BMC Evolutionary Biology

Research article

Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids

Gildas Le Corguillé[†1,2], Gareth Pearson[†3], Marta Valente[3], Carla Viegas[3], Bernhard Gschloessl[4,5], Erwan Corre[1,2], Xavier Bailly[1,2], Akira F Peters[4,5], Claire Jubin[6,7,8], Benoit Vacherie[6], J Mark Cock[4,5] and Catherine Leblanc*[4,5]

Address: [1]CNRS, FR2424, Computer and Genomics Resource Centre, Station Biologique, Roscoff, France, [2]UPMC Univ. Paris 06, FR2424, Computer and Genomics Resource Centre, Station Biologique, Roscoff, France, [3]Centre of Marine Sciences, University of Algarve, Marine Ecology and Evolution, Faro, Portugal, [4]CNRS, UMR7139, Marine Plants and Biomolecules, Station Biologique, Roscoff, France, [5]UPMC Univ. Paris 06, UMR7139, Marine Plants and Biomolecules, Station Biologique, Roscoff, France, [6]CEA, DSV, Institut de Génomique, Genoscope, Evry, France, [7]CNRS, UMR 8030, Evry, France and [8]Université d'Evry, Evry, France

Email: Gildas Le Corguillé - lecorguille@sb-roscoff.fr; Gareth Pearson - gpearson@ualg.pt; Marta Valente - msvalente@ualg.pt; Carla Viegas - caviegas@ualg.pt; Bernhard Gschloessl - bernhard.gschloessl@univ-rennes1.fr; Erwan Corre - corre@sb-roscoff.fr; Xavier Bailly - bailly@sb-roscoff.fr; Akira F Peters - akirapeters@gmail.com; Claire Jubin - claire.jubin@gmail.com; Benoit Vacherie - bvacheri@genoscope.cns.fr; J Mark Cock - cock@sb-roscoff.fr; Catherine Leblanc* - leblanc@sb-roscoff.fr

* Corresponding author   †Equal contributors

Abstract

Background: Heterokont algae, together with cryptophytes, haptophytes and some alveolates, possess red-algal derived plastids. The chromalveolate hypothesis proposes that the red-algal derived plastids of all four groups have a monophyletic origin resulting from a single secondary endosymbiotic event. However, due to incongruence between nuclear and plastid phylogenies, this controversial hypothesis remains under debate. Large-scale genomic analyses have shown to be a powerful tool for phylogenetic reconstruction but insufficient sequence data have been available for red-algal derived plastid genomes.

Results: The chloroplast genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*, have been fully sequenced. These species represent two distinct orders of the Phaeophyceae, which is a major group within the heterokont lineage. The sizes of the circular plastid genomes are 139,954 and 124,986 base pairs, respectively, the size difference being due principally to the presence of longer inverted repeat and intergenic regions in *E. siliculosus*. Gene contents of the two plastids are similar with 139-148 protein-coding genes, 28-31 tRNA genes, and 3 ribosomal RNA genes. The two genomes also exhibit very similar rearrangements compared to other sequenced plastid genomes. The tRNA-Leu gene of *E. siliculosus* lacks an intron, in contrast to the *F. vesiculosus* and other heterokont plastid homologues, suggesting its recent loss in the Ectocarpales. Most of the brown algal plastid genes are shared with other red-algal derived plastid genomes, but a few are absent from raphidophyte or diatom plastid genomes. One of these regions is most similar to an apicomplexan nuclear sequence. The phylogenetic relationship between heterokonts, cryptophytes and haptophytes (collectively referred to as chromists) plastids was investigated using several datasets of concatenated proteins from two cyanobacterial genomes and 18 plastid genomes, including most of the available red algal and chromist plastid genomes.

Conclusion: The phylogenetic studies using concatenated plastid proteins still do not resolve the question of the monophyly of all chromist plastids. However, these results support both the monophyly of heterokont plastids and that of cryptophyte and haptophyte plastids, in agreement with nuclear phylogenies.

Background

The endosymbiotic captures of free-living prokaryotes, leading to the evolution of two types of organelles, mitochondria and plastids, are considered to be key events in the establishment and success of extant eukaryotic lineages [1,2]. If all mitochondria are likely to be derived from an α-proteobacterium-like ancestor, possibly due to a single and ancient endosymbiotic event, the history of plastid acquisition in the diverse photosynthetic eukaryotic lineages seems to be more complex [3-6]. It is now largely accepted that a single primary endosymbiotic event involving the capture of a cyanobacterium led to an ancestral primary plastid, which subsequently gave rise to the green plastids of the terrestrial plants and chlorophytes, the rhodoplasts of red algae and the cyanelles of the glaucophytes. Once established, primary red or green algal plastids later spread independently to other eukaryote lineages via secondary or tertiary endosymbioses, whereby a photosynthetic eukaryote was engulfed by another eukaryote. Subsequently, plastids have also been independently lost and/or replaced in several eukaryote lineages, making the reconstruction of plastid evolution very difficult.

The current consensus of eukaryote phylogeny recognizes six putative super-clusters: Opisthokonta, Amoebozoa, Plantae, Chromalveolata, Rhizaria, and Excavata [7,8], but this division is still debated [9,10]. The three primary plastid-containing lineages, Viridiplantae, Rhodophyta and Glaucophyta form the "Plantae" or "Archaeplastida" supergroup. Photosynthetic eukaryotes with secondary or tertiary plastids have evolved independently in the Chromalveolata, Rhizaria, and Excavata [3,5]. Among the secondary plastids, chlorophyll c-containing plastids have been shown to be derived from an ancestral red alga via a secondary endosymbiotic process that took place around one billion years ago [11,12]. This type of plastid is found in Cryptophyta, Haptophyta, Heterokonta (also called stramenopiles) and Dinophyceae algae [3,4]. Cryptophyta, Haptophyta and Heterokonta eukaryotic lineages have been grouped under the name of "Chromista" by Cavalier-Smith [13], and were later associated with the Alveolata, which includes the apicomplexans, dinoflagellates and ciliates, to form the "Chromalveolata" supergroup. In 1999, Cavalier-Smith proposed that all the chlorophyll c-containing plastids were derived from a single secondary endosymbiotic event and that the common ancestor of chromalveolates was originally photosynthetic [14]. During diversification of the four extant chromalveolates lineages, photosynthetic capacity and/or the plastid organelle would then have been independently lost several times in different eukaryotic lineages, such as oomycetes (non-photosynthetic heterokonts), apicomplexa or ciliates (non-photosynthetic alveolates). According to this so-called "chromalveolate" hypothesis, plastid and nuclear genomes have similar evolutionary histories and one would expect monophyly of chromalveolate lineages in both nuclear and plastid phylogenies. This hypothesis has been extensively debated over the last ten years (for recent references, [5,6,15-17]), in part because of incongruence between plastid and nuclear phylogenies [9].

At the nuclear level, both the monophyly of heterokonts and alveolates and that of cryptophytes and haptophytes have received increasing support in recent years (for recent review and references therein, [6]). Two contemporary phylogenetic analyses based on expressed sequences tag surveys of the cryptomonad *Guillardia theta* and the haptophyte *Emiliania huxleyi* supported the close relationship of cryptophyte and haptophyte host lineages [18,19]. In nuclear phylogenies alveolates and heterokonts often form a sister group [9,20]. Unexpectedly, several large scale nuclear phylogenies have also shown a very robust relationship between members of Rhizaria, cercozoans, and these two main clades of the "chromalveolates", but with the exclusion of haptophytes and cryptophytes [18,21,22]. The debate is becoming more complex with the emergence of this new putative SAR (stramenopiles/alveolata/rhizaria) supergroup, as proposed by Burki [23]. Recent phylogenetic studies employing large gene- and taxon-rich datasets continue to question the reality of the "chromalveolate" supergroup, by placing the haptophyte-cryptophyte clade as a sister group to the Plantae [24,25] or by having them emerging independently and separately from the SAR supergroup [10]. It is however well known that reconstructing the evolution of host cell lineages can be difficult, especially because of the chimeric nature of nuclear genomes and because large-scale horizontal gene transfers have occurred in some lineages during evolution [26].

Plastid genomes are less affected by horizontal gene transfer, with some rare exceptions [27]. At the plastid level, the monophyly of chromist plastids is supported by analyses of single genes [28], of small numbers of concatenated plastid genes [12,29], and of larger datasets of plastid-associated genes, i.e. plastid and nuclear-encoded plastid-targeted genes [30-35]. The relationships among chlorophyll c-containing plastids are, however, particularly hard to resolve and the results obtained are sometimes incongruent with host cell phylogenies [9]. Haptophyte plastid genes more often group with the heterokont/dinoflagellate clade, than with those of cryptophytes [30,31,33,34]. A clade grouping haptophyte and cryptophyte species has been inferred from some plastid gene phylogenies [31,33-35]. This clustering was not strongly supported and was highly dependent on the plastid gene dataset used [31,35] and/or on taxon-sampling [33,34]. Other variant topologies have included the placing of dinoflagellates either as a sister-group to haptophyte plastids [30,33] or to heter-

okont plastids [34,35]. However, a close evolutionary relationship between haptophyte and cryptophyte plastids would be consistent with the presence of a unique laterally transferred bacterial *rpl36* gene in both plastid genomes [27]. Other multigene analyses produced alternative results, such as low support for the chromist clade [29] or paraphyly of red-algal derived plastids [35,36].

The inability to recover congruencies between plastid and nuclear phylogenies, especially concerning haptophyte and cryptophyte monophyly, may be explained by poor taxon sampling of red algal and chromist species [31,36]. Until now, insufficient sequence data have been available for the chromalveolates, in terms of both nuclear and plastid genome sequences. In public databases, more than 110 complete plastid genomes are available from land plants and green algae, whereas less than 15 sequences belong to red algae or photosynthetic chromalveolate species. Only five complete plastid sequences have been reported for red algal species [36-39]. For the chromalveolates, with the exception of the highly diverged red-algal derived plastid genomes of non-photosynthetic apicomplexans [40] and those of dinoflagellates [41,42], complete plastid sequences have been published for two cryptomonads, *Guillardia theta* and *Rhodomonas salina* [11,31], one haptophyte, *Emiliania huxleyi* [43], 3 diatoms, *Odontella sinensis*, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* [44,45], one raphidophyte *Heterosigma akashiwo* [46] and one xanthophyte *Vaucheria litorea* [47].

Here we report the complete sequences of the plastid genomes of *Ectocarpus siliculosus* and *Fucus vesiculosus*. These sequences represent the first fully characterized plastid genomes from two distinct orders of Phaeophyceae, namely Ectocarpales and Fucales [48]. We have performed phylogenetic studies using large sets of genes and different reconstruction methods. The results still do not resolve the question of the monophyly of chromist plastids. However the topologies of concatenated plastid protein phylogenetic trees support both the monophyly of heterokont plastids and that of cryptophyte and haptophyte plastids, in agreement with nuclear phylogenies.

Results
Structure and gene content of the phaeophyte plastid genomes
The plastid genomes of *E. siliculosus* and *F. vesiculosus* are 139,954 and 124,986 base pairs (bp) in size, respectively, and both contain two inverted repeat regions (IR). These IRs divide the circular molecules into large (LSC) and small single copy (SSC) regions (Figure 1 and see general features of the two plastid genomes in additional file 1, Table S1). The size difference between the genomes was partly due to the presence of longer IRs of 8,615 bp in the

E. siliculosus cpDNA. The 4,863 bp *F. vesiculosus* IRs contain only the ribosomal RNA operons. Another reason for the difference in size between the two genomes is the presence of longer intergenic regions in the *E. siliculosus* cpDNA. These sequences represent about 20% of the genome, whereas only 14.5% of the *F. vesiculosus* cpDNA is intergenic. The overall GC content is 30.7% for *E. siliculosus* and 28.9% for *F. vesiculosus*. In both *Fucus* and *Ectocarpus*, the cpDNA IRs contain two ribosomal operons encoding 16S, 23S and 5S rRNA. The *F. vesiculosus* and *E. siliculosus* plastid genomes are predicted to encode a total of 139 and 144 protein-coding genes, and 26 and 27 tRNA genes, respectively, when the duplicated genes in the IRs are counted only once. An intron was identified in the *F. vesiculosus trnL2* gene, which encodes tRNA-Leu. Interestingly, its closest homologue in *E. siliculosus* cpDNA (93% nucleotide identity) does not possess an intron. The other tRNA-Leu genes in these plastid genomes, *trnL1_1* of *E. siliculosus* and *trnL1* of *F. vesiculosus*, present 98% sequence identity to each other and also lack the intron (Figure 2).

Gene organisation is highly similar between the two genomes and around two thirds of both molecules are conserved with respect to both gene identity and order. About 50% of each genome is incorporated into two large, locally collinear blocks. One block contains a large proportion of ribosomal protein-coding genes and covers up to 24% of the plastid genomes. The second block extends between *trnM* and *atpA* and covers 26-27.5% of each genome (Figure 1 and see the MAUVE analysis, provided in additional file 1, Figure S1). When compared to other heterokont plastid genomes, the number of genome rearrangements since the common ancestor of *E. siliculosus* and *F. vesiculosus* is comparable to the number of rearrangements that have occurred since the divergence of the three diatom species (see the reversal distance matrix provided in additional file 1, Table S2). This number increases more than twofold when higher taxonomic levels are considered (e.g., xanthophyte, raphidophyte or diatoms vs. brown algae).

The two plastid genomes are also very similar in terms of total gene content (Table 1). As already found in most of the green and red photosynthetic plastid genomes, excluding those of dinoflagellates [43], they possess the common core set of 44 genes, but with the exception of the *psbZ* gene (listed in additional file 2, Table S3). They also contain 42 additional protein-coding genes, which are only found in red algal and chromist plastid genomes, giving a total of 86 genes that are shared with the red plastid lineage (Table 1). These genes mainly encode essential plastid proteins, involved in transcription, protein synthesis and transport, and photosynthetic metabolism, such as components of ATP synthase, cytochrome, photosystem I and II complexes. Nine genes are shared by all the chro-
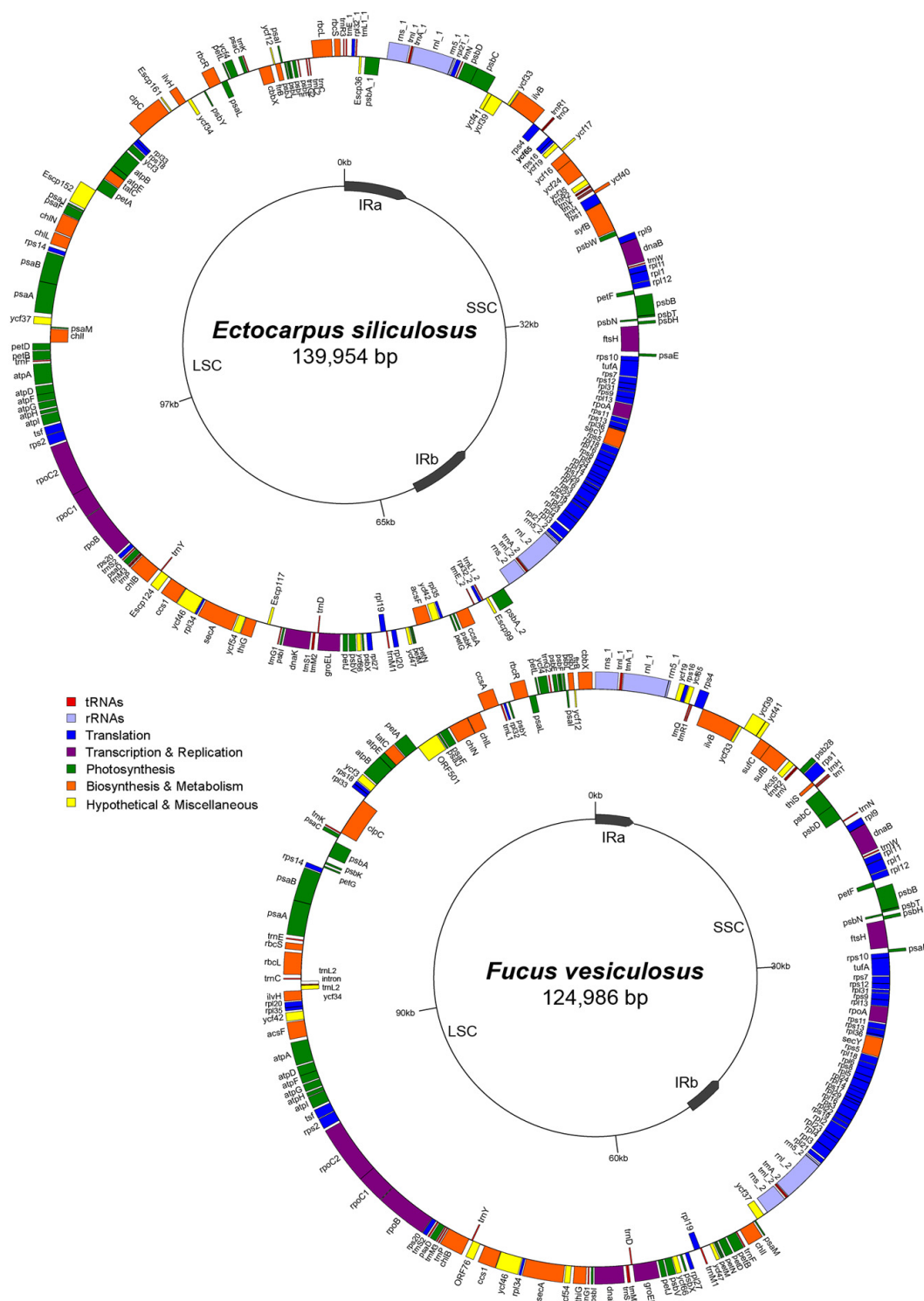
Figure 1
Plastid genome maps of *E. siliculosus* and *F. vesiculosus*. Genes on the outside of the circles are transcribed clockwise, whereas those on the inside counter clockwise. Annotated genes are colored according to the functional categories shown in the legend and the tRNA genes are indicated by the single-letter code of the corresponding amino-acid. Abbreviations: IR, inverted repeats; SSC, small single-copy region; LSC, large single-copy region.

A

```
              Stem      Stem  D-loop  Stem   Stem  Anticod Stem
trn           -------- T- -Y-- AR/GG/-A- --R-  R ----- YT--R- -----
Es_trnL1_1    -GCCAACC-TC-GCGG-AATTGGCAG-ACGC-G-CTAGA-TTTAGGT-TCTAG
Fv_trnL1      TACGAACC-TC-GCGG-AATTGGCAG-ACGC-G-CTAGA-TTTAGGT-TCTAG
Es_trnL2      -GGCGGTC-TC-GCGG-AATCGT-AG-ACGC-A-ACGGA-CTTAAAA-TCCGT
Fv_trnL2      -GGCAGCC-TC-GCGG-AATCGT-AG-ACGC-A-ACGGA-CTTAAAA-TCCGT
                                                        △
```

```
              Extra arm       Stem  Loop   Stem  Stem
trn           --/--Y---------  ---G TTCRA-Y C----  ------ -
Es_trnL1_1    TAAGGTTCTTGT----GAGAG-TTCAAGT-CTCTC-CGTTCGT-A
Fv_trnL1      TAAGATTCTTGT----GAGAG-TTCAAGT-CTCTC-CGTTCGT-A
Es_trnL2      TGATATATATAATTGT-GAGCG-TTCAAGT-CCTCG-CACCCC-A
Fv_trnL2      TGATATTTATAATCGT-GAGCG-TTCAAGT-CCTCG-CGCTCC-A
```

B

Figure 2
The canonical group I intron in the plastid tRNA-Leu (*trnL*) gene. (A) Multiple alignments of plastid *trnL* genes from *E. siliculosus* (Es) and *F. vesiculosus* (Fv), showing the position of the group I intron located in the *F. vesiculosus trnL2* gene. Structural features of tRNA are indicated on the two first lines. (B) Schematic phylogeny of the phaeophytes (redrawn from [48]), showing the presence (+) and absence (-) of the *trnL* intron in the different orders.

mist plastid genomes, but not with all the red algal plastid genomes (Table 1). Another 27 genes are encoded by most heterokont plastid genomes, but are not consistently present in the plastid genomes of haptophytes, cryptophytes and red algae. Of the 17 remaining genes that are common to *E. siliculosus*, *F. vesiculosus* and *V. litorea* cpDNAs, nine are present in the raphidophyte plastid genome, but all are absent from the diatom cpDNAs (Table 1).

Among the unknown plastid proteins, the conserved open reading frames (ORFs) *Ectocarpus Escp124* and *Fucus ORF76* encode putative proteins of 222 and 229 amino-acids, with 48% identity between species. Both protein sequences are predicted to possess five transmembrane helices. A homolog of these plastid proteins is also encoded by the plastid genome of the xanthophyte *V. litorea*. Interestingly, the most similar protein in the public databases is a nuclear-encoded protein, Tic20, found in several apicomplexa species, including *Toxoplasma* and *Plasmodium*. The C-terminal ends of these proteins also share weak similarity with the conserved hypothetical

plastid proteins encoded by the *ycf60* genes of plastid genomes from *E. huxleyi*, *G. tenuistitipata* and Cyanidiales (see partial multiple alignment provided in additional file 3, Figure S2).

Phylogenetic analyses
For phylogenetic analyses, three concatenated amino acid datasets were constructed (see additional file 2, Table S3) and analysed using maximum likelihood (ML), neighbour joining (NJ) and Bayesian inference (BI) methods. For the ML analyses, cpREV and JTT amino acid substitution matrices gave the same tree topologies (data not shown). Trees were constructed using a dataset of 44 proteins (8,652 amino-acid positions) from a broad range of species, including 13 taxa of red-algal type plastids, 4 taxa of Viriplantae, the glaucophyte *Cyanophora*, and two cyanobacteria (see additional file 2, Table S4 for species list). Plastid sequences of chlorophyll-c-containing dinoflagellates were not included in the analyses because this would have resulted in a significantly reduced common protein dataset. All but four of the nodes in the trees were well resolved and supported by the three different methods (Figure 3). As observed in previous studies, the red-algal and red-derived type plastid sequences grouped together, whereas green plastids formed a separate monophyletic group, derived from the cyanobacterial sequences. In all our analyses, the glaucophyte plastid from *Cyanophora* emerged at the base of the green plastids, with high confidence in the BI analysis but with low bootstrap support in the ML and NJ analyses (56 and 66%). Among the green plastids, the order of branching of *Mesostigma* and *Arabidopsis* was not fully resolved, but the phylogenetic position of *Mesostigma* within the Streptophyta has been studied recently, with expanded taxon sampling of the Viridiplantae [49]. In the other part of the tree, the Cyanidiales grouped together outside a strongly supported clade that includes the Florideophyceae and Bangiophyceae, together with the heterokont, the haptophyte and the cryptophyte plastids. The trees also strongly grouped all heterokont plastids together, with a split between diatom plastid sequences and those of the raphidophyte and phaeophytes. The Florideophyceae and Bangiophyceae branched together with high confidence using all the methods, as did the two species of cryptophytes. In these phylogenetic studies, the haptophyte *E. huxleyi* emerged as the closest branch to cryptophytes in the BI analysis but this topology had low bootstrap support in the ML analysis (67%), and no support in the NJ analysis. The order of branching of the following three major groups: heterokonts, (florideophyte+bangiophyte), and (cryptophytes+haptophyte), was also uncertain. In fact, the clade of heterokonts and (cryptophyte+haptophyte) plastids was only well-supported by the BI analysis, and very poorly (49%) or not supported by the ML and NJ analyses, respectively.

Table 1: Gene content comparisons between plastid genomes.

Genes	No.	Heterokonts							Hapto	Crypto		Rhodophytes			
		Esil	Fves	Vlit	Haka	Ptri	Osin	Tpse	Ehux	Gthe	Rsal	Gten	Ppur	Ccal	Cmer
83 genes listed in Table S3*, rpl36, rbcR, ccsA	86	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpoC1, rpoC2, ycf35, ycf46	4	+	+	+/-	+	+	+	+	+	+	+	+	+	+/-	-
cbbX, ccs1, tatC	3	+	+	+	+	+	+	+	+	+	+	+	-	+/-	+
petL, petM	2	+	+	+	+	+	+	+	+	+	+	-	+	-	+
thiG, thisS	2	+	+	+	+	+	+	+	+	-	-	+	+/-	+	+
dnaB, ftsH, petF, psaE, psbX, rpl1, rpl4, rpl11, rpl12, rpl13, rpl18, rpl24, rpl29, rpl32, rpl35, rps20, sufC, ycf33	18	+	+	+	+	+	+	+	-	+	+	+	+	+/-	+
psbY	1	+	+	+	+	+	-	+	-	-	+	-	-	-	+
ycf41, ycf42, ycf66	3	+	+	+	+/-	+	+/-	+	-	-	-	-	-	-	-
psbW, secG, tsf	3	+	+	+/-	-	+	+	+	+/-	+	+	+/-	+/-	+/-	+/-
ftrB, ilvB, ilvH	3	+	+	+	+	-	-	-	-	+	+	+	+	+/-	+/-
ycf65	1	+	+	+	+	-	-	-	+	-	-	+	-	+	+
acsF, petJ, rps1, ycf34, ycf54	5	+	+	+/-	+	-	-	-	-	-	-	+/-	+/-	+/-	+/-
chlB, chlL, chlN, rpl9	4	+	+	+	-	-	-	-	-	-	-	+/-	+/-	+/-	+/-
ycf19, ycf37	2	+	+	+	-	-	-	-	+/-	+	+	+	+	+	+/-
Escp124 (FvORF76) ~ ycf60	1	+	+	+	-	-	-	-	+	-	-	+	-	+	+
Escp152 (FvORF501)	1	+	+	-	-	-	-	-	-	-	-	-	-	-	-
hlip ~ ycf17	1	+	-	+	-	-	-	-	-	+	+	-	+	+	+
syfB	1	+	-	-	-	+	-	-	-	-	-	+	+	-	-
Escp36, Escp117, Escp161	3	+	-	-	-	-	-	-	-	-	-	-	-	-	-

E. siliculosus (Esil) and *F. vesiculosus* (Fves) cpDNAs were compared with those of other heterokonts, haptophyte (Hapto), cryptophytes (Crypto) and rhodophytes. Presence (+), absence (-) or presence of only a subset (+/-) of the genes listed in the first column is indicated. Species abbreviations: *Vaucheria litorea* (Vlit), *Heterosigma akashiwo* (Haka), *Phaeodactylum tricornutum* (Ptri), *Odontella sinensis* (Osin), *Thalassiosira pseudonana* (Tpeu), *Emiliania huxleyi* (Ehux), *Guillardia theta* (Gthe), *Rhodomonas salina* (Rsal), *Gracilaria tenuistitipata* (Gten), *Porphyra purpurea* (Ppur), *Cyanidium caldarium* (Ccal), *Cyanidioschyzon merolae* (Cmer).
(*) see additional file 2.

To strengthen the topology of branching in the region of the tree corresponding to the red-alga derived plastids, we decided to increase the protein dataset by focusing the phylogenetic studies on 13 species. A full dataset of 83 plastid-encoded proteins (16,738 amino acid positions) was analyzed in parallel with a sub-dataset of 33 slowly-evolving plastid proteins, excluding the fast-evolving proteins (Figure 4). Using the PhyloBayes software, the values of the saturation index have been calculated for each dataset. The observed and predicted homoplasy rates are, respectively, 1.98 ± 0.05 and 2.00 ± 0.05 for the 83-protein dataset, and 1.01 ± 0.03 and 1.00 ± 0.04 for the 33-protein dataset. These results show that the exclusion of the fast-evolving proteins tends to decrease the global level of saturation. Both trees still showed two well-supported plastid groups, corresponding to heterokonts and the Cyanidiales. Globally, the branches that were strongly supported by the 44-protein dataset were maintained. Interestingly, the group formed by haptophyte and cryptophyte plastids had greater support in the ML analysis (97% bootstrap value) but little support with NJ method

with the 83-protein dataset (Figure 4A) and was strongly supported by the three methods in the analyses of the slowly-evolving proteins (Figure 4B). Compared to the 44-protein trees, the 83- and 33-protein trees differed in their branching patterns with respect to the (florideophyte+bangiophyte) and the (cryptophytes+haptophyte). Both the ML and NJ trees built with the dataset of 83 proteins clustered these two groups with high bootstrap values, whereas the red algal plastids were found outside the clade of heterokont/(cryptophyte+haptophyte) plastids in the 33-protein trees. This latter topology was strongly supported in the ML, NJ and BI analyses (Figure 4B).

To further test these phylogenetic positions, we compared different topologies by performing the approximately unbiased (AU) and Shimodaira-Hasegawa (SH) tests (Figure 5). Four topologies were selected to evaluate two hypotheses: 1) Are chromist plastids indeed monophyletic; 2) Are haptophyte plastids specifically related to cryptophyte plastids to the exclusion of heterokont or (florideophyte+bangiophyte) plastids? Our analyses
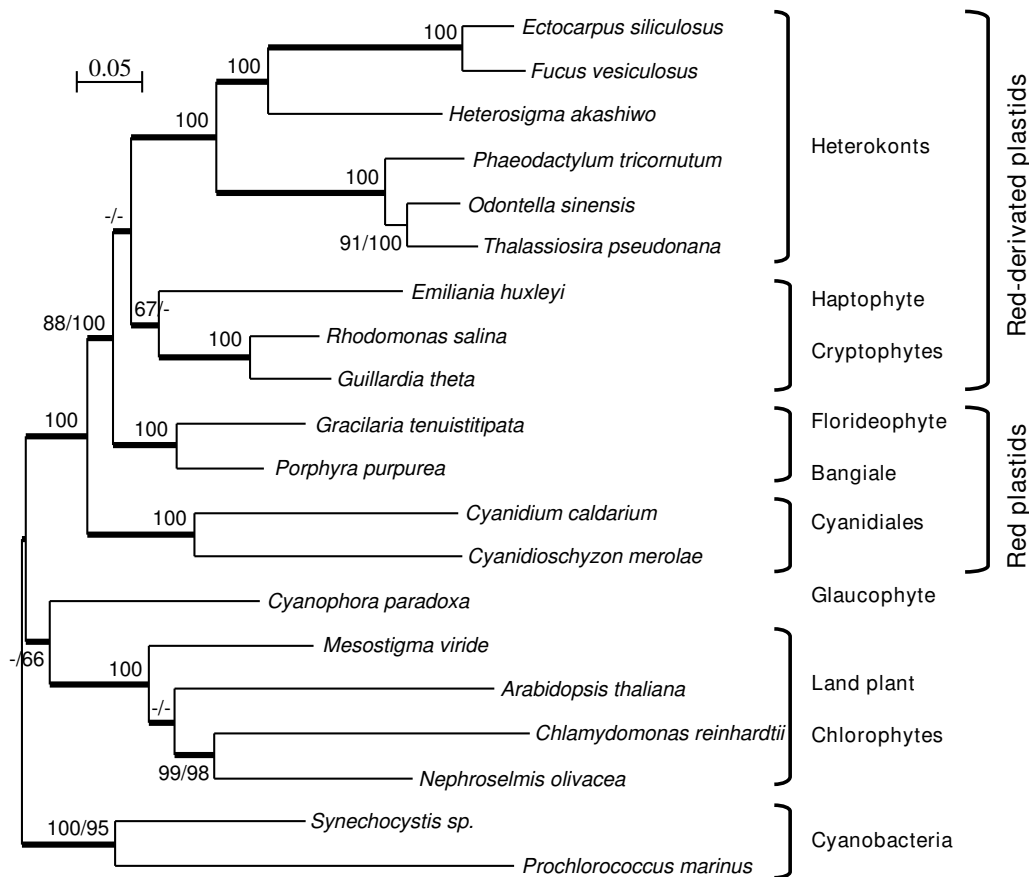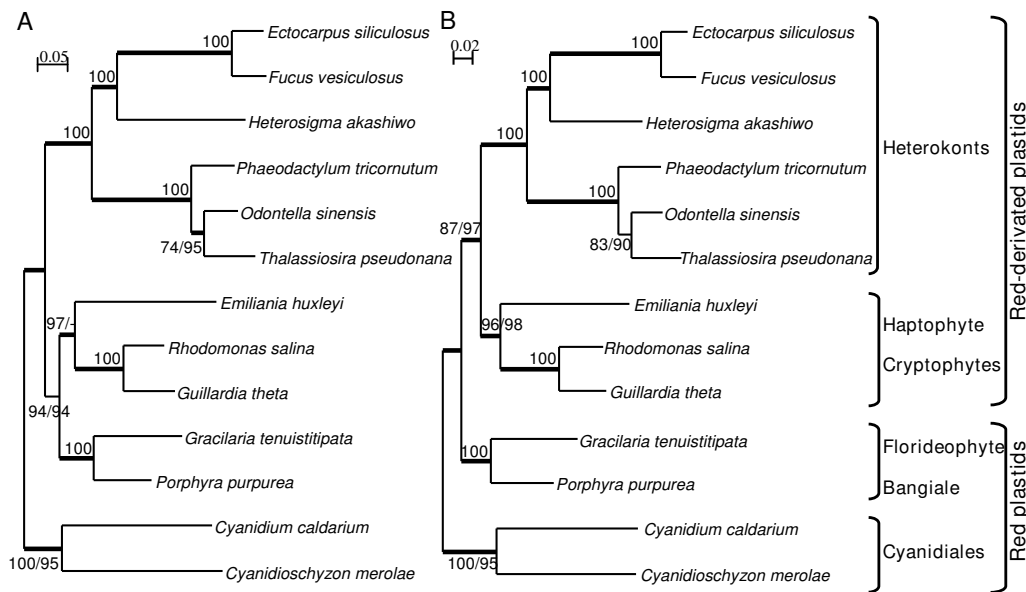
Figure 3
Maximum likelihood tree constructed from a dataset of 44 concatenated proteins from 20 plastid or cyanobacterial complete genomes. PHYML and Neighbour Joining trees were constructed based on 8,652 amino-acid sites using cpREV and JTT matrices, respectively. When above 65% and different, bootstrap values (1000 replicates) are provided for PHYML (first value) and NJ (second value) analyses. The thick branches represent ≥ 0.9 posterior probability for Bayesian inference analysis.

showed that, for the 83- and 33-protein datasets, the best topologies correspond to the trees shown in Figure 4A (topology I) and 4B (topology II), respectively. Considering the two datasets, these two topologies had a much higher likelihood in AU and SH tests, than topologies that place either the haptophyte plastid outside a (cryptophyte+(florideophyte+bangiophyte)) clade (topology III) or that propose that the closest relationship is between heterokont and haptophyte plastids (topology IV). For the 83-protein dataset, the three topologies (II, III and IV) were significantly rejected with p value under 0.05 for AU tests, but not for SH tests. For the 33-protein dataset, the topology I could not be significantly rejected by both tests ($P = 0.09$; $P = 0.24$), whereas the other topologies were refuted with P values below the significance level.

Discussion

Monophyly and evolution of heterokont plastid genomes

Until very recently, all of the plastid genomes available for the heterokont lineage were from diatoms (*O. sinensis*, *P. tricornutum* and *T. pseudonana*), and these genomes featured conserved gene content and gene clusters [45]. Along with the recently published plastid genomes of two strains of the raphidophyte *H. akashiwo* [46]) and the xanthophyte *V. litorea* [47], the complete sequences of the *E. siliculosus* and *F. vesiculosus* plastid genomes presented here significantly increase the number and diversity of heterokont plastid genomes available, allowing a more extensive comparison of these genomes. Our results support a unique origin for all heterokont plastids, based on similarity in terms of gene content (Table 1) and on their forming a strongly supported group in all our phylogenetic analyses (Figures 3 and 4). These analyses were, therefore, consistent with the well established mono-

Figure 4
Maximum likelihood trees constructed from two datasets of concatenated proteins from 13 completed plastid genomes of red algal and chromist species. A) The full dataset of 83 proteins (16,738 amino-acid sites), and B) the 33 slow-evolving proteins dataset (8,404 amino-acid sites) were used for PHYML and Neighbour Joining analyses, using cpREV and JTT matrices, respectively. When above 65% and different, bootstrap values (1000 replicates) are provided for PHYML (first value) and NJ (second value) analyses. The thick branches represent ≥ 0.9 posterior probability for Bayesian inference analysis.

phyletic origin of the heterokont host cell [10,21,23]. However, despite their common origin, genome comparisons revealed specific traits in the evolution of heterokont plastids during the diversification of the different heterokont orders.

All the Xanthophyceae or Phaeophyceae plastid genomes analyzed to date, including that of *F. vesiculosus* described here, contain a tRNA-Leu gene with a single intron [47,50]. This canonical group I intron is thought to have been acquired from the ancestral cyanobacterial endosymbiont and to have been lost independently in several lineages of plastids, including the red algae and almost all their secondary plastid derivatives, except the Xanthophyceae/Phaeophyceae lineage [50]. Given the high sequence similarities found between these plastid tRNA-Leu genes in *V. litorea*, *F. vesiculosus* and *E. siliculosus* (86 to 93% sequence identity), they are probably derived from the same ancestral tRNA-Leu gene, containing the endosymbiotic derived intron. In the *E. siliculosus* gene, its loss is likely to be recent because it is still present in the plastid tRNA-Leu genes of Laminariales species and of two Ectocarpales, *Pylaiella littoralis* and *Scytosiphon lomentaria* (Figure 2) [50]. This feature is evidence for continued evolution of brown algal plastid genomes within the recently-derived order Ectocarpales [48,51].

In terms of gene content, the brown algal plastid genomes seem to be more closely related to those of *V. litorea* and of *H. akashiwo* than to those of diatoms and this is consistent with evolutionary relationships of the nuclear compartment [51,52]. Although the structural organisation of plastid genomes is highly conserved within the brown algae (additional file 1, Figure S1) and within diatoms [45], there is evidence of intensive gene rearrangements having occurred earlier in evolution after the separation of diatoms from raphidophytes, xanthophytes and phaeophytes. Moreover, more extensive gene losses seem to have occurred in diatom plastid genomes than in other heterokonts (Table 1). These genes could have been transferred to the nucleus or replaced by bacterial counterparts, functionally-integrated through horizontal gene transfer as often seen in the diatom nuclear genome [53]. All these data, together with the topologies of plastid phylogenetic trees (Figure 3 and 4) support a relatively ancient split between diatoms and the raphidophyte-phaeophyte clade, in agreement with the early divergence of the Bacillariophyceae from the other photosynthetic heterokont lineages in nuclear phylogenies [51,52].

What is the closest relative of the heterokont plastid clade?
A critical step for the transformation of the endosymbiont into a permanent organelle was the establishment of an efficient protein targeting and translocation system from
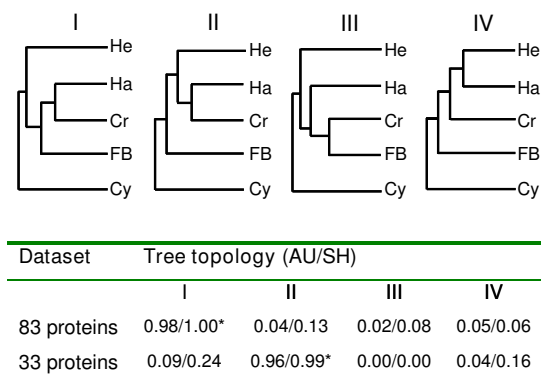
Figure 5
Likelihood AU and SH tests of four alternative tree topologies, using the two different datasets. In tree topologies, the abbreviations used are: He, Heterokont plastids; Ha, Haptophyte plastids; Cr, Cryptophyte plastids; FB, Florideophyte+Bangiophyte plastids; Cy, Cyanidiales plastids. In the table, the best tree is indicated by a star. Boldface type corresponds to *P* < 0.05.

the nucleus to the plastid [1,4]. The canonical Tic/Toc protein import complex of secondary plastids was inherited from the first red-algal endosymbiont, with components of both eukaryotic and eubacterial origin [1,54,55]. Both brown algal plastid genomes have a gene (*Escp124* in *Ectocarpus* and *ORF76* in *Fucus*) that shares similarity with the Tic20-like genes in xanthophyte, haptophyte and red algal plastid genomes. There are no homologues of this gene in raphidophyte, diatom and cryptophyte plastid genomes (Table 1). This plastid-encoded Tic20 gene (also called *ycf60*) encodes a small membrane protein and is thought to be endosymbiont-derived with a cyanobacterial origin [1,54,55]. Interestingly, the highest similarity scores of brown algal and xanthophyte plastid ORFs were found with a homologous protein encoded in the nucleus of several apicomplexan species, including *Toxoplasma* and *Plasmodium*. In *T. gondii*, this Tic20-like protein has been shown to be essential for protein import into the apicoplast [56] and is therefore likely to be linked to apicoplast evolution [4]. Escp124 and ORF76 protein sequences are also predicted to have five transmembrane regions, suggesting a putative location in the plastid membrane. It is now widely accepted that alveolates and heterokonts are derived from a common host cell ancestor. *Escp124* and *ORF76* could be footprints of a common photosynthetic ancestor of heterokonts and apicomplexans. This hypothesis is in agreement with several recently published studies suggesting that contemporary alveolates are derived from a photosynthetic ancestor. These studies include the characterization of a photosynthetic alveolate closely related to apicomplexan parasites [57], the identification

of plastid-derived genes in a non-photosynthetic alveolate [58] and the identification of remnant algal-related genes in ciliates [59].

Is the monophyly of chromist plastids still in doubt?

All the phylogenetic analyses carried out in this study suggest that the red algal ancestor of chromist plastids was more closely related to the more recently evolved red algae (Florideophyceae and Bangiophyceae) than to Cyanidiales, confirming the report by Sanchez-Puerta et al. [33]. It is worth mentioning that Cyanidiales are extremophile unicellular red algae and have been shown to be the earliest diverging red algal group. They emerge very distinctly from the other multi-cellular red algal taxa in nuclear phylogenies [29]. Within the chromist plastid clade, most plastid phylogenies have hitherto featured a clade grouping haptophyte and heterokont plastids [29,30] and the relationship between haptophyte and cryptophyte plastids was never strongly recovered in previous studies [31,33-35]. These conflicting results have been discussed in the light of taxon- or data-sampling limitations [31,34]. Our results do not support a preferential link between heterokont and haptophyte plastids, neither in terms of gene content (Table 1) nor phylogenetic relationship. Moreover, these phylogenetic analyses strongly support the monophyly of haptophyte and cryptophyte plastids (Figure 4). In general, addition of taxa has been shown to reduce support for previously robust clades, whereas the addition of more positions has been shown to increase support regardless of the topology [60]. Indeed this topology has high confidence, especially when the dataset of genes was increased or slowly-evolving proteins were selected. Moreover, whatever the datasets used, with or without fast-evolving proteins, AU tests significantly rejected topologies separating haptophyte and cryptophyte plastids. The monophyly of haptophyte and cryptophyte plastids is in complete agreement with recent nuclear phylogenies that support a common origin of their host cells [18,19] and with a previous study that identified a unique, laterally transferred bacterial gene in plastid genomes from these two groups [27].

Horizontal gene transfers into plastid genomes happened only rarely after the establishment of the endosymbiont within the host cell. The major events which can affect the structure of the organelle genome are gene transfer to the nucleus and/or gene loss. Indeed, red algal plastid genomes possess more than 230 protein-coding genes while those derived from a red-algal endosymbiont encode less than 150, of which more than half are shared by all the genomes (Table 1). An exceptional case is the drastic reduction of plastid minicircular genomes of peridinean dinoflagellates [41]. In other plastid genomes derived from a red algal endosymbiont, the remaining pool of genes is the result of losses that have occurred

independently in the different lineages and of retention that could constitute interesting fingerprints of ancestral plastid gene contents. A comparison of gene content did not reveal any particular relationships between heterokonts and cryptophytes/haptophytes and therefore did not provide support for a common history. For the phylogenetic analyses, whereas the use of the complete dataset supported a different red-algal origin for heterokont plastids (Figure 4A), monophyly of all chromist plastids was recovered when the most conservative data was used in the phylogenetic reconstruction (Figure 4B), as previously observed [33,36]. Other studies have also shown the disruption of the monophyly of chromist plastids [31,33,35]. Our dataset and taxa sampling are not sufficient to completely refute or confirm the polyphyly of chromist plastids, given that the monophyletic topology does not significantly exclude the polyphyletic one when using the slow-evolving proteins (Figure 5). The slowly evolving proteins may reflect more ancient divergences, but the exclusion of fast-evolving proteins decreases the number of analysed amino-acid positions by a factor of two and the issue of dataset size is critical in plastid multi-gene phylogenetic studies [34]. In the context of the chromalveolate hypothesis, the major separation between cryptophyte/haptophyte and heterokont/alveolate host cells is more likely to have occurred very early after the secondary endosymbiosis. An alternative origin of heterokont/alveolate plastids has recently been proposed, with laterally transferred red-algal derived plastids from the haptophyte/cryptophyte clade into the heterokont/alveolate lineage [5,61]. The monophyly of all chromist plastids is also consistent with this tertiary endosymbiosis hypothesis, if the heterokont plastids were captured before the divergence between the haptophyte and cryptophyte host lineages. It is however clear that plastid phylogenies alone will not resolve these currently discussed questions about vertical or lateral inheritance of red-algal derived plastids [16,17].

It has been shown that plastid metabolism could also involve a significant number of nuclear-encoded proteins recruited from diverse origins, such as laterally transferred genes from Chlamydiae [62] or green algae [63-65]. Phylogenies based on nuclear-encoded plastid-targeted proteins could then trace and reflect complex evolutionary pathways, whereas phylogenies based on complete sets of plastid-encoded genes should better reflect the evolution of the organelle since its engulfment by the host cell. As illustrated by the high resolution of the heterokont plastid clade, additional plastid genomes from haptophytes, cryptophytes and dinoflagellates, but certainly also from other evolved red algae will be required to fully resolve chromist plastid phylogenies and, subsequently, test the different hypotheses concerning red-algal derived plastid origin(s).

Conclusion

In conclusion, this study of two novel plastid genomes belonging to brown algal species has shown the importance of increased taxon sampling when analysing phylogenetic relationships based on large datasets. As expected, the phylogenetic analyses showed that heterokont plastids are monophyletic, although very diverse in terms of gene arrangement. There is also evidence that some heterokont (phaeophyte and xanthophyte) plastids have retained finger-prints indicating a common ancestory with alveolate plastids. Moreover, monophyly of haptophyte and cryptophytes plastids was strongly recovered whatever the dataset or the method used, in complete agreement with large-scale nuclear phylogenies.

Methods

Algal material and DNA extraction

E. siliculosus strain Ec32 (CCAP1310/4) was cultivated under laboratory conditions as previously described [66] and total DNA was prepared according to the method of Apt et al. [67].

F. vesiculosus was collected from the field (Ria Formosa Natural Park, Portugal) and DNA was extracted from isolated plastids. Briefly, 20 g apical tissue free from visible epiphytes was cleaned by 2 min exposure in bleach (1% in filtered natural seawater), rinsed and homogenized in 100 mL cold extraction buffer containing 0.05 M MES (pH 6.1), 0.5 M sorbitol, 1 mM $MgCl_2$, 1 mM MnCl, 0.5 mM K_2HPO_4, 5 mM EDTA, 1% BSA, 2% PVP, and 2 mM Na-ascorbate. The homogenate was passed through cotton gauze and 1 μm nylon mesh, centrifuged for 2 min at 2000 × *g* at 4°C. The supernatant was transferred to new 50 mL tubes and centrifuged at 5000 × *g* for 5 min. The pellet containing plastids was gently resuspended in a total of 10 mL of extraction buffer and re-centrifuged (5 min, 5000 × *g*, 4°C). The pellet was resuspended in new extraction buffer and applied to a 30:50% sucrose step gradient. After centrifugation for 45 min at 5000 × g (4°C), the plastids were removed from the 30 and 50% sucrose interface, carefully resuspended in a buffer containing 0.05 M HEPES (pH 7.5), 0.5 M sorbitol, 1 mM $MgCl_2$, 1 mM MnCl, 0.5 mM K_2HPO_4. After observation under the microscope to determine the quality of the plastid preparation, plastids were centrifuged again for 10 min at 5000 × g. The supernatant was removed and plastids were stored at -80°C prior to DNA extraction using the CTAB method [68].

Genome Sequencing, Assembly and Annotation

For *E. siliculosus*, several scaffolds corresponding to plastid DNA were detected by similarity to other plastid genomes in an assembly of shotgun sequenced total genomic DNA produced by Genoscope http://www.genoscope.cns.fr/spip/-Ectocarpus-siliculosus-.html. These scaffolds were

removed from the rest of the sequence data and the sequence of the circular genome was completed by manual assembly and PCR amplification of gap regions. The plastid genome was annotated using the GenDB interface [69], available through the bioinformatics' facilities of the Marine Genomics Europe Network of Excellence.

For *F. vesiculosus*, two main strategies were used to obtain the full genome sequence: 1) Plastid-enriched DNA (cpDNA) was digested (*Hind*III), and cloned into pBluescript II (SK-) (Stratagene). Positive colonies were randomly picked and those with inserts > 1 Kb after digestion were end-sequenced. 2) Plastid DNA was used to make uncloned, adaptor-ligated libraries for a genome-walking approach using long-distance PCR (GenomeWalker kit, Clontech, Palo Alto, USA). Gaps in the genome were filled by PCR, based on predicted gene organization in red-lineage plastids. The *F. vesiculosus* plastid genome was assembled using CodonCode Aligner (CodonCode Corp., USA). Protein coding genes and putative open reading frames (ORFs) were identified by database comparison (Blastx, [70]) and online tools (ORF Finder, NCBI). Ribosomal and tRNA genes were identified using RNAmmer http://www.cbs.dtu.dk/services/RNAmmer/[71] and ARAGORN http://130.235.46.10/ARAGORN/[72], respectively.

The two plastid sequences are available under the following EMBL accession numbers: *E. siliculosus* (FP102296) and *F. vesiculosus* (FM957154). The physical maps of the circular genome were drawn using GenomeVx (freely available at wolfe.gen.tcd.ie/GenomeVx/).

Phylogenetic Analyses
For global gene content comparisons, the two brown algal plastid genomes were analysed together with those of the xanthophyte *V. litorea* [47] and the raphidophyte *H. akashiwo* [46] plus the 15 algal sequences and the two reference cyanobacterium genomes analysed by Khan et al. [31]. The phylogenetic analyses were conducted with a total of two cyanobacterium and 18 plastid genomes, including four complete genomes from red algae and nine from chromist species (see additional file 2, Table S4). Three concatenated protein datasets were constructed from these genomes (additional file 2, Table S3). The first dataset corresponded to the 44 plastid protein-coding genes shared by all 20 species. In addition, a larger dataset of 83 proteins was built using all the plastid proteins common to the 13 red, cryptophyte, haptophyte and heterokont algae. A list of gene synonyms used during this study is provided in additional file 2 (Table S5), together with complementary gene annotation information. Single and concatenated protein sequences were aligned using MUSCLE [73] and each alignment was further optimised using GBlocks [74]. Datasets for individual genes were

first analysed using maximum likelihood, in order to eliminate genes derived from horizontal transfer. Only the *rpl36* protein phylogeny suggests a non red-algal origin for the haptophyte and cryptophyte genes, which grouped far outside the red algal and heterokont cluster, as previously reported [27]. This gene was therefore eliminated from the full 83-protein dataset. The average distance was calculated for each protein with Tree-Puzzle [75]. We excluded 50 "fast-evolving" protein sequences to produce a dataset of 33 "slowly-evolving" proteins, which present an average distance under the threshold of 0.6. This value was chosen in order to conserve at least half of the analysed positions for the 33-protein dataset.

Phylogenetic analyses of concatenated protein data were carried out on 8,652, 16,738 and 8,404 amino acids corresponding, respectively, to the 44-, 83- and 33-protein datasets. A Maximum Likelihood (ML) approach was used to reconstruct phylogenetic trees using PHYML [76] under both cpREV [77] and JTT [78] amino acid substitution matrices with 4 gamma-distributed rate categories and estimated invariable sites. The neighbor-joining (NJ) method was performed with JTT amino acid substitution matrix using the Phylip software package [79]. For both the ML and NJ methods, bootstrap analyses of 1,000 replicates were used to provide confidence estimates for the phylogenetic tree topologies. Finally, Bayesian inference (BI) analyses were performed with PhyloBayes 3.1d [80] using 4 gamma-distributed rate categories. PhyloBayes was run using the site-heterogeneous CAT model as described in Lartillot et al. [81] and two independent chains with a total length up to 25,000 cycles, discarding the first 25% as burn-in and calculating the posterior consensus tree. Furthermore, a saturation test was performed on the different datasets to calculate the observed and predicted homoplasy rates as described in the PhyloBayes user manual.

To statistically test the topologies of the trees, approximately unbiased (AU) and Shimodaira-Hasegawa (SH) analyses were performed on four topologies. These were selected to reflect the relative positions of haptophyte, cryptophyte and heterokont plastids and were generated by rearrangement of ML and NJ trees (if required). Site likelihoods for each topology were calculated using Tree-Puzzle on the two different concatenated datasets and the AU/SH tests were performed using CONSEL 0.1 [82].

Authors' contributions
GLC, BG, CL annotated the *E. siliculosus* cpDNA. GLC, CL carried out the phylogenetic analysis. GP, MV, CV sequenced and assembled the cpDNA of *F. vesiculosus*. GP annotated the *F. vesiculosus* genome. GLC, GP, CL performed the comparative genomic analyses on both plastid genomes. AFP obtained and provided *E. siliculosus* cul-

tures. CJ, BV sequenced and provided plastid contigs of *E. siliculosus*. EC, XB participated in design of phylogenetic and statistical approaches. GLC, GP, JMC contributed to manuscript writing. JMC helped to supervise the project. CL conceived and designed the project, wrote the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Additional data and analyses on the plastid genomes of E. siliculosus *and* F. vesiculosus*. Tables S1 and S2, general features of the two phaeophyte plastid genomes and reversal distance matrix for pairwise comparisons between heterokont plastid genomes. Figure S1, MAUVE genome comparison between* E. siliculosus *and* F. vesiculosus *plastid genomes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-253-S1.PDF]

Additional file 2

Genes and genomes used in the phylogenetic studies. Tables S3 to S5, lists of genes, accession numbers of plastid and bacterial genomes and complementary information about gene synonyms and about some protein-encoded genes used in the phylogenetic studies.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-253-S2.PDF]

Additional file 3

Partial multiple alignment of Tic20 and yfc60 protein homologs. Figure S2 showing partial multiple alignment of Tic20 and yfc60 protein homologs from red alga-derived plastid and apicomplexan genomes.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-253-S3.PDF]

References
1. Dyall SD, Brown MT, Johnson PJ: **Ancient invasions: from endosymbionts to organelles.** *Science* 2004, **304:**253-257.
2. Bhattacharya D, Archibald JM, Weber APM, Reyes-Prieto A: **How do endosymbionts become organelles? Understanding early events in plastid evolution.** *BioEssays* 2007, **29:**1239-1246.
3. Reyes-Prieto A, Weber APM, Bhattacharya D: **The origin and establishment of the plastid in algae and plants.** *Annu Rev Genet* 2007, **41:**147-168.
4. Gould SB, Waller RF, McFadden GI: **Plastid Evolution.** *Annu Rev Plant Biol* 2008, **59:**491-517.
5. Archibald JM: **The Puzzle of Plastid Evolution.** *Curr Biol* 2009, **19:**R81-R88.
6. Keeling PJ: **Chromalveolates and the evolution of plastids by secondary endosymbiosis.** *J Eukaryot Microbiol* 2009, **56:**1-8.
7. Keeling PJ: **Diversity and evolutionary history of plastids and their hosts.** *Am J Bot* 2004, **91:**1481-1493.
8. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR: **The new higher level classification of eukaryotes with emphasis on the taxonomy of protists.** *J Eukaryot Microbiol* 2005, **52:**399-451.
9. Parfrey L, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, Patterson D, Katz L: **Evaluating support for the current classification of eukaryotic diversity.** *PLoS Genet* 2006, **2:**e220.
10. Yoon HS, Grant J, Tekle YI, Wu M, Chaon BC, Cole JC, Logsdon JMJ, Patterson DJ, Bhattacharya D, Katz LA: **Broadly sampled multigene trees of eukaryotes.** *BMC Evol Biol* 2008, **8:**14.
11. Douglas S, Penny S: **The plastid genome of the cryptophyte alga,** *Guillardia theta*: **complete sequence and conserved synteny groups confirm its common ancestry with red algae.** *J Mol Evol* 1999, **48:**236-244.
12. Yoon HS, Hackett JD, Pinto G, Bhattacharya D: **The single, ancient origin of chromist plastids.** *Proc Natl Acad Sci USA* 2002, **99:**15507-15512.
13. Cavalier-Smith T: **A revised six-kingdom system of life.** *Biol Rev* 1998, **73:**203-266.
14. Cavalier-Smith T: **Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree.** *J Eukaryot Microbiol* 1999, **46:**347-366.
15. Braun EL, Phillips N: **Phylogenomics and secondary plastids: a look back and a look ahead.** *J Phycol* 2008, **44:**2-6.
16. Bodyl A, Stiller J, Mackiewicz P: **Chromalveolate plastids: direct descent or multiple endosymbioses?** *Trends Ecol Evol* 2009, **24:**119-121.
17. Lane CE, Archibald JM: **Reply to Body, Stiller and Mackiewicz: "Chromalveolate plastids: direct descent or multiple endosymbioses?".** *Trends Ecol Evol* 2009, **24:**121-122.
18. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D: **Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates.** *Mol Biol Evol* 2007, **24:**1702-1713.
19. Patron NJ, Inagaki Y, Keeling PJ: **Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages.** *Curr Biol* 2007, **17:**887-891.
20. Baldauf S, Roger A, Wenk-Siefert I, Doolittle W: **A kingdom-level phylogeny of eukaryotes based on combined protein data.** *Science* 2000, **290:**972-977.
21. Not F, Valentin K, Romari K, Lovejoy C, Massana R, Töbe K, Vaulot D, Medlin LK: **Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes.** *Science* 2007, **315:**253-255.
22. Rodríguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF: **Toward resolving the eukaryotic tree: the phylogenetic positions of Jakobids and Cercozoans.** *Curr Biol* 2007, **17:**1420-1425.
23. Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J: **Phylogenomics reshuffles the eukaryotic supergroups.** *Plos One* 2007, **2:**e790.
24. Burki F, Shalchian-Tabrizi K, Pawlowski J: **Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes.** *Biol Rev* 2008, **4:**366-369.
25. Hampl V, Hug L, Leigh J, Dacks J, Lang BF, Simpson AG, Roger A: **Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups".** *Proc Natl Acad Sci USA* 2009, **106:**3859-3864.
26. Huang J, Gogarten JP: **Concerted gene recruitment in early plant evolution.** *Genome Biol* 2008, **9:**R109.
27. Rice D, Palmer J: **An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters.** *BMC Biol* 2006, **4:**31.
28. Harper JT, Keeling PJ: **Nucleus-encoded, plastid-targeted Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids.** *Mol Biol Evol* 2003, **20:**1730-1735.
29. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D: **A molecular timeline for the origin of photosynthetic eukaryotes.** *Mol Biol Evol* 2004, **21:**809-818.

30. Bachvaroff TR, Sanchez Puerta MV, Delwiche CF: **Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages.** *Mol Biol Evol* 2005, **22:**1772-1782.

31. Khan H, Parks N, Kozera C, Curtis B, Parsons B, Bowman S, Archibald J: **Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny.** *Mol Biol Evol* 2007, **24:**1832-1842.

32. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ: **The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts.** *Mol Biol Evol* 2007, **24:**54-62.

33. Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF: **Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids.** *Mol Phylogen Evol* 2007, **44:**885-897.

34. Iida K, Takishita K, Ohshima K, Inagaki Y: **Assessing the monophyly of chlorophyll-c containing plastids by multi-gene phylogenies under the unlinked model conditions.** *Mol Phylogen Evol* 2007, **45:**227-238.

35. Wang Y, Joly S, Morse D: **Phylogeny of dinoflagellate plastid genes recently transferred to the nucleus supports a common ancestry with red algal plastid genes.** *J Mol Evol* 2008, **66:**175-184.

36. Hagopian JC, Reis M, Kitajima JP, Bhattacharya D, de Oliveira MC: **Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. liui provides insights into the evolution of rhodoplasts and their relationship to other plastids.** *J Mol Evol* 2004, **59:**464-477.

37. Reith M, Munholland J: **Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome.** *Plant Mol Biol Rep* 1995, **13:**333-335.

38. Glöckner G, Rosenthal A, Valentin K: **The structure and gene repertoire of an ancient red algal plastid genome.** *J Mol Evol* 2000, **51:**382-390.

39. Ohta N, Matsuzaki M, Misumi O, Miyagishima S-y, Nozaki H, Tanaka K, Shin-i T, Kohara Y, Kuroiwa T: **Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*.** *DNA Res* 2003, **10:**67-77.

40. Waller RF, McFadden GI: **The Apicoplast: a review of the derived plastid of Apicomplexan parasites.** *Curr Issues Mol Biol* 2005, **7:**57-80.

41. Zhang Z, Green BR, Cavalier-Smith T: **Single gene circles in dinoflagellate chloroplast genomes.** *Nature* 1999, **400:**155-159.

42. Koumandou VL, Nisbet RER, Barbrook AC, Howe CJ: **Dinoflagellate chloroplasts - where have all the genes gone?** *Trends Gen* 2004, **20:**261-267.

43. Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF: **The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes.** *DNA Res* 2005, **12:**151-156.

44. Kowallik K, Stoebe B, Schaffran I, Kroth-Pancic P, Freier U: **The chloroplast genome of a chlorophyll a+c- containing alga, *Odontella sinensis*.** *Plant Mol Biol Rep* 1995, **13:**336-342.

45. Oudot-Le Secq M-P, Grimwood J, Shapiro H, Armbrust EV, Bowler C, Green BR: **Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage.** *Mol Genet Genom* 2007, **277:**427-439.

46. Cattolico RA, Jacobs MA, Zhou Y, Chang J, Duplessis M, Lybrand T, McKay J, Ong HC, Sims E, Rocap G: **Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains.** *BMC Gen* 2008, **9:**211.

47. Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, Moustafa A, Manhart JR: **Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug *Elysia chlorotica*.** *Proc Natl Acad Sci USA* 2008, **105:**17867-17871.

48. Phillips N, Burrowes R, Rousseau F, de Reviers B, Saunders GW: **Resolving evolutionary relationships among the brown algae using chloroplast and nuclear genes.** *J Phycol* 2008, **44:**394-405.

49. Rodríguez-Ezpeleta N, Philippe H, Brinkmann H, Becker B, Melkonian M: **Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of *Mesostigma* in the Streptophyta.** *Mol Biol Evol* 2007, **24:**723-731.

50. Simon D, Fewer D, Friedl T, Bhattacharya D: **Phylogeny and self-splicing ability of the plastid tRNA-Leu group I intron.** *J Mol Evol* 2003, **57:**710-720.

51. Riisberg I, Orr RJS, Kluge R, Shalchian-Tabrizi K, Bowers HA, Patil V, Edvardsen B, Jakobsen KS: **Seven gene phylogeny of Heterokonts.** *Protist* 2009, **160:**191-204.

52. Kai A, Yoshii Y, Nakayama T, Inouye I: **Aurearenophyceae classis nova, a new class of Heterokontophyta based on a new marine unicellular alga *Aurearena cruciata* gen. et sp. nov. inhabiting sandy beaches.** *Protist* 2008, **159:**435-457.

53. Bowler C, Allen AE, Badger JH, (co-authors), *et al.*: **The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456:**239-244.

54. Kalanon M, McFadden GI: **The chloroplast protein translocation complexes of *Chlamydomonas reinhardtii*: a bioinformatic comparison of Toc and Tic components in plants, green algae and red algae.** *Genetics* 2008, **179:**95-112.

55. Gross J, Bhattacharya D: **Revaluating the evolution of the Toc and Tic protein translocons.** *Trends in Plant Science* 2009, **14:**13-20.

56. van Dooren GG, Tomova C, Agrawal S, Humbel BM, Striepen B: **Toxoplasma gondii Tic20 is essential for apicoplast protein import.** *Proc Natl Acad Sci USA* 2008, **105:**13574-13579.

57. Moore RB, Obornik M, Janouskovec J, Chrudimsky T, Vancova M, Green DH, Wright SW, Davies NW, Bolch CJS, Heimann K, *et al.*: **A photosynthetic alveolate closely related to apicomplexan parasites.** *Nature* 2008, **451:**959-963.

58. Slamovits CH, Keeling PJ: **Plastid-derived genes in the nonphotosynthetic alveolate *Oxyrrhis marina*.** *Mol Biol Evol* 2008, **25:**1297-1306.

59. Reyes-Prieto A, Moustafa A, Bhattacharya D: **Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic.** *Curr Biol* 2008, **18:**956-962.

60. Rokas A, Carroll S: **More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy.** *Mol Biol Evol* 2005, **22:**1337-1344.

61. Sanchez-Puerta MV, Delwiche CF: **A hypothesis for plastid evolution in chromalveolates.** *J Phycol* 2008, **44:**1097-1107.

62. Moustafa A, Reyes-Prieto A, Bhattacharya D: **Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions.** *Plos One* 2008, **3:**e2205.

63. Li S, Nosenko T, Hackett JD, Bhattacharya D: **Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates.** *Mol Biol Evol* 2006, **23:**663-674.

64. Bhattacharya D, Nosenko T: **Endosymbiotic and horizontal gene transfer in Chromalveolates.** *J Phycol* 2008, **44:**7-10.

65. Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier U, Grossman AR, Bhattacharya D, Lohr M: **Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis.** *Mol Biol Evol* 2008, **25:**2653-2667.

66. Peters A, Scornet D, Ratin M, Charrier B, Monnier A, Merrien Y, Corre E, Coelho S, Cock J: **Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga *Ectocarpus siliculosus*.** *Dev* 2008, **135:**1503-1512.

67. Apt KE, Clendennen SK, Powers DA, Grossman AR: **The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*.** *Mol Gen Genet* 1995, **246:**455-464.

68. Doyle JJ, Doyle JL: **Natural interspecific hybridization in eastern North American Claytonia.** *Am J Bot* 1988, **75:**1238-1246.

69. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, *et al.*: **GenDB-an open source genome annotation system for prokaryote genomes.** *Nucleic Acids Res* 2003, **31:**2187-2195.

70. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.

71. Lagesen K, Hallin P, Rødland E, Staerfeldt H, Rognes T, Ussery D: **RNAmmer: consistent and rapid annotation of ribosomal RNA genes.** *Nucleic Acids Res* 2007, **35:**3100-3108.

72. Laslett D, Canback B: **ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.** *Nucleic Acids Res* 2004, **32:**11-16.

73. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792-1797.

74. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17:**540-552.
75. Strimmer K, von Haeseler A: **Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies.** *Mol Biol Evol* 1996, **13:**964-969.
76. Guindon S, Gascuel O: **A simple, fast and accurate algotithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52:**696-704.
77. Adachi J, Waddell P, Martin W, Hasegawa M: **Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA.** *J Mol Evol* 2000, **50:**348-358.
78. Jones D, Taylor W, Thornton J: **The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8:**275-282.
79. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5:**164-166.
80. Lartillot N, Philippe H: **A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process.** *Mol Biol Evol* 2004, **21:**1095-1109.
81. Lartillot N, Brinkmann H, Philippe H: **Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogenious model.** *BMC Evol Biol* 2007, **7:**S4.
82. Shimodaira H, Hasegawa M: **CONSEL: for assessing the confidence of phylogenetic tree selection.** *Bioinformatics* 2001, **17:**1246-1247.