

RESEARCH ARTICLE

Open Access



Comparative plastid genomics of Synurophyceae: inverted repeat dynamics and gene content variation

Jong Im Kim¹, Hyunmoon Shin¹, Pavel Škaloud², Jaehee Jung³, Hwan Su Yoon⁴, John M. Archibald^{5*} and Woongghi Shin^{1*}

Abstract

Background: The Synurophyceae is one of most important photosynthetic stramenopile algal lineages in freshwater ecosystems. They are characterized by siliceous scales covering the cell or colony surface and possess plastids of red-algal secondary or tertiary endosymbiotic origin. Despite their ecological and evolutionary significance, the relationships amongst extant Synurophyceae are unclear, as is their relationship to most other stramenopiles.

Results: Here we report a comparative analysis of plastid genomes sequenced from five representative synurophycean algae. Most of these plastid genomes are highly conserved with respect to genome structure and coding capacity, with the exception of gene re-arrangements and partial duplications at the boundary of the inverted repeat and single-copy regions. Several lineage-specific gene loss/gain events and intron insertions were detected (e.g., *cemA*, *dnaB*, *syfB*, and *trnL*).

Conclusions: Unexpectedly, the *cemA* gene of Synurophyceae shows a strong relationship with sequences from members of the green-algal lineage, suggesting the occurrence of a lateral gene transfer event. Using a molecular clock approach based on silica fossil record data, we infer the timing of genome re-arrangement and gene gain/loss events in the plastid genomes of Synurophyceae.

Keywords: Algae, Stramenopiles, Synurophyceae, Plastid genomes, Lateral gene transfer

Introduction

The Synurophyceae, a class of photosynthetic stramenopile (or heterokont) algae, is a morphologically diverse lineage with plastids derived from red algae via secondary or tertiary endosymbiosis. They are motile organisms with two parallel emergent flagella, one or two plastids, a cell coat in which the siliceous scales cover the entire cell, and lack of an eyespot [1, 2]. Synurophyceae are presently assigned to one of three genera: *Mallomonas*, *Neotessella* and *Synura*. Members of the genus *Synura* are colonial flagellates characterized by cells having two visible and unequal flagella, two plastids, and an external covering of siliceous scales. Members of the single celled genus

Mallomonas have silica scales and bristles, while the colonial genus *Neotessella* is characterized by an oval-shaped scale structure and a single scale case that surrounds the whole colony.

The presence of four membranes surrounding Synurophycean plastids provides direct evidence for the hypothesis that their plastids are derived by eukaryote-eukaryote endosymbiosis, a process that is thought to have given rise to photosynthesis in several other protist lineages (e.g., cryptophytes, haptophytes, euglenoids, chlorarachniophytes; [3–5]). The outermost membrane of synurophycean plastids is continuous with the endoplasmic reticulum (referred to as the chloroplast ER; CER), but unlike the red-algal derived plastids of other ‘chromists’ such as haptophytes and cryptophytes, a linkage between this membrane and the outer nuclear envelope is either totally lacking or marginal [1]. The silica deposition vesicles (SDVs), in which siliceous scales

* Correspondence: John.Archibald@Dal.Ca; shinw@cnu.ac.kr

⁵Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

¹Department of Biology, Chungnam National University, Daejeon 34134, South Korea

Full list of author information is available at the end of the article



form, are produced from the CER on the outer side of the plastid. The mature scales are brought to the cell surface via Golgi body vesicles and placed in position alongside pre-existing scales [6, 7].

Molecular sequence datasets that include combinations of nuclear, plastid, and mitochondrial genes have provided insight into the branching order amongst the three recognized synurophyte genera and their phylogenetic relationship to other algae [8–10]. Recently, the plastid genome of the chrysophycean alga *Ochromonas* sp. CCMP1393 was reported [5]. The *Ochromonas* species genome was ‘conservative’ in possessing a large single copy region (LSC), a very short single copy region (SSC), and two inverted repeats (IR) with 15 functional protein-coding genes and ribosomal RNA operons. Recent phylogenomic studies of photosynthetic stramenopiles based on plastid genome data have focused mainly on Bacillariophyceae (diatoms) and Phaeophyceae (brown algae) with one or a few species from six additional classes—Bolidophyceae, Chrysophyceae, Eustigmatophyceae, Pelagophyceae, Raphidophyceae, Xanthophyceae—as well as plastid-bearing alveolates [5, 11]. While photosynthetic stramenopiles consist of at least 15 classes, phylogenetic relationships amongst them, including Synurophyceae, are still unresolved. Investigation of organellar genome structure and coding capacity from new protists has the potential to complement phylogenetic analyses by reinforcing observed relationships and helping to resolve phylogenetic issues.

Lateral gene transfer (LGT; also known as horizontal gene transfer) is the movement of genetic material from one species into the genome of an unrelated species. LGT provides a potentially important source of genetic variation in mitochondrial and plastid genomes, many of which display intron gain/loss and the presence of chimeric genes created by gene conversion. LGT appears to be rare in algal plastid genome but a few probable cases of bacterial derived genes have nevertheless been documented. These include the leucine biosynthesis (*leuC/D*) operon and RuBisCO genes of red algae [4, 12], the *dnaX* gene and group II introns in cryptophytes [13–16], the *rpl36* gene in the haptophyte and cryptophyte plastid genomes [13] and the *ebo* operon in Eustigmatophyceae [17]. In addition, the plastid genomes of several angiosperms show evidence for LGT of one or more genes from mitochondria to plastids [18–20]. Although in most cases the underlying mechanisms are not known, taken together these examples strongly suggest that LGT from bacteria to organelles and from one organelle to another can occur.

Here, we present five complete plastid genome sequences from the three morphologically distinct genera of synurophycean algae: *Mallomonas*, *Neotessella*, and *Synura*. To better understand the relationships among synurophytes as well as the broader insight into organellar

genome evolution of plastids in stramenopiles, we performed comparative and phylogenetic analyses of these genomes in the context of publicly available plastid genome sequence data, including that of the chrysophycean alga, *Ochromonas* sp. CCMP1393. Our results reveal highly conserved features of plastid genomes amongst the synurophyceans. We also uncovered several examples of gene loss/gain, duplication and gene rearrangement. Our results provides important insights into the evolutionary history of organelle genomes via lateral gene transfer (LGT) from green-algal lineages into the Synurophyceae, as well as divergence time estimates using molecular clock approaches based on silica fossil records. Collectively, our data contribute to a better understanding of the evolutionary history of the Synurophyceae.

Results

General features of Synurophyceae plastid genomes

Five new plastid genomes (ptDNA) were sequenced from the synurophycean genera *Mallomonas*, *Neotessella*, and *Synura* (Table 1, Fig. 1). The structure and coding capacity of these ptDNAs were then compared to the published genome of the related chrysophycean alga, *Ochromonas* sp. CCMP1393 [5]. The plastid genome sizes of the Synurophyceae ranged from ~130 kbp (*S. sphagnicola*) to ~147 kbp (*M. splendens*) and the overall GC content ranged from 37.5 to 42.4%. The overall organization of the five synurophytes and *Ochromonas* sp. CCMP1393 was found to be conserved: they each contain a large single copy region, a very short single copy region, and two inverted repeats. The plastid genomes of Synurophyceae share a core set of 134 functional protein-coding genes including genes in the IR regions (Table 2). The plastid genome IR sequence length of the Synurophyceae and *Ochromonas* sp. CCMP1393 ranged from 22.5 kbp to 31.6 kbp with 15 functional protein-coding genes (*ccs1*, *ccsA*, *chlL*, *petI*, *petM*, *petN*, *psaC*, *psaM*, *psbA*, *psbC*, *psbD*, *rpl21*, *rpl27*, *rpl34*, and *secA*), 3 rRNAs and 5 tRNAs. Introns and a pseudogene were found in the tRNA genes. The *trnR*^{UCU} in *S. petersenii*, *S. uvella*, and *M. splendens* and *trnE*^{UUC} in *N. volvocina* is present as a pseudogene; it has a low hidden Markov model score (HMM score = 0) and a secondary structure-only score (2Str Score < 40) predicted by tRNAscan-SE. *Synura* and *Mallomonas* have an intron in *trnL*^{UAA}, whereas *N. volvocina* has introns in *trnP*^{UGG} and *trnS*^{GCU} (Table 3, Fig. 2). The five newly determined plastid genomes showed a high degree of structural conservation relative to the representative species, *Synura petersenii* (Fig. 1 and Additional file 1: Figure S2). Gene order among Synurophyceae and *Ochromonas* ptDNAs was conserved, with the exception of *N. volvocina*, which has two inversions that differ from other synurophycean species (Fig. 1).

Table 1 Characteristics of Synurophyceae plastid genomes analyzed in this study

General characteristics	<i>Synura petersenii</i> S114.C7, CZ	<i>Synura sphagnicola</i> FBCC200022	<i>Synura uvella</i> FBCC200023	<i>Mallomonas splendens</i> CCMP1872	<i>Neotessella volvocina</i> CCMP1871	<i>Ochromonas</i> species CCMP1393
Key characteristics for genus classification	colonized cells covered with silica scale on each cell	colonized cells covered with silica scale on each cell	colonized cells covered with silica scale on each cell	single cell covered with silica scale	colonized cells covered with silica scale on colony	naked single cell
Size (bp)	133,059	129,699	133,257	146,918	130,705	126,746
Inverted repeat (IR)	23,151	22,505	23,691	31,611	24,064	22,906
Small single-copy region	1,135	1,191	2,939	711	2,432	805
Large single-copy region	85,622	83,498	82,936	82,985	80,145	80,129
G+C (%)	37.89	38.76	38.19	42.39	37.54	30.9
Total gene (include RNAs)	182	181	189	187	186	183
No. of protein-coding genes	144	144	151	150	149	144
tRNAs	34	33	34	33	33	33
rRNA operons	2	2	2	2	2	2
Introns	<i>trnL</i>	<i>trnL</i>	<i>trnL</i>	<i>trnL</i>	<i>trnP</i> , <i>trnS</i>	-
Unknown ORFs	8	3	9	11	9	7
pseudogene	-	-	<i>trnR</i>	<i>trnR</i>	<i>trnE</i>	-
partial copied gene	-	<i>dnaK</i>	-	-	<i>dnaK</i>	-
specific encoded genes	<i>dnaB/ cemA</i>	<i>dnaB/ cemA</i>	<i>dnaB/ cemA</i>	<i>dnaB/ cemA</i>	<i>cemA</i>	-
missing gene	<i>syfB</i>	<i>syfB</i>	-	-	-	-
GenBank accession	MH795128	MH795129	MH795130	MH795131	MH795132	KJ877675

Gene arrangements of IR and SSC regions

The synurophycean plastid genomes exhibit different gene order patterns in six distinct regions (Fig. 1a-e). First, three different gene order patterns were found at the boundaries between single-copy and repetitive regions (Fig. 1a). The most common pattern is that shared by *Ochromonas* sp. CCMP1393, *N. volvocina* and *M. splendens*, in which a particular block of genes (*trnS-psaD-trnM-ycf36-petM-petN*) lies between the *dnaK* and *chlI* genes at the IRA/LSC junction (Fig. 1a). Within this syntenic block, *S. uvella* and *S. sphagnicola* have one open reading frame (ORF) between *ycf36* and *petM*, whereas *S. petersenii* is distinct in the loss of *trnS-psaD-trnM-ycf36* (Fig. 1a). Second, four different gene order patterns were observed at the IR/SSC junction (Fig. 1b). Gene content in this region of the genome is conserved, but four genes (*psaC*, *ccsA*, *psbA*, *trnN*) are dynamically rearranged in the IR regions of each species. The gene rearrangements in the plastid genome of *S. uvella* are distinct from other synurophyceans. The IRB/LSC junctions of the synurophyte plastid genomes also showed three different gene order

patterns (Fig. 1d). The first pattern, shared by *Ochromonas* sp. CCMP1393, *M. splendens* and *S. petersenii*, involves the *ilvB* gene linked to the *trnS-psaD-trnM* genes. It probably represents the ancestral gene order in the synurophycean algae. The second pattern is seen only in *S. uvella*, one in which there is an ORF in the plus orientation between *ycf36* and *petM*. The third pattern is shared by *N. volvocina* and *S. sphagnicola*; it involves the presence of a partial *dnaK* gene (the duplicated region starts at the 440th amino acid) between the *ilvB* and *trnS-psaD-trnM* genes.

The small single copy (SSC) region is exceptionally short, ranging from 711 bp to 2939 bp, and includes only two protein-coding genes: *ycf54* and *psbY* (Fig. 1c). The SSC flanking regions have slightly different patterns of gene duplication and location of the genes in each species. The *psbY* gene is located to the left side of *ycf54* gene in *S. petersenii* and *Ochromonas* sp. CCMP1393, while it is located on the right side in *M. splendens*. *Synura sphagnicola* has duplicated *psbY* genes on both sides of *ycf54*, but the *ycf54* gene is absent in plastid genome of *N. volvocina*.

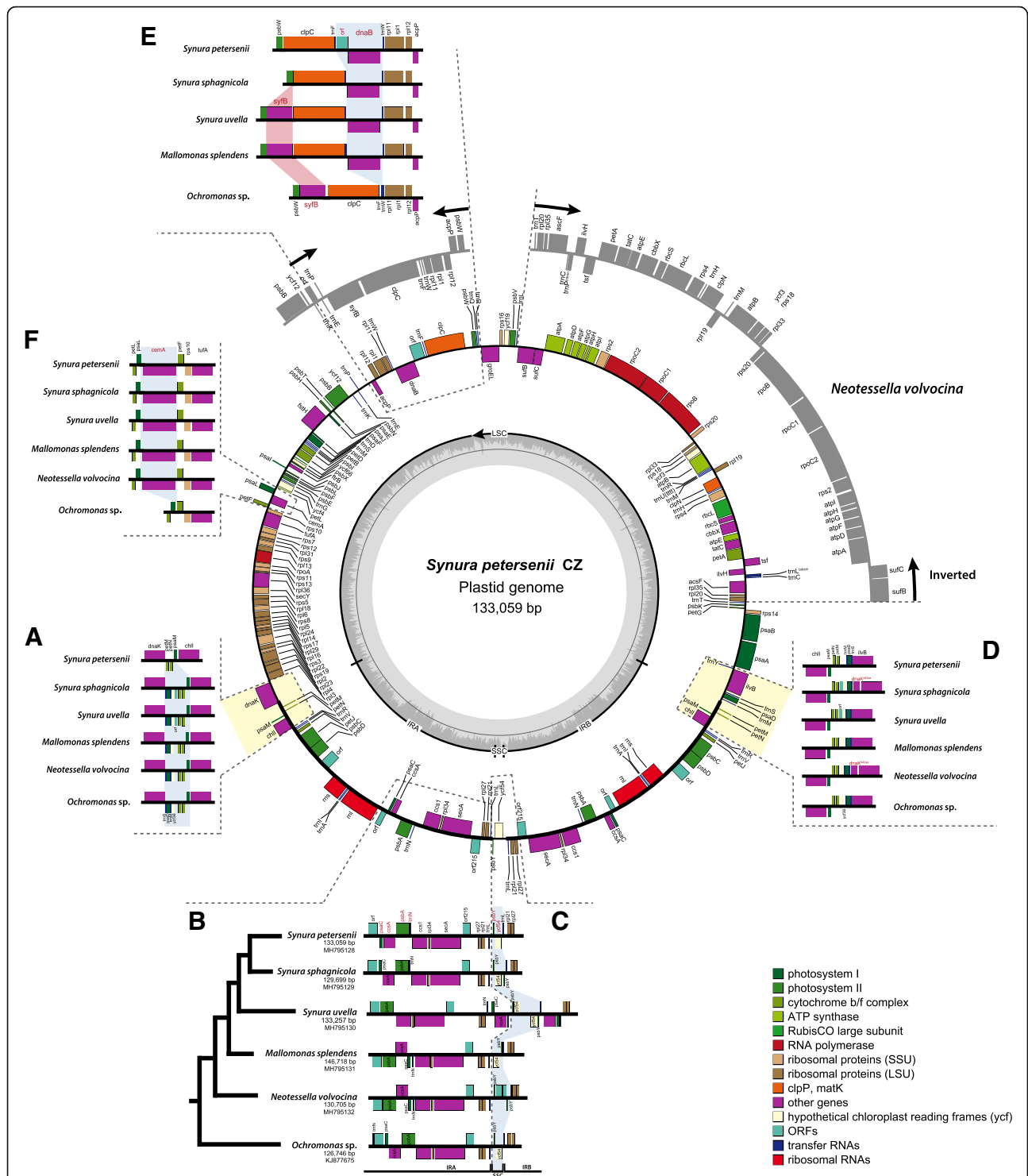


Fig. 1 Circular map of the plastid genome of *Synura petersenii*. The gene content and arrangement of the synurophycean plastid genomes examined herein are identical, with the exception of the six syntenic regions shown as (a-f). Regions of the *N. volvocina* genome are shown in gray. The protein coding genes, rRNA and tRNA genes are labeled inside or outside of the circle. The genes are color-coded according to the functional categories in the index

Table 2 List of genes in the synurophycean plastid genome

Classification	Genes								
Genetic systems									
Maintenance	dnaB ^{S,M}								
RNA polymerase	rpoA	rpoB	rpoC1	rpoC2					
Translation	tsf	tufA	syfB ^{Su,M,N}						
Protein quality control	clpC	clpN	dnaK	ftsH	groEL				
Transport									
Transport	cemA ^{S,M, N}	secA ²	secY	sufB	sufC	tatC			
ATP synthesis									
ATP synthase	atpA	atpB	atpD	atpE	atpF	atpG	atpH	atpI	
Ribosomal proteins									
Large subunit	rpl1	rpl2	rpl3	rpl4	rpl5	rpl6	rpl11	rpl12	
	rpl13	rpl14	rpl16	rpl18	rpl19	rpl20	rpl21 ²	rpl22	
	rpl23	rpl24	rpl27 ²	rpl29	rpl31	rpl33	rpl34 ²	rpl35	
	rpl36								
Small subunit	rps2	rps3	rps4	rps5	rps7	rps8	rps9	rps10	
	rps11	rps12	rps13	rps14	rps16	rps17	rps18	rps19	
	rps20								
Metabolism									
Carbohydrates	rbcl	rbcS							
Lipids	acpP								
Nucleotides									
Amino acids	ilvB	ilvH							
Cofactors	ascF	chlI ²							
Photosystems									
Photosystems I	psaA	psaB	psaC ²	psaD ²	psaE	psaF	psaI	psaJ	
	psaL	psaM ²	ycf3	ycf4					
Photosystems II	psbA ²	psbB	psbC ²	psbD ²	psbE	psbF	psbH	psbI	
	psbJ	psbK	psbL	psbN	psbT	psbV	psbW	psbX	
	psbY ²	ycf12							
Cytochrome complex	cbbX	ccs1 ²	ccsA ²	petA	petB	petD	petF	petG	
	petJ ²	petL	petM ²	petN ²					
Redox system	ftrB								
Unknown									
Conserved ORFs	orf215 ²	ycf19	ycf36 ²	ycf54	ycf66				

Note. ²present as repeated genes, ^Mpresent in *Mallomonas splendens*, ^Npresent in *Neotessella volovocina*, ^Spresent in three species of genus *Synura*, ^{Su}present only in *Synura uvella* in plastid genome

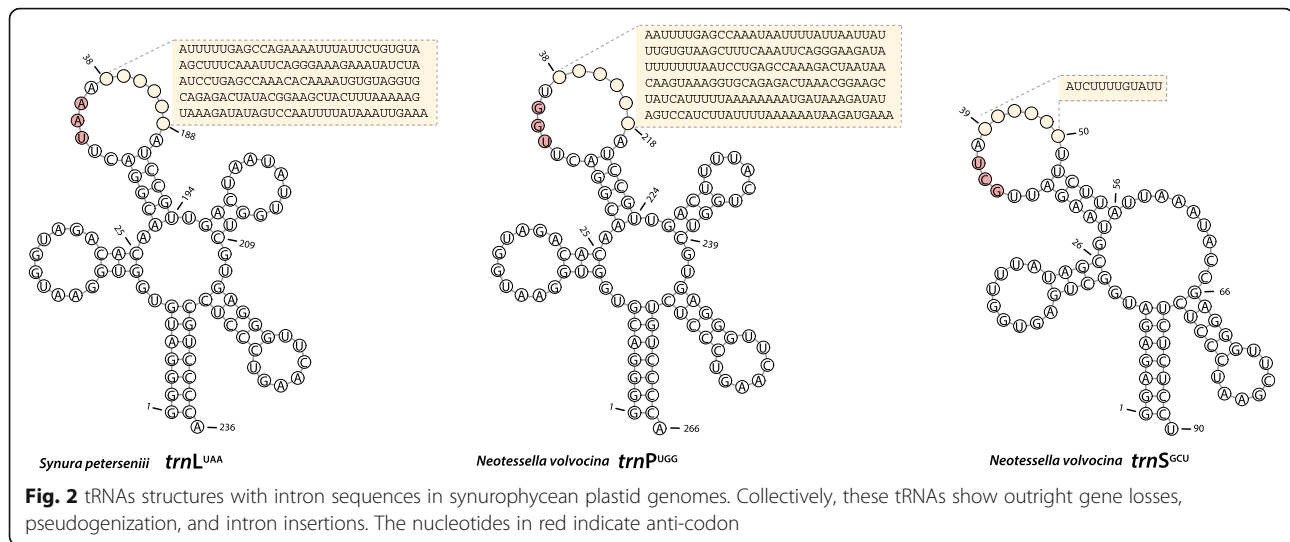
Synura uvella has duplicated *psbY-ycf54* genes in extended IR regions.

Discussion

Expansions and contractions of IR region

The expansions and contractions of IR region have occurred frequently during the evolutionary history of Synurophyceae, leading not to changes in gene content, but to gene rearrangements and gene duplications in

our results. Such events can alter gene order through inversion, expansion/contraction of the IR, gene duplication/loss, or transposition. IR boundary shifts are a common phenomenon, which is thought to be caused by inversions or recombination between repeated sequences resulting in gene order changes in plastid genomes [21, 22]. Contractions, expansions and small-scale changes in IR and SSC regions appear to be common in diatoms and green algae, leading to dynamic gene rearrangements



and changes in gene content [23–26]. Rearrangements at the IR boundary is likely one of the factors contributing to the extensive genome rearrangements in the Synurophyceae as well.

Lineage specific gene gain and loss

Previous work has shown that in red-algal derived secondary plastids, most of the lineage-specific plastid genes show complex distribution patterns suggesting independent losses across a broad range of phylogenetic depths [16]. Although the plastid genomes of Synurophyceae and *Ochromonas* sp. CCMP1393 studied herein are generally highly conserved in structure and gene content, three genes were identified as being lineage specific: *dnaB*, *syfB*, and *cemA* (Fig. 1e-f). To understand the evolutionary distribution and phylogenetic relationships of these genes among eukaryotes, we performed phylogenetic analyses of homologs obtained from the plastid genomes of major photosynthetic eukaryotic groups with their cyanobacterial homologs.

The *dnaB* gene encodes a DNA helicase that is involved in organelle division [27, 28] and is found in the plastid genome of cryptophytes, some dinoflagellates (i.e., those with diatom-derived plastids), and specific subgroups of stramenopiles and rhodophytes (Additional file 2: Figure S3; [16]). Interestingly, *dnaB* in Synurophyceae was only found in the genera *Synura* and *Mallomonas* (Fig. 1e and Additional file 2: Figure S3). The synurophyte sequences branch at the base of the photosynthetic stramenopiles in the algal *dnaB* gene tree (Additional file 2: Figure S3). In stramenopiles, *dnaB* is present only in Bacillariophyceae (except *Synedra acus*), Phaeophyceae, Raphidophyceae, *Triparma*, Synurophyceae and Xanthophyceae, but absent in Pelagophyceae, Eustigmatophyceae, and Chrysophyceae [16]. Of particular note, the ‘dinotoms’ *Durinskia baltica* and *Krptoperidinium foliaceum*, which

are dinoflagellates harboring a diatom endosymbiont, contain a *dnaB* gene in their plastid genomes [29]. The cryptophytes, which also harbor a complex red-algal derived plastid, branch with the main red algal lineage including *Galdieria sulphuraria*. The *dnaB* gene appears to have been present in the plastid genome of the red algal common ancestor; if it was present in the common ancestor of all primary plastid-bearing algae, it was lost in green algae and glaucophytes, and independently in many complex red-algal derived plastid genomes (Additional file 2: Figure S3).

The *syfB* gene encodes the β subunit of phenylalanyl-tRNA synthetase [30]. While the *syfB* and *syfH* genes are retained in primary plastid-bearing organisms, the *syfH* gene is absent in complex red-algal plastid genomes. Furthermore, *syfB* has been lost in almost all red algal-derived plastid genomes, with the exception of the diatom *Triparma* and Chrysophyceae in stramenopiles [5, 11, 24, 31], as well as rhodophytes [32]. In the Synurophyceae and Chrysophyceae, *syfB* remains in most species, but is not present in *S. petersenii* and *S. sphagnicola*, suggesting that it was lost recently in a common ancestor shared by these two species (Fig. 1e, and Additional file 3: Figure S4). The dinotoms *Durinski baltica* and *Krptoperidinium foliaceum* also have a *syfB* gene in their plastid genomes. It is likely that the *syfB* gene has a similar history as *dnaB*, i.e., being ancestrally present and lost independently in specific groups.

Lateral gene transfer from green algae into photosynthetic Stramenopiles

The *cemA* gene encodes a chloroplast inner membrane protein [33]; it is conserved in almost all green algae, liverworts, land plants and rhodophytes, but is not found in glaucophytes [34, 35]. In the green algal order Bryopsidales, the

Table 3 tRNAs present in Synurophyceae plastid genomes.

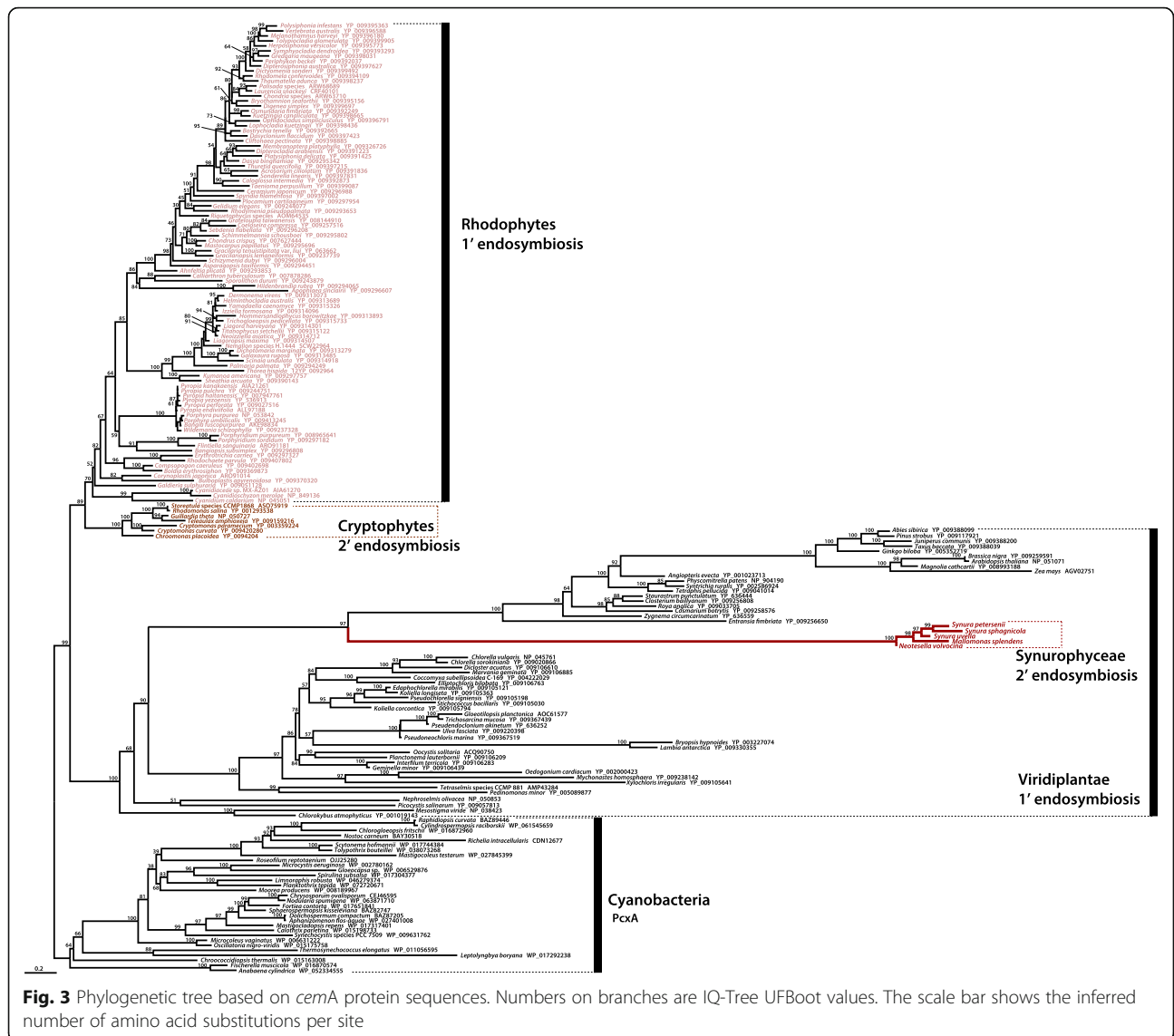
	<i>Synura petersenii</i> S114.C7	<i>Synura sphagnicola</i> FBCC200022	<i>Synura uvella</i> FBCC200023	<i>Mallomonas splendens</i> CCMP1782	<i>Neotessella volvocina</i> CCMP1781	<i>Ochromonas</i> sp. CCMP1393
trnA(UGC) ²	2	2	2	2	2	2
trnC(GCA)	1	1	1	1	1	1
trnD(GUC)	1	1	1	1	1	1
trnE(UUC)	1	1	1	-	1 ^ψ	1
trnF(GAA)	1	1	1	1	1	1
trnG(UCC)	1	1	1	1	1	1
trnH(GUG)	1	1	1	1	1	1
trnI(GAU) ²	2	2	2	2	2	2
trnK(UUU)	1	1	1	1	1	1
trnL(UAA)	1 ¹	1 ¹	1 ¹	1 ¹	-	1
trnL(CAA)	1	-	1	1	-	-
trnL(UAG) ²	2	2	2	2	2	2
trnM(CAU) ²	3	4	4	4	4	3
trnM(CAU)	-	-	-	-	-	1
trnN(GUU) ²	2	2	2	2	2	2
trnP(UGG)	1	1	1	1	1+1 ¹	1
trnQ(UUG)	1	1	1	1	1	1
trnR(ACG)	1	1	1	1	1	1
trnR(CCG)	1	-	-	-	-	-
trnR(UCU) ²	2 ^ψ	2	2 ^ψ	2 ^ψ	2	2
trnS(GCU)	-	1	1	1	1 ¹	1
trnS(UGA) ²	2	2	2	2	2	2
trnT(UGU)	1	1	1	1	1	1
trnU(UUUU)	1	-	-	-	-	-
trnV(UAC) ²	2	2	2	2	2	2
trnW(CCA)	1	1	1	1	1	1
trnY(GUA)	1	1	1	1	1	1
Total	34	33	34	33	33	33

Note. ²present as repeated tRNA genes, ¹present as intron encoded tRNA genes, ^ψpresent as tRNA pseudogene in plastid genome

cemA gene appears to have been lost in two clades [36]. Among the groups of algae with red-algal derived complex plastids, the *cemA* gene is thus far only found in cryptophytes and Synurophyceae (Figs. 1f and 3), and our phylogenetic analyses suggest that the *cemA* homologs in these two lineages have different origins: the cryptophyte protein shows a strong phylogenetic relationship with rhodophytes, whereas the synurophycean protein groups within Viridiplantae (Fig. 3). At face value, this is consistent with the hypothesis that the synurophycean plastid *cemA* gene was derived from a member of the green lineage through LGT. It is, however, not possible to be more specific than this; the synurophycean homologs are extremely divergent and branch sister to long-branching *cemA* proteins in streptophytes, rather than those of chlorophytes.

Evolution of the *trnL*^{UAA} intron

The *trnL*^{UAA} group I intron of algae is thought to have been acquired from the ancestral cyanobacterial endosymbiont that gave rise to the plastid. The existence of related introns in the *trnL*^{UAA} gene has been reported in most green algal plastid genomes, as well as some stramenopiles [37]. A phylogenetic analysis of the intron suggests that it was present in the cyanobacterial ancestor of the three primary plastid-bearing lineages i.e., Rhodophyta, Viridiplantae, and Glaucophyta (Additional file 4: Figure S5). The *trnL*^{UAA} group I intron is absent from red, cryptophyte and haptophyte algae, and found only in some stramenopiles, i.e., Phaeophyceae, Phaeothamniophyceae, Xanthophyceae, and Eustigmatophyceae ([5, 38, 39], this study). Given the high degree of intron sequence similarity between these four



subgroups of stramenopiles, the *trnL^{UAA}* gene is probably derived from the same ancestral archaeplastidal sequence. One notable feature is the presence of a predicted group I intron in all *trnL^{UAA}* and *trnD^{UGG}* (*Neotessella volvocina*) genes in the synurophycean plastid genome (Fig. 2). The group I intron sequences are more closely related to homologs in the green algal lineage and chlorarachniophytes (which have a green algal secondary plastid), rather than other stramenopiles (Additional file 4: Figure S5). However, it is not possible to infer the origin *trnL^{UAA}* group I intron with certainty, as the structure of the *trnL^{UAA}* group I intron is generally very poorly supported.

Evolutionary history of plastid genomes in Synurophycean algae

Phylogenomic analysis using 91 genes of plastid genome data showed a monophyletic, strongly supported (MLB =

100%) synurophycean clade; internal relationships among the three genera were also well resolved (Fig. 4). In our maximum likelihood (ML) phylogeny, the genus *Neotessella* is the deepest branching synurophycean lineage, with *Mallomonas* and *Synura* splitting off thereafter. Furthermore, our phylogenomic investigations show that the Synurophyceae form a strongly supported sister relationship with the chrysophytes (Fig. 4), which is congruent with previous multigene phylogenetic studies [9, 10, 40–42]. Interestingly, the chromerids *V. brassicaformis* and *C. velia* form a strongly supported sister relationship with Eustigmatophyceae in Fig. 4. This topology is consistent with recent studies suggesting that the eustigmatophytes could be the source of the chromerid plastid [5, 43].

Synurophycean algae are characterized by the presence of distinctive siliceous scales that produce a highly organized covering around the cell [8, 44]. The

fossil record is rich in Synurophyceae containing silicious scales and cysts, which are resting stages produced by species of the Synurophyceae as well as Chrysophyceae [8, 45]. According to Siver et al. [8], the Synurophyceae originated in the Jurassic, approximately 157 million years ago (Ma), with the clade containing *Mallomonas* and *Synura* diverging during the Early Cretaceous at 130 Ma.

Using molecular clock data and our plastid genome phylogenies, we inferred the timing of gene gains, losses, and rearrangements in the plastid genomes of the synurophycean lineage. *N. volvocina* is predicted to have lost the *dnaB* gene in the plastid genome between ~156 Ma

and the present, after the major synurophycean lineages diverged (Fig. 5. ①). The *syfB* gene loss may have occurred during the Early Cretaceous at 130 Ma, after the divergence of colonial *Synura* and unicellular *Mallomonas*; this is inferred because the gene is found in the plastid genomes of *Mallomonas*, *Synura*, *Neotessela* and ochrophytes (Fig. 5. ②). The *cemA* gene, hypothesized to have been derived from a member of the Viridiplantae by LGT, and the intron of the *trnL^{UAA}* (*Synura* and *Mallomonas*) or *trnD^{UGG}* (*N. volvocina*) genes appear to have been acquired during the Jurassic approximately 156 Ma, before the divergence of the *Mallomonas*, *Neotessela*, and *Synura* genera (Fig. 5. ③-④). The *psaD* gene, located near

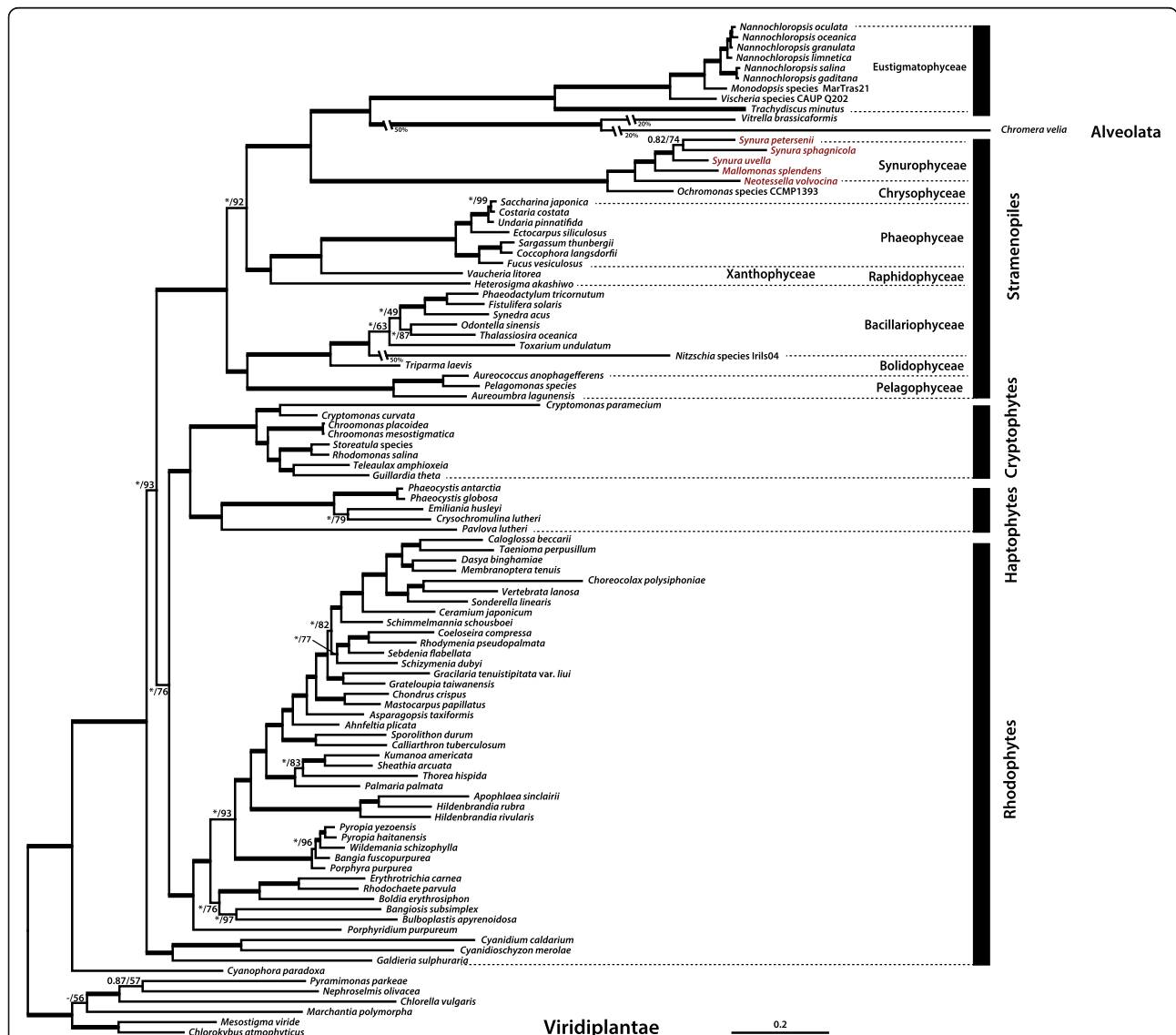
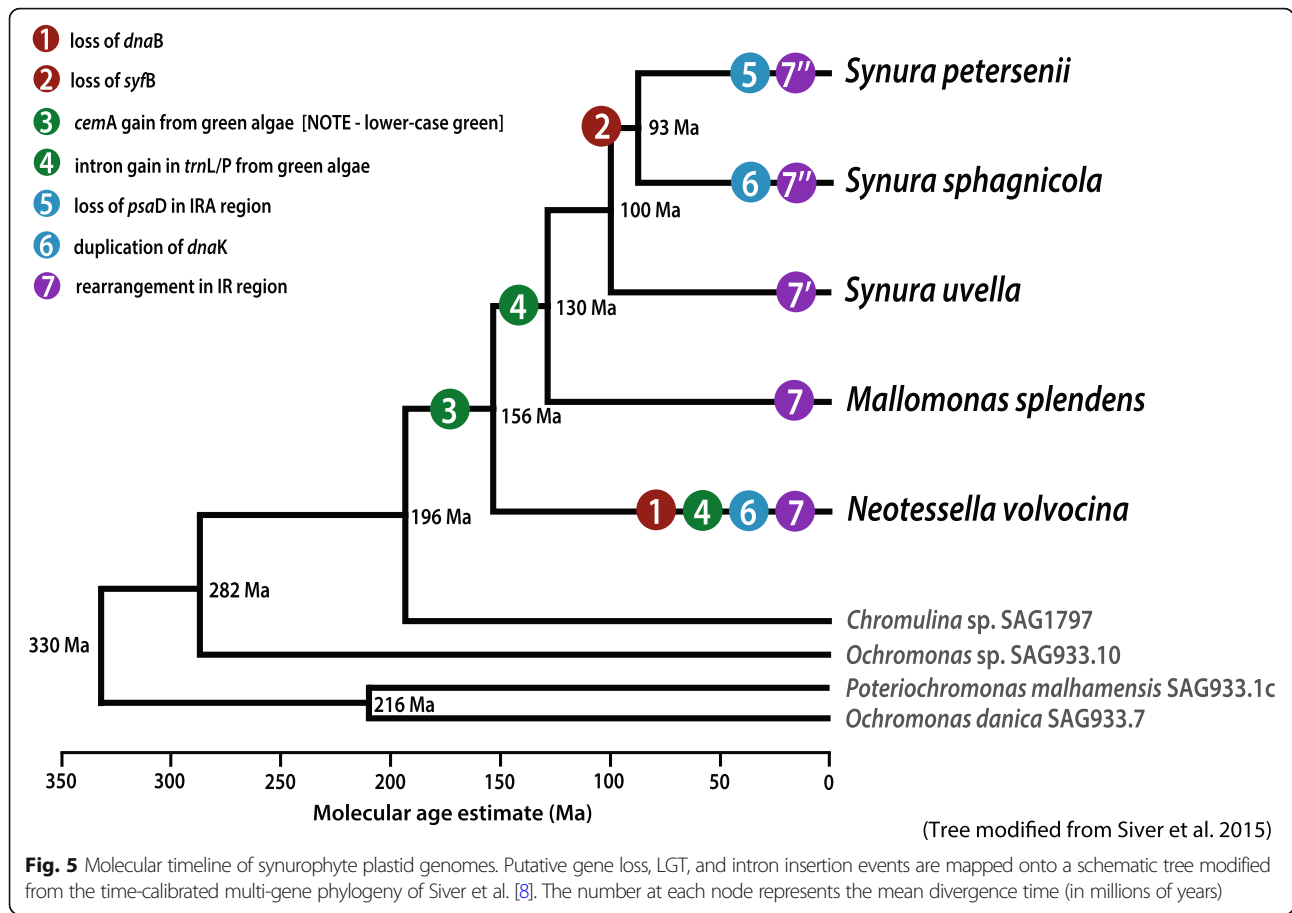


Fig. 4 Phylogenetic tree of synurophyte plastids. This tree was constructed using a dataset of 91 concatenated proteins (18,250 amino acids). The numbers on each node represent ultrafast bootstrap approximation (UFboot) values (left) calculated using IQ-Tree and posterior probabilities (right). The bold branch indicates strong supported values (ML = 100 / PP = 1.00). The scale bar indicates the number of substitutions/site



dnaK in the LSC/IRA junction, appears to have been lost ~ 93 Ma before present because the gene is found in all genera except *S. petersenii* (Fig. 5. ⑤). The partial duplication of the *dnaK* gene near the *ilvB* gene in the SSC/IRB junction might have been duplicated or truncated recently given that the genes are present only in *S. uvella* and *N. volvocina* (Fig. 5. ⑤-⑥). The gene rearrangements in the IR regions and duplications/translocations in the SSC regions are the result of species-specific events (Fig. 5. ⑦).

Conclusions

We have sequenced five synurophyte plastid genomes from morphologically distinct genera: the colonial genus *Synura*, whose individual cells are covered with silica scales; the single-celled genus *Mallomonas* covered with silica scales and bristles; and the colonial genus *Neotessella*, whose entire colony is covered with a single, large silica case. The overall organization of the plastid genome shows a high degree of conservation among the five Synurophyceae and *Ochromonas* sp. CCMP1393, but *N. volvocina* has two inversions relative to the other synurophycean species. The IR and SSC boundaries are particularly variable from species to species. Instances of lineage specific gene loss/gain and intron insertions were

also detected (e.g., *cemA*, *dnaB*, *syfB*, and *trnL*). The *dnaB* and *syfB* genes appear to have been lost independently in different synurophyceans. Both the *trnL* intron sequences and *cemA* gene of Synurophyceae appear most closely related to their counterparts in green algae, suggestive of LGT. However, their sequences are divergent and should thus be interpreted with caution. All things considered, the extent to which LGT has contributed to the plastid genomes of Synurophyceae and other algae remains to be seen. Multi-gene phylogenetic analyses show that Synurophyceae group together with Chrysophyceae among the stramenopiles. Combined with molecular clock data, our phylogenetic tree allows us to infer the timing of gene gains, losses, duplications and rearrangements in the plastid genome of the synurophycean lineage.

Materials and methods

Cultures and sequencing

Cultures of *Neotessella volvocina* CCMP1871 and *Mallomonas splendens* CCMP1872 were obtained from the Culture Collection of the National Center for Marine Algae and Microbiota (NCMA). Three species of *Synura* were collected from natural habitats: *Synura petersenii* from Sweden (36° 30' N, 126° 47' E), *Synura sphagnicola*

from Cheongyang, Korea (36° 30' N, 126° 47' E), and *Synura uvella* from Gahang, Korea (35° 30' N, 128° 23' E). *S. petersenii* strain S114.C7 has been deposited as CAUP B713 in the Culture Collection of Algae of Charles University in Prague, Czech Republic. The strains *S. sphagnicola* FBCC200022 and *S. uvella* FBCC200023 are available from the Freshwater Bioresources Culture Collection at the Nakdong-gang National Institute of Biological Resources Korea, respectively. All freshwater cultures were grown in DY-V medium [46] with distilled water and were maintained at 17 °C under conditions of a 14:10 light:dark cycle with 30 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ from cool white fluorescent tubes. All cultures were derived from a single-cell isolate for unialgal cultivation. Total genomic DNAs were extracted using the QIAGEN DNEasy Blood Mini Kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. Next-generation sequencing was carried out using the MiSeq (Illumina, San Diego, CA, USA). The amplified DNA was fragmented and tagged using the NexteraXT protocol (Illumina), indexed, size selected, and pooled for sequencing using the small amplicon targeted resequencing run, which performs paired end 2 × 300 bp sequencing reads using the MiSeq Reagent Kit v3 (Illumina), according to the manufacturer's recommendations.

Assembly and annotation of plastid genomes

Sequence data were trimmed, assembled using the SPAdes 3.7 assembler (<http://bioinf.spbau.ru/spades>), and mapped to the assembled contigs. The contigs were deemed to be of plastid genome origin as follows: (1) BLAST searches against the entire assembly using commonly known plastid genes as queries resulted in hits to these contigs [47, 48] and (2) the predicted genome sizes were similar to the previously published 127 kbp plastid genome of *Ochromonas* sp. CCMP1393 (KJ877675). For each genome we verified the sequence and structure of both inverted repeat positions and SSC regions with specific primers using standard Sanger sequencing (Additional file 5: Figure S1). Annotation of protein coding genes, rRNA genes, and tRNA genes were identified using data from all previously sequenced synurophycean plastid genomes according to the methods described in Kim et al. [16]. Genome sequences were deposited to the NCBI GenBank database under the accession numbers shown in Table 1.

Phylogenetic analysis

Phylogenetic analyses were carried out on amino acid sequence datasets created by combining 91 protein coding genes from 99 plastid genomes (Additional file 6: Table S1). The sequences of six Viridiplantae and one glaucophyte species were used as outgroup taxa for rooting purposes.

The datasets were aligned and concatenated (18,250 amino acid sequences) using MacGDE2.6 [49].

ML phylogenetic analyses of individual protein alignments and concatenated alignments were conducted using IQ-TREE Ver. 1.5.2 [50] with 1000 bootstrap replications. The best evolutionary model for each tree was automatically selected using the $-m$ LG + I + G option incorporated in IQ-TREE. RAxML version 8.0.0 [51] with the general time-reversible plus gamma (GTR + GAMMA) model was used for nucleotide data of the intron within *trnL*^{UAA}. The model parameters with gamma correction values and the proportion of invariable sites in the combined dataset were obtained automatically by the program. ML bootstrap support values (MLB) were calculated using 1000 replicates with the same substitution model. Bayesian analyses were carried out using MrBayes 3.2.6 [52] with two simultaneous runs (nrns = 2). Four Metropolis-coupled Markov chain Monte Carlo (MC³) chains ran for 2 × 10⁶ generations, sampling every 1000 generations. The burn-in point was determined by examining the trace files using Tracer v.1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>). This analysis was repeated twice independently, and both analyses resulted in the same tree. The trees were visualized using the FigTree v.1.4.2 program, available at <http://tree.bio.ed.ac.uk/software/figtree/>.

Additional files

Additional file 1: Figure S2. Overview of Synurophyceae plastid genomes. Linearized maps of five novel complete plastid genomes are compared with *Ochromonas* sp. CCMP1393. The color coded syntenic blocks are shown above each genome, and the gene maps are shown below each genome. The syntenic blocks above the horizontal line are on the same strand, and those below the line are on the opposite strand. The horizontal bars inside the syntenic blocks show sequence conservation. The block boundaries correspond to the sites where inversion events occurred. In the gene maps, the genes above the horizontal line are transcribed from left to right, and those below the horizontal line are transcribed from right to left. The rRNA operons are shown in red. (PDF 1586 kb)

Additional file 2: Figure S3. Phylogenetic tree based on *dnaB*. Numbers on branches are IQ-Tree UFBoot values. The scale bar shows the inferred number of amino acid substitutions per site. (PDF 283 kb)

Additional file 3: Figure S4. Phylogenetic tree based on *syfB*. Numbers on branches are IQ-Tree UFBoot values. The scale bar shows the inferred number of amino acid substitutions per site. (PDF 227 kb)

Additional file 4: Figure S5. Phylogenetic tree based on intron sequence within *trnL*^{UAA}. Numbers on branches are RAxML bootstrap values. The scale bar indicates the number of substitutions/site. (PDF 178 kb)

Additional file 5: Figure S1. The positions of gap filling with primer information. (PDF 489 kb)

Additional file 6: Table S1. The concatenated dataset of protein sequences used to infer the reference phylogenetic tree. (XLSX 53 kb)

Abbreviations

CER: Chloroplast ER; IR: Inverted repeats; LGT: Lateral gene transfer; LSC: Large single copy region; Ma: Million years ago; ML: Maximum likelihood;

MLB: Maximum likelihood bootstrap; ORF: Open reading frame; PheRS: Phenylalanyl-tRNA synthetase; ptDNA: Plastid genomes; SDVs: Silica deposition vesicles; SSC: Short single copy region

Acknowledgements

The authors thank Gangman Yi for bioinformatic assistance and Sergio A. Muñoz-Gómez for assistance of phylogenetic analyses.

Funding

This research was supported by the National Research Foundation (NRF) of Korea (NRF-2015R1D1A1A01057899 and 2018R1D1A1B07050727) to JJK; NRF (NRF-2016R1C1B1007929) to JJ; the Czech Science Foundation (grant number 17-13254S) to PŠ; the Collaborative Genome Program (20140428) funded by the Ministry of Oceans and Fisheries, Korea and NRF (NRF-2017R1A2B3001923) to HSY; the Natural Sciences and Engineering Research Council of Canada to JMA; and NRF (2015R1A2A2A01003192 and 2015M1A5A1041808) to WS. These funding organizations were not involved in the design of the study, in the collection, analysis and interpretation of the data, or in writing the manuscript.

Availability of data and materials

All algal cultures are available from the National Center for Marine Algae and Microbiota (NCMA), and the Freshwater Bioresources Culture Collection at the Nakdong-Gang National Institute of Biological Resources, Korea. The plastid genome datasets in this project has been deposited in the NCBI GenBank database under the accession numbers; MH795128, MH795129, MH795130, MH795131, MH795132.

Authors' contributions

JJK, HSY, WS, JMA conceived and designed the experiments: JJK, HS, JJ, PŠ performed the experiments and analyzed the data: JJK, HSY, WS, JMA wrote the manuscript: All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Biology, Chungnam National University, Daejeon 34134, South Korea. ²Department of Botany, Faculty of Science, Charles University, Benátská 2, CZ-12800 Prague 2, Czech Republic. ³Department of General Education, Hongik University, Seoul 04066, South Korea. ⁴Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, South Korea. ⁵Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada.

Received: 18 January 2018 Accepted: 4 December 2018

Published online: 11 January 2019

References

- Andersen RA. Synurophyceae classis Nov., a new class of algae. *Am J Bot.* 1987;74:337–53.
- Siver PA. *The biology of Mallomonas: morphology, taxonomy and ecology.* Dordrecht: Kluwer Academic Publishers; 1991.
- Bhattacharya D, Medlin L. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J Phycol.* 1995;31:489–98.
- Delwiche CF, Palmer JD. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol Biol Evol.* 1996;13:873–82.
- Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, et al. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep.* 2015;5:10134.
- Schnepf E, Deichgräber G, Koch W. Über das Vorkommen und den Bau gestielter "Hüllen" bei *Ochromonas malhamensis* Pringsheim und *Ochromonas sociabilis* nom. prov. Pringsheim. *Arch Mikrobiol.* 1968;63:15–25.
- Kristiansen K, Škaloud P. Chrysophyta. In: *Handbook of the Protists*; 2017. p. 331–66.
- Siver PA, Jo BY, Kim JI, Shin W, Lott AM, Wolfe AP. Assessing the evolutionary history of the class Synurophyceae (Heterokonta) using molecular, morphometric, and paleobiological approaches. *Am J Bot.* 2015;102:921–41.
- Jo BY, Kim JI, Škaloud P, Siver PA, Shin W. Multigene phylogeny of *Synura* (Synurophyceae) and description of four new species based on morphological and DNA evidence. *Eur J Phycol.* 2016;51:413–30.
- Andersen RA, Graf L, Malakhov Y, Yoon HS. Rediscovery of the *Ochromonas* type species *Ochromonas triangulata* (Chrysophyceae) from its type locality (Lake Veysove, Donetst region, Ukraine). *Phycologia.* 2017;56:591–604.
- Ruck EC, Nakov T, Jansen RK, Theriot EC, Alverson AJ. Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. *Genome Biol Evol.* 2014;6:644–54.
- Janoušková J, Liu SL, Martone PT, Carré W, Leblanc C, Collén J, Keeling PJ. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One.* 2013;8:e59001.
- Rice DW, Palmer JD. 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.
- Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM. 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;24:1832–42.
- Khan H, Archibald JM. Lateral transfer of introns in the cryptophyte plastid genome. *Nucleic Acids Res.* 2008;36:3043–53.
- Kim JI, Moore CE, Archibald JM, Bhattacharya D, Yi G, Yoon HS, Shin W. Evolutionary dynamics of cryptophyte plastid genomes. *Genome Biol Evol.* 2017;9:1859–72.
- Yurchenko T, Ševčíková T, Strnad H, Butenko A, Eliáš M. The plastid genome of some eustigmatophyte algae harbours a bacteria-derived six-gene cluster for biosynthesis of a novel secondary metabolite. *Open Biol.* 2016;6:160249.
- Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biol Evol.* 2013;5:1872–85.
- Ma PF, Zhang YX, Guo ZH, Li DZ. Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. *Sci Rep.* 2015;5:11608.
- Gandini CL, Sanchez-Puerta MV. Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. *Sci Rep.* 2017;7:43402.
- Goulding S, Olmstead R, Morden C, Wolfe K. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet.* 1996;252:195–206.
- Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol.* 2008;8:36.
- Jansen R, Ruhlman T. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. *Genomics of chloroplasts and mitochondria.* Netherlands: Springer; 2012. p. 103–26.
- Sabir JSM, Yu M, Ashworth MP, Baeshen NA, Baeshen MN, Bahieldin A, Theriot EC, Jansen RK. Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *PLoS One.* 2014;9:e107854.
- Turmel M, Otis C, Lemieux C. Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. *Genome Biol Evol.* 2015;7:2062–82.
- Turmel M, Otis C, Lemieux C. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophyceae green algae. *Sci Rep.* 2017;7:994.
- Douglas SE, Penny SL. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved syntenic groups confirm its common ancestry with red algae. *J Mol Evol.* 1999;48:236–44.
- Ohta N, Matsuxaki M, Misumi O, Miyagishima S, Nozaki H, et al. Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Res.* 2003;10:67–77.
- Imanian B, Pombert J-F, Keeling PJ. The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS One.* 2010;5:e10711.

30. Saforo M, Moor N, Lavrik O. Phenylalanyl-tRNA synthetases. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000. <http://www.ncbi.nlm.nih.gov/books/NBK6321/>.
31. Tajima N, Saitoh K, Sato S, Maruyama F, Ichinomiya M, et al. Sequencing and analysis of the complete organellar genomes of *Parmales*, a closely related group to Bacillariophyta (diatoms). *Curr Genet*. 2016;62:887–96.
32. Lee JM, Cho CH, Park SI, Choi JW, et al. Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol*. 2016;14:75.
33. Sasaki Y, Sekiguchi K, Nagano Y, Matsuno R. Chloroplast envelope protein encoded by chloroplast genome. *FEBS Lett*. 1993;316:93–8.
34. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*. 1986;322:572–4.
35. Grzebyk D, Schofield O, Vetriani C, Falkowski PG. The mesozoic radiation of eukaryotic algae: The portable plastid hypothesis. *J Phycol*. 2003;39:259–67.
36. Cremen MCM, Leliaert F, Marcelino VR, Verbruggen H. Large diversity of nonstandard genes and dynamic evolution of chloroplast genomes in siphonous green algae (Bryopsidales, Chlorophyta). *Genome Biol Evol*. 2018;10:1048–61.
37. Besendahl A, Qiu YL, Lee J, Palmer JD, Bhattacharya D. The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. *Curr Genet*. 2000;37:12–23.
38. Simon D, Fewer D, Friedl T, Bhattacharya B. Phylogeny and self-splicing ability of the plastid tRNA-Leu group I intron. *J Mol Evol*. 2003;57:710–20.
39. Le Corquillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, et al. Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. *BMC Evol Biol*. 2009;9:253.
40. Brown JW, Sorhannus U. A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *PLoS One*. 2010;5:e12759.
41. Yang EC, Boo GH, Kim HJ, Cho SM, Boo SM, Andersen RA, Yoon HS. Supermatrix data highlight the phylogenetic relationships of photosynthetic stramenopiles. *Protist*. 2012;163:217–31.
42. Derelle R, López-García P, Timpano H, Moreira D. A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Mol Biol Evol*. 2016;33:2890–8.
43. Füssy Z, Obornik M. Chromerids and their plastids. *Adv Bot Res*. 2017;84:187–218.
44. Kristiansen, J. Golden algae: a Biology of chrysophytes. A.R.G. Gantner Verlag, Königstein; 2005. p. 167.
45. Nicholls KH, Gerrath JF. The taxonomy of *Synura* (Chrysophyceae) in Ontario with special reference to taste and odour in water supplies. *Can J Bot*. 1985;63:1482–93.
46. Andersen RA, Berges JA, Harrison PJ, Watanabe MM, Appendix A. Recipes for freshwater and seawater media. In: Andersen RA, editor. *Algal culturing techniques*. Burlington, Massachusetts: Elsevier Academic Press; 2005. p. 429–538.
47. Jung J, Kim JI, Jeong Y-S, Yi G. A robust method for finding the automated best matched genes based on grouping similar fragments of large-scale references for genome assembly. *Symmetry*. 2017;9:192.
48. Jung J, Kim JI, Jeong Y-S, Yi G. AGORA: organellar genome annotation from the amino acid and nucleotide references. *Bioinformatics*. 2018;34:2661–3.
49. Smith SW, Overbeek R, Woese CR, Gilbert W, Gillevet PM. The genetic data environment: an expandable GUI for multiple sequence analysis. *Comput Appl Biosci*. 1994;10:671–5.
50. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
51. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
52. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

