introduction of EFL (in the ancestor of green algae or earlier), both genes co-existed in green algal genomes for some time before one or the other was lost on multiple occasions.

### Background

Horizontal or lateral gene transfer (HGT or LGT) is the non-sexual movement of genetic information between two species. This process has been amply documented to affect prokaryotic genomes: although the frequency and

importance of such events is debated, there is little debate that it plays some role [1-4]. The importance of HGT in the evolution of eukaryotic genomes is not as well appreciated, in part because comparative nuclear genomics has lagged behind that of prokaryotes. Nevertheless, a

number of genome-wide surveys have now shown the process has had a broad impact in several species [5-9], and a number of single gene transfers have also been investigated and found to have interesting implications for ecology, protein function, or genome evolution [10-12].

One potential gene transfer with interesting functional implications is EFL, or elongation factor-like protein [13]. EFL is member of a large GTPase superfamily containing translation initiation, elongation, and termination factors. EFL is specifically related to a subset of proteins found exclusively in eukaryotes, including elongation factor-1 alpha (EF-1 $\alpha$ ), eukaryotic translation release factor-3, and HSP70-binding protein HBS1. EFL is itself only known from eukaryotes, but not all eukaryotes have it. It has presently been reported from eight eukaryotic lineages [13-15], nearly all of which have close relatives which lack it (many of which have completely sequenced genomes supporting the absence of EFL). Moreover, most or all the lineages where EFL has been reported to date appear to lack EF-1 $\alpha$  (again, this includes both organisms with completely sequenced genomes and some with large scale EST sequencing surveys). Altogether there is generally a mutually exclusive distribution of EFL and EF-1 $\alpha$ , and both are scattered around the tree of eukaryotes. This is of interest for several reasons. First, EF-1 $\alpha$  was formerly considered essential because of the important and universal role it plays in translation, but this is clearly not the case since some EFL containing organisms clearly don't use EF-1 $\alpha$ . Second, the distribution suggests EFL might substitute or partially substitute for EF-1 $\alpha$  function. Third, the punctate distribution of EFL suggests it may be moving between eukaryotes [13]. As more lineages with EFL are found [14,15], the alternative that EFL is ancestral to all the lineages that possess it, and that either EFL or EF-1 $\alpha$  has been lost many times in different lineages becomes more defensible. However, the number of such events is very large since the lineages possessing EFL are scattered across the tree of eukaryotes, and the number of lineages with EFL remains far fewer than those with EF-1 $\alpha$ . Based on the current distribution we favour the movement of EFL between eukaryotes, but this must be reassessed as the distribution becomes better known.

Within each group where EFL has been found, the distribution of the protein is sometimes simple but uninformative; for example, all dinoflagellates appear to use EFL [14]. Conversely, its distribution among other groups is more complex but difficult to interpret; for example, some fungi, ichthyosporeans, and choanoflagellates have EFL while others do not and the relationships between these is unclear [13,15-17]. A group where the distribution appeared to be most precisely explained was the green algae: here the well-sampled plants lack EFL and

instead have EF-1 $\alpha$ , whereas the well-sampled chlorophytes and trebouxiophytes have EFL and lack EF-1 $\alpha$ . The single ulvophycean eukaryotic elongation factor, from *Acetabularia acetabulum*, was found to use EF-1 $\alpha$  but not EFL, altogether leading to the conclusion that the green algal EFL originated in a common ancestor of chlorophytes and trebouxiophytes {Keeling, 2004 #652}. This makes green algae a good model for the study of the fine scale distribution of EFL for two reasons. First, EFL seems to have arisen within the group rather than prior to its diversification, so the age of this event could in principle be predicted. Second, while the phylogeny of green algae is not perfectly known, it is better understood than most other eukaryotic groups [18], so the distribution can be interpreted according to a reasonable phylogeny. A better idea of this distribution can help us answer several questions. In particular, if EFL is introduced into a lineage by HGT, what happens to EF-1 $\alpha$ ? Do they both persist for some time, and does lineage sorting take place? Only a fine-scale survey of a group where EFL exists in some but not all members could address such questions, and these questions are important because they are one way to determine the possible functional relationship between EFL and EF-1 $\alpha$ .

Here we have sampled EFL and EF-1 $\alpha$  from representatives of all major missing lineages of green algae. The viridiplantae are divided into two main groups, streptophytes and chlorophytes [18]. The streptophytes include land plants, charophytes, and the enigmatic scaly green flagellate *Mesostigma*. The chlorophytes include four subdivisions, prasinophytes, ulvophytes, trebouxiophytes and chlorophytes (we will use the suffix '-ceans' so as to distinguish the larger group chlorophytes from the subgroup chlorophytes). The expectation, based on the relatively simple distribution originally described, is that EF-1 $\alpha$  but not EFL should be found in all green lineages except chlorophytes and trebouxiophytes. In contrast, however, we have found that EFL is abundant in chlorophytes and EF-1 $\alpha$  comparatively rare. Indeed, among chlorophyte green algae, *Acetabularia* is the only lineage in which we could find EF-1 $\alpha$ : EFL was found in all other chlorophyte lineages, and in the streptophyte *Mesostigma*. Phylogenetic reconstruction shows the *Acetabularia* EF-1 $\alpha$  is related to homologues from charophytes and land plants, and therefore ancestral (and the only known chlorophyte EF-1 $\alpha$  to be retained). This implies that EFL and EF-1 $\alpha$  were likely both present in the ancestor of streptophytes and chlorophytes, and that lineage sorting led to the current distribution of EFL and EF-1 $\alpha$ . This has implications for the functional co-existence of the two proteins, and also suggests that EFL was more frequently retained than was EF-1 $\alpha$ .

## Results & discussion

### Characterisation of new EFL and EF-1 $\alpha$ genes from green algae

Previously, EFL was found in a number of species of chlorophytes and trebouxiophytes, but not in any land plants or the ulvophycean *A. acetabulum*. This led to the conclusion that it originated in a common ancestor of chlorophytes and trebouxiophytes [13], which in turn leads to the prediction that all other green algae should only contain EF-1 $\alpha$ . In contrast, however, we have found that most green algal lineages possess EFL. Genes encoding EFL were characterized from three prasinophytes (*M. pusilla*, *T. tetrathele*, and *O. tauri*), two ulvophytes (*U. fenestrata* and *U. intestinalis*), an additional chlorophyte (*Chlorococcum* sp.), and the only mesostigmatophyte (*M. viride*). From none of these lineages were we able to detect the presence of a gene for EF-1 $\alpha$ . Moreover, EF-1 $\alpha$  was not present in the complete genome of *O. tauri* or in EST surveys of *M. viride* [19,20]. An EST survey of a third ulvophyte, *U. linza* [21], was also searched and in agreement with our results contained a short fragment of EFL but no detectable EF-1 $\alpha$ . Conversely, EF-1 $\alpha$  was characterized from two charophytes (*C. australis* and *Spirogyra* sp.), and we were not able to detect the presence of EFL in either of these species. EFL fragments were identified in EST or genomic survey projects from four land plants, *Pinus taeda*, *Oryza sativa*, *Lactuca saligna*, and *Triticum aestivum*. With the exception of the first two, however, these do not form a group in phylogenetic analyses (Suppl. Figure 1) as one would expect if they were ancestrally land plant sequences. Moreover, most are closely related to either a subset of green algae or fungi, and are poorly represented in the surveys where they are found. In addition, in cases where substantial genomic data are available from the same organism, the EFL is only found in EST sequences and not in the genomic data. The best cases are *O. sativa* and *T. aestivum*, where Blast searches against their respective genome databases revealed no EFL sequence, and specifically no match to the EST purportedly from that organism. On balance, we conclude that these are contaminant sequences until evidence that they are truly encoded in the plant genomes is available (which, according to the phylogeny, would suggest several recent acquisitions in certain plants).

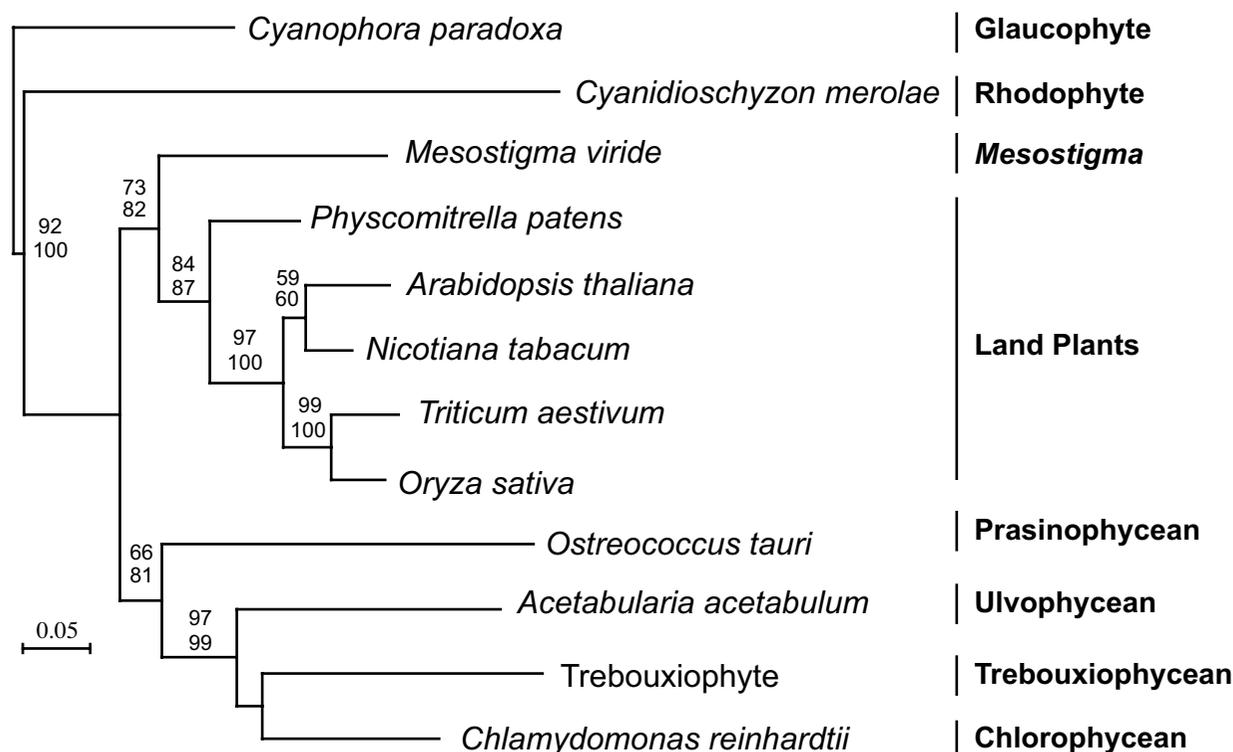
In general, organisms from the streptophyte lineage mostly possess EF-1 $\alpha$  while organisms from the chlorophyte lineage mostly possess EFL. However, *M. viride* and *A. acetabulum* are important exceptions because, based on their currently accepted taxonomic positions, one would predict that they should use EF-1 $\alpha$  and EFL respectively, whereas EST sampling from both reveals the opposite. It is possible that both genes are present in these genomes, although this would still suggest the expected gene is not expressed as highly as the one characterized here, espe-

cially in *M. viride*, where there is a large sample of 10,395 ESTs [19]. It is also possible that these species are misclassified; that *M. viride* is really a chlorophyte and *A. acetabulum* is really a streptophyte. The phylogenetic position of both genera, and in particular *M. viride*, have been analysed many times using a variety of data [18,19,22-27], making this somewhat unlikely. Nevertheless, we carried out a phylogeny based on nine concatenated protein-coding genes from the EST surveys to re-confirm their phylogenetic affinities to streptophytes or chlorophytes. The phylogeny (Figure 1) is consistent with previous analyses that place *M. viride* at the base of the streptophytes and *A. acetabulum* within the chlorophytes. This rules out the possibility that the possession of EF-1 $\alpha$  and EFL by *A. acetabulum* and *M. viride*, respectively, reflects their relationship to other groups that possess the same gene. In short, the distribution of EFL and EF-1 $\alpha$  is not strictly consistent with the phylogenetic relationships of the organisms where the proteins are found.

### Phylogenies of EFL and EF-1 $\alpha$

The complex distribution of EFL and EF-1 $\alpha$  could have arisen from multiple transfers of EFL to green algae, reversions by re-introduction of EF-1 $\alpha$ , or an extended period of time where both genes co-existed in the same ancestral lineage followed by differential loss in the descendants of this lineage. Each of these hypotheses is significant for different reasons, and each leads to predictions for the phylogeny of EFL and EF-1 $\alpha$ . If EFL invaded green algae multiple times, the green algal EFLs will not form a single well-supported clade. If EF-1 $\alpha$  was re-introduced to one or more groups, then the green algal and plant EF-1 $\alpha$  genes will not form a single well-supported clade. Lastly, if lineage sorting has taken place, then the green lineage should form a single clade in both genes.

The first possibility cannot be confidently accepted or rejected because the phylogeny of green algal EFL genes is not sufficiently well resolved. Figure 2 shows a phylogeny of all full-length EFL genes with the exception of the highly divergent gene from the chlorarachniophyte *Bigeloviella natans*; this gene has been included in analyses previously [13], and we similarly found no indication that it was related to green algal EFL (Suppl. Figure 1). The green algal EFL sequences fall into three groups. One large and weakly supported clade includes representatives of all four major chlorophyte subgroups (noteworthy among which are the ulvophytes since *Acetabularia* is currently believed to be related to this group). A second strongly-supported clade includes marine picoplanktonic prasinophytes (noteworthy among which is a previously unidentified environmental sequence from the Sargasso Sea that is specifically related to *M. pusilla*, consistent with the source of this material and the size of cells collected: Suppl. Figure 1). Lastly, *M. viride* branches independently



**Figure 1**

**Bayesian phylogeny of concatenated proteins from green algae and plants with maximum likelihood branch lengths.** Numbers at nodes correspond to bootstrap support from protein maximum likelihood methods ProML (top) and PhyML (bottom). Major groups are labeled to the right. Note that the relationships between *M. viride* and streptophytes and *A. acetabulum* and chlorophytes are both recovered and supported by bootstrap. The 'Trebouxiphyte' is a composite of sequences from *Chlorella*, *Prototheca*, and *Helicosporidium* (for which genes come from each taxon, see methods).

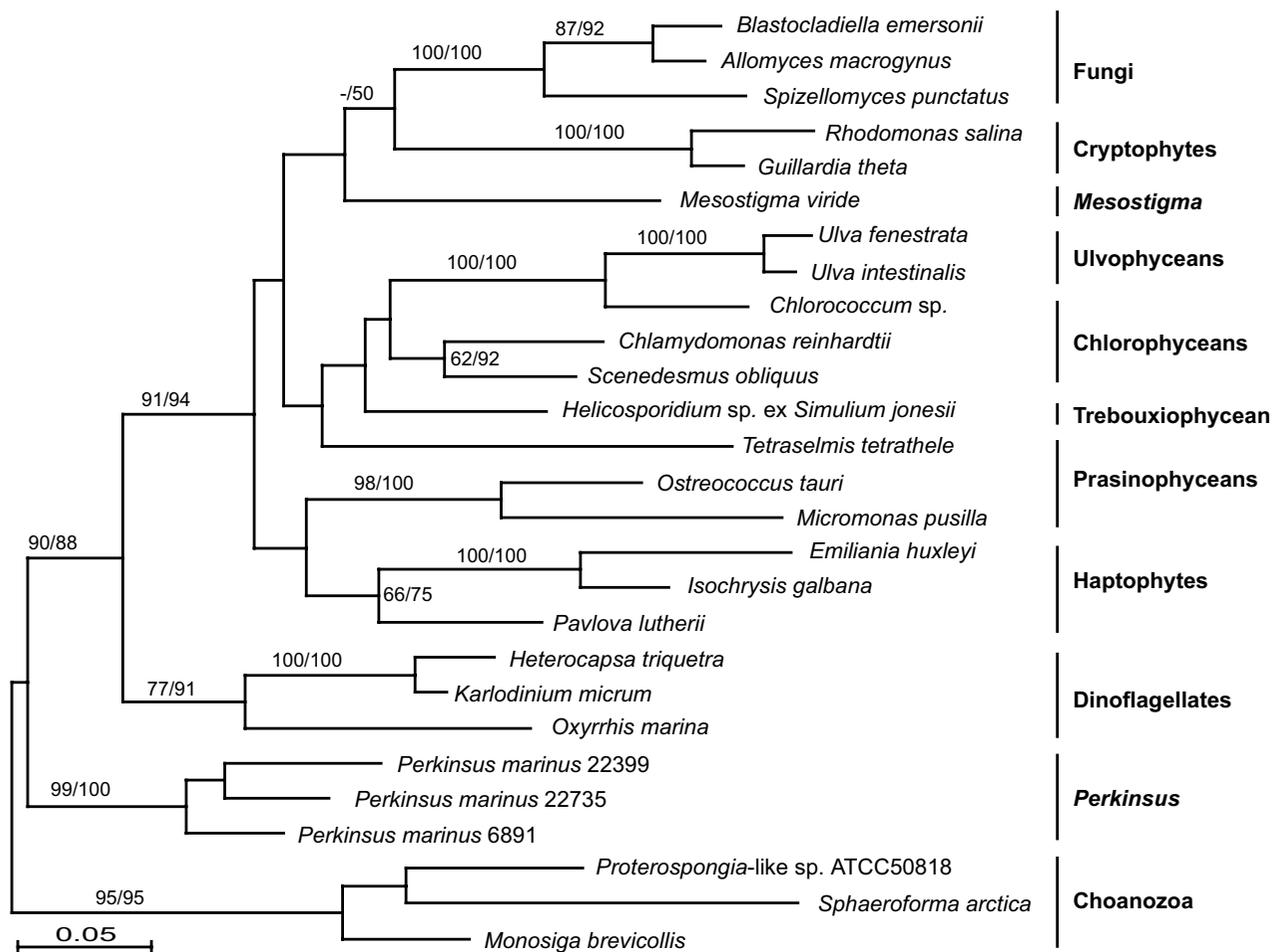
of other green algae without any support for its position, which is unfortunate since it would be expected to possess EF-1 $\alpha$  rather than EFL. None of these three groups of green algae is related to any other group or to one another with strong support in any analysis.

The phylogeny of EF-1 $\alpha$  is much more informative (Figure 3). In contrast to EFL, the positions of all members of the green lineage, *A. acetabulum*, charophyceans, and land plants, are resolved with strong support, and the relationships of the EF-1 $\alpha$  genes match those expected for the organisms. The green clade is well supported overall (91–98%), with *A. acetabulum* sister to all streptophytes with 94–100% support. The charophyceans are paraphyletic with *Spirogyra* sister to land plants, although this branching order is not supported by bootstrap analyses. The position of *A. acetabulum* is of particular importance, because the presence of EF-1 $\alpha$  is unexpected in this ulvophycean, given chlorophytes as a whole, including other

ulvophyceans, do not appear to possess EF-1 $\alpha$ . In previous analyses based on a truncated EST the position of the *A. acetabulum* EF-1 $\alpha$  was not resolved with strong support, but it did not branch with plants [13]. However, having now sequenced an additional 561 bp of the 5' end (about 50% of the alignable sequence), it is clear that the *A. acetabulum* EF-1 $\alpha$  is indeed sister to other members of the green lineage. This suggests very strongly that *A. acetabulum* has not 'reacquired' EF-1 $\alpha$  secondarily (e.g., through HGT), and by extension that its EF-1 $\alpha$  gene is the single known representative of the ancestral chlorophyte EF-1 $\alpha$ .

#### **The evolution and distribution of EFL in the green algal lineage**

The range of green algae that possess EFL is much broader than expected based on previous data, and more importantly its distribution is not easily reconciled with the evolutionary relationships among the organisms that possess it. On the whole, streptophytes possess EF-1 $\alpha$ , with the



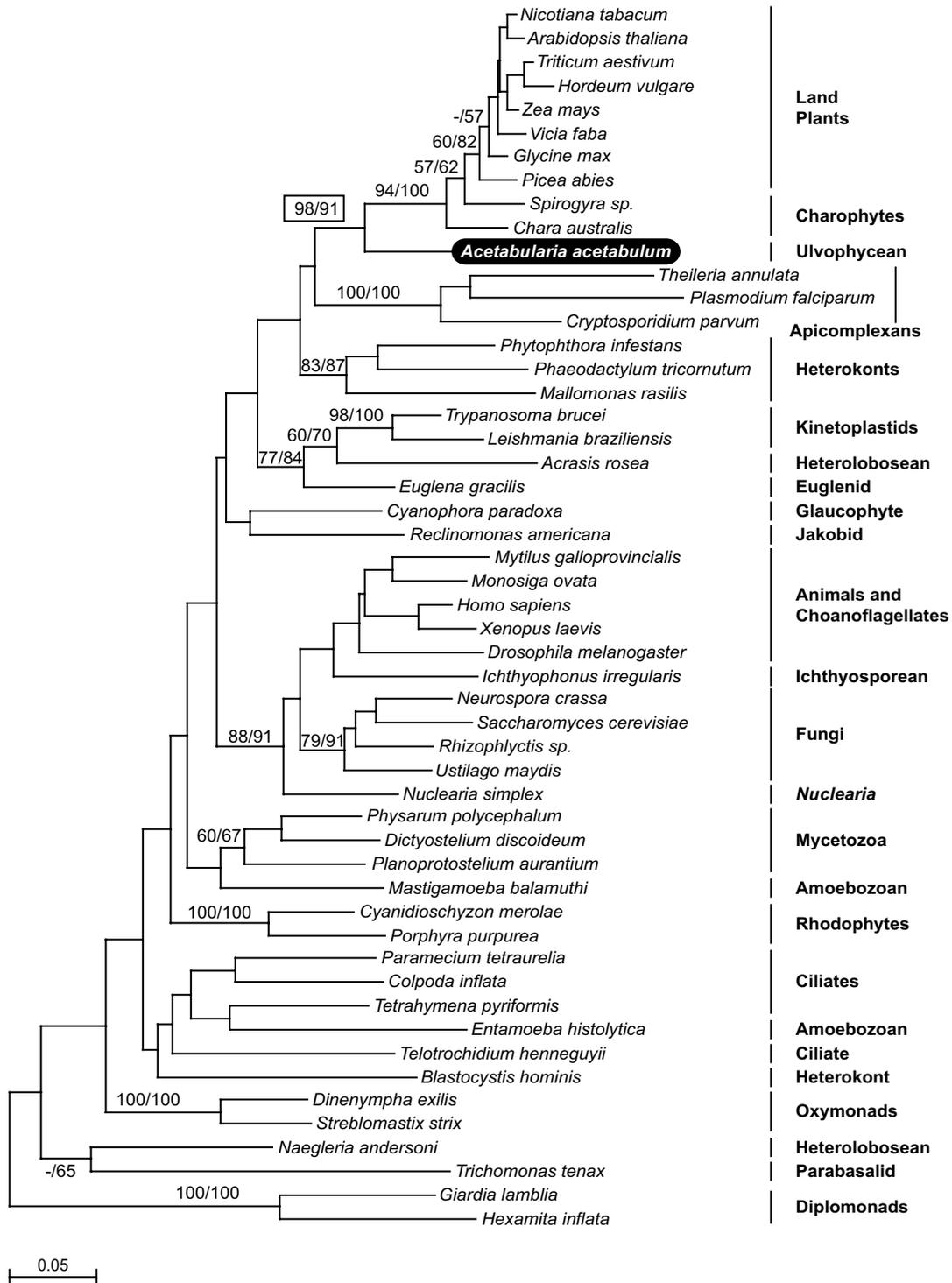
**Figure 2**  
**Bayesian phylogeny of EFL proteins with maximum likelihood branch lengths.** Numbers at nodes correspond to bootstrap support from protein maximum likelihood methods ProML (left) and PhyML (right). Major groups are labeled to the right. Note that the relationships of many of the green algal groups are not well supported, and that they fall into three poorly separated groups.

exception of *M. viride*, whereas chlorophytes possess EFL, with the exception of *A. acetabulum*. Figure 4 shows a schematic phylogeny of the green lineage with the distribution of EFL and EF-1 $\alpha$ . The phylogeny of EF-1 $\alpha$  demonstrates that EF-1 $\alpha$  existed in the ancestor of the green lineage, and that the land plants and charophytes as well as the chlorophyte *A. acetabulum* have all retained this ancestral gene.

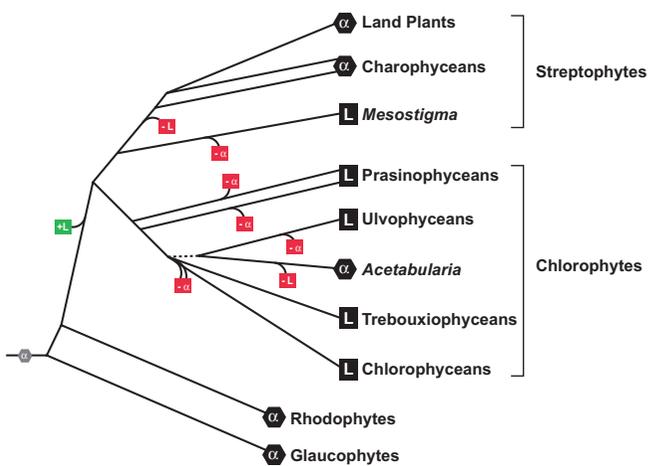
It is possible that the distribution of EFL is due to multiple independent invasions of the gene to different subgroups

of the green lineage. This would require three or perhaps more origins of EFL (this is unclear because the phylogeny of basal chlorophytes is not resolved), which would suggest that EFL was very mobile indeed. While this is consistent with the present data (because the EFL tree does not reject such a scenario), it seems unlikely that one group would acquire such a gene many times.

Another possibility is that EFL and EF-1 $\alpha$  were both present in the common ancestor of streptophytes and chlorophytes and co-existed for some extended period of



**Figure 3**  
**Bayesian phylogeny of EF-1α proteins with maximum likelihood branch lengths.** Numbers at nodes correspond to bootstrap support from protein maximum likelihood methods ProML (left) and PhyML (right). Major groups are labeled to the right. Note that *A. acetabulum* is sister to the streptophytes and that the charophyceans are sister to plants, in agreement with previous molecular phylogenies of the viridiplantae.



**Figure 4**  
**Schematic of evolutionary relationships within the green algal lineage showing distribution of EFL and EF-1 $\alpha$  and possible events explaining the distribution.**

A phylogeny of the Plantae, with the known presence of either EFL or EF-1 $\alpha$  in extant lineages plotted at the tips of each branch (EFL is a square containing "L" and EF-1 $\alpha$  is an octagon containing an " $\alpha$ "). All known rhodophytes and glaucophytes contain EF-1 $\alpha$ , as do most other eukaryotic lineages, so the ancestor is inferred to have contained EF-1 $\alpha$  as well (indicated by a gray octagon at the base of the tree). Also plotted is one possible explanation for the current distribution of the two genes in the green lineage, where EFL was gained once in the ancestor of the green viridiplantae (indicated by the green box), and subsequently either EFL or EF-1 $\alpha$  lost in several lineages each (indicated by red boxes with indicated which gene would have been lost). Other models to explain this distribution that include multiple origins of EFL are also possible, but are not shown for simplicity. The tree of the green lineage is based on Figure 1 and other analyses [18].

time. Assuming that both genes do not still co-exist in any one green algal genome, this would mean EFL was lost once in the ancestor of charophyceans and land plants, and once in an ancestor of *A. acetabulum*, whereas EF-1 $\alpha$  was lost once in an ancestor of *M. viride* and several times in chlorophyte lineages. Because the phylogeny of chlorophytes is still uncertain it is impossible to state how many times EF-1 $\alpha$  would have to have been lost, but it is interesting that EF-1 $\alpha$  loss would likely have been more frequent than EFL loss. It is also possible that both genes persist in some green algae, although we know this is not the case in several plants, *O. tauri*, and *C. reinhardtii*, and have good reason to suspect it is not true in the many green algae for which there are large EST surveys. A long period of co-existence followed by differential loss of one or the other gene may seem unlikely at first glance, but if one considers the function of these proteins it may not be

so. More specifically, EFL has been argued to share some functional overlap with EF-1 $\alpha$  in translation [13]. However, EF-1 $\alpha$  is a multifunctional protein, so even if EFL is capable of taking over its role in translation elongation, it stands to reason that EF-1 $\alpha$  would have to persist until its other roles in the cell were also obsolete or at least non-essential. Accordingly, a period of co-existence is probably not only possible, but necessary, and during such a time either protein could conceivably be lost. This has the interesting implication that plants, although they now only have EF-1 $\alpha$ , once had EFL as well. It also bolsters expectations that organisms with both proteins should exist. The zygomycete fungus *Spizellomyces punctatus* is reported to have both genes [17], and we have detected both in the diatom *Thalassiosira pseudonana* (G. Gile and PJK, unpublished data), but so far co-existence of the two proteins is curiously rare.

## Conclusion

We have shown that the apparently mutually exclusive distribution of the GTPase genes EFL and EF-1 $\alpha$  in the green lineage is not consistent with the phylogeny of the organisms in which the genes are found. Specifically, members of the streptophyte lineage tend to encode EF-1 $\alpha$  and not EFL, except for the basal genus *Mesostigma* where the opposite is found. Conversely, members of the chlorophytes tend to encode EFL and not EF-1 $\alpha$ , except for *Acetabularia* where the opposite is found. The phylogeny of EFL does not resolve the origin of the *Mesostigma* gene, but the phylogeny of EF-1 $\alpha$  clearly shows the *Acetabularia* gene to be a relict, ancestral chlorophyte EF-1 $\alpha$  rather than an independent re-acquisition. Altogether, we suggest that the ancestor of the green lineage encoded EF-1 $\alpha$ , but at some point EFL was introduced and the two genes co-existed in green algal genomes for some time before one or the other was lost on multiple occasions.

## Methods

### Strains

Multiple whole plants of *Ulva fenestrata* and *Ulva intestinalis* were collected from the intertidal zone at Spanish Banks, Vancouver, BC Canada. *Micromonas pusilla* (strain NEPCC 29), *Chlorococcum* sp. (strain NEPCC 478), *Tetraselmis tetrathele* (strain NEPCC 500) were grown in natural seawater medium [28] at 16° with a 16:8 light:dark cycle. *Chara australis* (strain FWAC 7144) and *Spirogyra* sp. (strain FWAC 125) were grown in soilwater medium [29] at 16° with a 16:8 light:dark cycle. A cDNA library from *Acetabularia acetabulum* (strain Aa0005) was generously provided by D. Mandoli.

### Characterisation of EFL and EF-1 $\alpha$ genes

Material from all cultures was harvested by centrifugation and total RNA and DNA was isolated using protocols described previously [30]. RNA and DNA was used as a tem-

plate for RT-PCR and PCR reactions, respectively, using primers for each of EFL and EF-1 $\alpha$ . EFL primers were CTGTCGATCGTCATHGTGCGNCAAYGTNGA and CTTGATRTTNAGNCCNACRTRTRCNCC, EF-1 $\alpha$  primers were AACATCGTCGTGATHGGNCAAYGTNGA and CTTGATCACNCCNACNGCNACNGT or CAACATCGTCGTATCGGNCAAYGTNGA and GCCGCGCACGTTGAANCCNACRTRTC. Products of the expected size (or larger in the case of genomic DNA amplifications) were cloned and multiple clones completely sequenced. In all cases the RT-PCR approach was more successful, leading us to conclude EFL and EF-1 $\alpha$  may encode a number of introns. The *A. acetabulum* EFL was partially characterized as part of an EST survey [13], and to improve the resolution of its phylogenetic position we characterized the 5' portion of the gene by amplification from a cDNA library using the specific primer TTCCGACCGGCACGGTCCAATCCG and the 5' degenerate EF-1 $\alpha$  primer AACATCGTCGTGATHGGNCAAYGTNGA. During the course of this work the complete nuclear genome of the prasinophycean *Ostreococcus tauri* and an EST survey from the mesostigmatophycean *Mesostigma viride* (strain NIES 476) both became available [19,20]. EFL and EF-1 $\alpha$  were both sought from this sequence data, from which only EFL sequences could be found. New sequences were deposited in GenBank under accession numbers [EF551321-EF551331](#).

#### Phylogenetic analyses

New EFL and EF-1 $\alpha$  sequences were added to existing amino acid alignments [14]. The two proteins were analysed separately since their relationship to one another has been examined previously, and separate analyses allow the inclusion of more unambiguously alignable characters, 425 and 407 for EFL and EF-1 $\alpha$ , respectively. An alignment of nine concatenated proteins from algae and plants was also constructed using actin, alpha-tubulin, beta-tubulin, RbcS, Rps10, Rps13, Rpl3, Rpl11, and Rpl13, for a total of 1,206 unambiguously alignable amino acid characters (with no missing data). The trebouxiophytes are represented by a composite of sequences from three species: actin, Rps10, and Rpl11 are from *Helicosporidium* sp., beta-tubulin, Rpl3, Rpl13 and Rps13 are from *Prototheca wickerhamii*, and alpha-tubulin, RbcS, and TufA are from *Chlorella vulgaris*. Trees were inferred using distance, maximum likelihood and Bayesian methods. Bayesian trees were inferred using Mr. Bayes 3.1 [31] employing the WAG substitution model with site-to-site rate variation modeled on a gamma distribution with 8 variable rate categories and one category of invariable sites, three heated chains and one cold one, and 1,000,000 generations with sampling every 1,000 generations. Log likelihoods were plotted and showed a rapid plateau after only five samples, so a burnin of 40 trees was removed before constructing the consensus (constructing a consensus of all trees resulted in the same topology). Maximum likelihood branch lengths for the consensus

topology were calculated using ProML 3.6 [32] with JTT, 8 gamma rate categories and one category of invariable sites. ProML trees were inferred using the same settings, and 100 bootstraps were inferred with the gamma shape parameter alpha and proportion of invariant sites estimated using Tree-Puzzle 5.2 [33]. ML trees and 1,000 bootstrap trees were also inferred using PhyML 2.4.4 [34] with the WAG model with 8 gamma rate categories and one category of invariable sites (when p-inv was not zero), and parameters estimated from the data. Distances were also calculated using Tree-Puzzle with the same settings (and parameters estimated from the data), and trees constructed using WEIGHBOUR 1.0.1a [35]. 100 bootstrap replicates were carried out using puzzleboot [36], which gave similar results as maximum likelihood (not shown).

#### Authors' contributions

G.P. Noble and P.J. Keeling characterized all new gene sequences. Phylogenetic analyses were performed by M.B. Rogers, G.P. Noble and P.J. Keeling. P.J. Keeling conceived of the study and wrote the manuscript.

#### Additional material

##### Additional file 1

Bayesian phylogeny of all known EFL genes and fragments with maximum likelihood branch lengths. Numbers at nodes correspond to bootstrap support from protein maximum likelihood methods ProML (top) and PhyML (bottom). Major groups are labeled to the right.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-82-S1.eps>]

#### Acknowledgements

This work was supported by a grant (227301) from the Natural Sciences and Engineering Research Council of Canada (NSERC). We thank D. Mandoli for providing the *A. acetabulum* cDNA library and C. Bowler and I. Grigoriev for unpublished data from *P. tricornutum*. GPN was supported by a student award from NSERC and PJK is a Fellow of the Canadian Institutes for Advanced Research and a Senior Scholar of the Michael Smith Foundation for Health Research.

#### References

- Lawrence JG, Hendrickson H: **Lateral gene transfer: when will adolescence end?** *Mol Microbiol* 2003, **50**:739-749.
- Doolittle WF: **Lateral genomics.** *Trends Cell Biol* 1999, **9**:M5-8.
- Gogarten JP, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19**:2226-2238.
- Kurland CG, Canback B, Berg OG: **Horizontal gene transfer: a critical view.** *Proc Natl Acad Sci USA* 2003, **100**:9658-9662.
- Andersson JO: **Lateral gene transfer in eukaryotes.** *Cell Mol Life Sci* 2005, **62**:1182-1197.
- Bergthorsson U, Adams KL, Thomason B, Palmer JD: **Widespread horizontal transfer of mitochondrial genes in flowering plants.** *Nature* 2003, **424**:197-201.
- Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ: **Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*.** *Proc Natl Acad Sci USA* 2003, **100**:7678-7683.

8. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, Nozaki T, Suh B, Pop M, Duchene M, Ackers J, Tannich E, Leippe M, Hofer M, Bruchhaus I, Willhoeft U, Bhattacharya A, Chillingworth T, Churcher C, Hance Z, Harris B, Harris D, Jagels K, Moule S, Mungall K, Ormond D, Squares R, Whitehead S, Quail MA, Rabbinowitsch E, Norbertczak H, Price C, Wang Z, Guillen N, Gilchrist C, Stroup SE, Bhattacharya S, Lohia A, Foster PG, Sicheritz-Ponten T, Weber C, Singh U, Mukherjee C, El-Sayed NM, Petri WA Jr., Clark CG, Embley TM, Barrell B, Fraser CM, Hall N: **The genome of the protist parasite *Entamoeba histolytica***. *Nature* 2005, **433**:865-868.
9. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM: **The genome of the African trypanosome *Trypanosoma brucei***. *Science* 2005, **309**:416-422.
10. Jenkins C, Samudrala R, Anderson I, Hedlund BP, Petroni G, Michailova N, Pinel N, Overbeek R, Rosati G, Staley JT: **Genes for the cytoskeletal protein tubulin in the bacterial genus *Prostheco bacter***. *Proc Natl Acad Sci USA* 2002, **99**:17049-17054.
11. Huang J, Mullanpudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC: **Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum***. *Genome Biol* 2004, **5**:R88.
12. Waller RF, Slamovits CH, Keeling PJ: **Lateral gene transfer of a multigene region from cyanobacteria to dinoflagellates resulting in a novel plastid-targeted fusion protein**. *Mol Biol Evol* 2006, **23**:1437-1443.
13. Keeling PJ, Inagaki Y: **A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor  $\alpha$** . *Proc Natl Acad Sci USA* 2004, **101**:15380-15385.
14. Gile GH, Patron NJ, Keeling PJ: **EFL GTPase in cryptomonads and the distribution of EFL and EF- $\alpha$  in chromalveolates**. *Protist* 2006, **157**:435-444.
15. Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ: **Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzakii* and *Sphaeroforma arctica***. *J Eukaryot Microbiol* 2006, **53**:379-384.
16. Ragan MA, Murphy CA, Rand TG: **Are Ichthyosporea animals or fungi? Bayesian phylogenetic analysis of elongation factor  $\alpha$  of *Ichthyophonus irregularis***. *Mol Phylogenet Evol* 2003, **29**:550-562.
17. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schussler A, Longcore JE, O'Donnell K, Mozley-Stanbridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkman-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Unteriner WA, Lucking R, Budel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hobbitt DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R: **Reconstructing the early evolution of Fungi using a six-gene phylogeny**. *Nature* 2006, **443**:818-822.
18. Lewis LA, McCourt RM: **Green algae and the origin of land plants**. *Am J Bot* 2004, **91**:1535-1556.
19. Simon A, Glockner G, Felder M, Melkonian M, Becker B: **EST analysis of the scaly green flagellate *Mesostigma viride* (Streptophyta): implications for the evolution of green plants (Viridiplantae)**. *BMC Plant Biol* 2006, **6**:2.
20. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piegou B, Ball SG, Ral JP, Bouget FY, Pignoneau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, Moreau H: **Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features**. *Proc Natl Acad Sci USA* 2006, **103**:11647-11652.
21. Stanley MS, Perry RM, Callow JA: **Analysis of expressed sequence tags from the green algal *Ulva linza* (Chlorophyta)**. *J Phycol* 2005, **41**:1219-1226.
22. Lemieux C, Otis C, Turmel M: **Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution**. *Nature* 2000, **403**:649-652.
23. Lemieux C, Otis C, Turmel M: **A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies**. *BMC Biol* 2007, **5**:2.
24. Rodriguez-Ezpeleta N, Philippe H, Brinkmann H, Becker B, Melkonian M: **Phylogenetic analyses of nuclear, mitochondrial and plastid multi-gene datasets support the placement of *Mesostigma* in the Streptophyta**. *Mol Biol Evol* 2006.
25. Petersen J, Teich R, Becker B, Cerff R, Brinkmann H: **The GapA/B gene duplication marks the origin of Streptophyta (charophytes and land plants)**. *Mol Biol Evol* 2006, **23**:1109-1118.
26. Nedelcu AM, Borza T, Lee RW: **A land plant-specific multigene family in the unicellular *Mesostigma* argues for its close relationship to Streptophyta**. *Mol Biol Evol* 2006, **23**:1011-1015.
27. Turmel M, Otis C, Lemieux C: **The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants**. *Mol Biol Evol* 2002, **19**:24-38.
28. Harrison PJ, Waters RE, Taylor FJR: **A broad spectrum artificial medium for coastal and open ocean phytoplankton**. *J Phycol* 1980, **16**:28-35.
29. Pringsheim EG: **The biphasic or soil-water culture method for growing algae and flagellata**. *J Ecol* 1946, **33**:193-204.
30. Keeling PJ, Leander BS: **Characterisation of a non-canonical genetic code in the oxymonad *Streblospio trix***. *J Mol Biol* 2003, **326**:1337-1349.
31. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**:1572-1574.
32. Felsenstein J: **PHYLIP (Phylogeny Inference Package)**. 3.5th edition. Seattle, J. Felsenstein, University of Washington; 1993.
33. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing**. *Bioinformatics* 2002, **18**:502-504.
34. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood**. *Syst Biol* 2003, **52**:696-704.
35. Bruno WJ, Succi ND, Halpern AL: **Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction**. *Mol Biol Evol* 2000, **17**:189-197.
36. www.tree-puzzle.de: . .

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.biomedcentral.com/info/publishing\_adv.asp

