



Comparative Plastid Genomics of Non-Photosynthetic Chrysophytes: Genome Reduction and Compaction

Jong Im Kim¹, Minseok Jeong¹, John M. Archibald² and Woongghi Shin^{1*}

¹ Department of Biology, Chungnam National University, Daejeon, South Korea, ² Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada

Spumella-like heterotrophic chrysophytes are important eukaryotic microorganisms that feed on bacteria in aquatic and soil environments. They are characterized by their lack of pigmentation, naked cell surface, and extremely small size. Although *Spumella*-like chrysophytes have lost their photosynthetic ability, they still possess a leucoplast and retain a plastid genome. We have sequenced the plastid genomes of three non-photosynthetic chrysophytes, *Spumella* sp. Baeckdong012018B8, *Pedospumella* sp. Jangsampo120217C5 and *Poteriospumella lacustris* Yongseonkyo072317C3, and compared them to the previously sequenced plastid genome of “*Spumella*” sp. NIES-1846 and photosynthetic chrysophytes. We found the plastid genomes of *Spumella*-like flagellates to be generally conserved with respect to genome structure and housekeeping gene content. We nevertheless also observed lineage-specific gene rearrangements and duplication of partial gene fragments at the boundary of the inverted repeat and single copy regions. Most gene losses correspond to genes for proteins involved in photosynthesis and carbon fixation, except in the case of *petF*. The newly sequenced plastid genomes range from ~55.7 kbp to ~62.9 kbp in size and share a core set of 45 protein-coding genes, 3 rRNAs, and 32 to 34 tRNAs. Our results provide insight into the evolutionary history of organelle genomes *via* genome reduction and gene loss related to loss of photosynthesis in chrysophyte evolution.

Keywords: genome reduction, leucoplast, non-photosynthesis, chrysophytes, plastid genome

OPEN ACCESS

Edited by:

Richard Dorrell,
École Normale Supérieure, France

Reviewed by:

Marek Eliáš,
University of Ostrava, Czechia
Anna Kamkowska,
University of Warsaw, Poland

*Correspondence:

Woongghi Shin
shinw@cnu.ac.kr

Specialty section:

This article was submitted to
Marine and Freshwater Plants,
a section of the journal
Frontiers in Plant Science

Received: 15 June 2020

Accepted: 24 August 2020

Published: 10 September 2020

Citation:

Kim JI, Jeong M, Archibald JM and
Shin W (2020) Comparative Plastid
Genomics of Non-Photosynthetic
Chrysophytes: Genome Reduction
and Compaction.
Front. Plant Sci. 11:572703.
doi: 10.3389/fpls.2020.572703

INTRODUCTION

Chrysophytes are a large algal group with diverse morphologies and various nutritional modes, including phototrophy, mixotrophy, and heterotrophy. Among them, mixotrophs and heterotrophs are ecologically important eukaryotes that feed on bacteria and other eukaryotes inhabiting freshwater, brackish, and marine environments. The phototrophic chrysophytes have golden-brown plastids with chlorophylls *a* and *c*. The chrysophyte plastid is derived from a red alga through secondary (i.e., eukaryote-eukaryote) endosymbiosis (Kim and Archibald, 2009). As stramenopiles, chrysophytes are closely related to the Synchronophyceae and Eustigmatophyceae. They are also phylogenetically grouped together with 16 classes of plastid-containing stramenopiles, including brown algae and diatoms (Han et al., 2019; Kim et al., 2019).

Non-photosynthetic chrysophytes appear to have evolved several times independently within the chrysophytes (Grossmann et al., 2016). *Spumella*-like species, which are non-photosynthetic nanoflagellates, are characterized by colorless cells with naked cell surfaces, a non-colonial lifestyle, heterokont flagella, and endogenous silicified stomatocysts. Recent taxonomic studies based on nuclear small subunit (SSU) rDNA sequence data suggest that *Spumella*-like flagellates are not monophyletic and have lost the ability to photosynthesize many times in different lineages, resulting in seven new *Spumella*-like genera (Boenigk and Stadler, 2004; Findenig et al., 2010; Grossmann et al., 2016). Molecular phylogenies clearly show that these independently evolved colorless chrysophyte species are related to photosynthetic lineages, including phototrophic and mixotrophic species, within chrysophytes (Grossmann et al., 2016; Andersen et al., 2017; Dorrell et al., 2019). Interestingly, non-pigmented, vestigial plastids (leucoplasts) have been reported in *Poteriospumella lacustris*, *Cornospumella fuschlensis*, *Spumella vulgaris*, and *Pedospumella encystans*, suggesting that phagotrophic *Spumella*-like flagellates are derived from ancestral phototrophic or mixotrophic species with photosynthetic ability (Grossmann et al., 2016).

The loss of photosynthesis in autotrophic organisms containing primarily and secondarily derived plastids has occurred multiple times during the course of eukaryotic evolution (Kim and Archibald, 2009; Hadariová et al., 2018; Maciszewski and Karnkowska, 2019; Sibbald and Archibald, 2020). Although the loss of most or all photosynthesis-related genes leads to a reduction in plastid genome size and complexity, most non-photosynthetic taxa retain a plastid genome. Such examples include the green algal pathogens *Helicosporidium* sp. (de Koning and Keeling, 2004) and *Prototheca* spp. (Knauf and Hachtel, 2002; Borza et al., 2005; Suzuki et al., 2018), the parasitic plant *Epifagus virginiana* (de Pamphilis and Palmer, 1990), members of the free-living green algal genus *Polytoma* (Nedelcu, 2001; Figueroa-Martinez et al., 2017), and the euglenoid alga *Euglena longa* (Gockel and Hachtel, 2000). Additionally, genome-containing non-photosynthetic plastids of red-algal secondary endosymbiotic origin are diverse in nature, and can be found in the malaria parasite *Plasmodium falciparum* (Wilson et al., 1996; Waller and McFadden, 2005), the cryptophyte *Cryptomonas paramecium* (Donaher et al., 2009; Tanifuji et al., 2020), the colorless diatom *Nitzschia* species. (Kamikawa et al., 2018), and the chrysophyte “*Spumella*” sp. NIES-1846 (Dorrell et al., 2019).

Many comparative genomic and phylogenomic studies have used plastid genome data to shed light on the evolution of photosynthetic stramenopiles (Ruck et al., 2014; Ševčíková et al., 2015; Han et al., 2019; Kim et al., 2019). In non-photosynthetic species of chrysophytes, the ultrastructures of remnant plastids have been reported (Grossmann et al., 2016), and a recent genome and transcriptome-based phylogenomic study suggested parallel reductive evolution of non-photosynthetic chrysophytes, including the genus *Paraphysomonas*, which appears to have undergone complete loss of the plastid genome (Dorrell et al., 2019). But there are still many unanswered questions about the loss of photosynthesis in the chrysophytes, questions that can be addressed by investigating the structure and coding capacity of plastid genomes in diverse chrysophyte species.

To that end, we sequenced three plastid genomes from the following representative non-photosynthetic chrysophyte algal genera: *Spumella*, *Pedospumella*, and *Poteriospumella*. We carried out a detailed comparative analysis of their plastid genome structures and gene contents relative to each other and to published photosynthetic chrysophyte plastid genome sequences for *Ochromonas* sp. CCMP1393, synurlean algae, and the non-photosynthetic “*Spumella*” sp. NIES-1846 (Ševčíková et al., 2015; Dorrell et al., 2019; Kim et al., 2019). Our results contribute to the growing body of knowledge relating to how gene content and genome structure changes in response to the loss of photosynthesis in chrysophytes and other algae.

MATERIALS AND METHODS

Cultures, DNA Isolation and Whole Genome Sequencing

Cultures of three *Spumella*-like species were established by a single cell isolation technique with glass pipetting from natural habitats: *Spumella* sp. Baeckdong012018B8 from freshwater, Jindo, Korea (34° 22' 39" N, 126° 11' 00" E); *Pedospumella* sp. Jangsampo120217C5 from the seashore, Taean, Korea (36° 25' 24" N, 126° 21' 33" E); and *Poteriospumella lacustris* Yongseonkyo072317C3 from freshwater, Gochang, Korea (35° 31' 38" N, 126° 35' 41" E). The three strains of *Spumella*-like flagellates are available from the Protist Culture Collection, Department of Biological Sciences, Chungnam National University, Korea. All cultures were grown in AF-6 medium (Andersen et al., 2005) with rice grains and distilled water for the freshwater strains or distilled seawater for the marine strain (*Pedospumella* sp. Jangsampo120217C5) and were maintained at 20°C under 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 single-cell isolates for unialgal cultivation. Total genomic DNA was extracted from exponentially growing cell cultures using the QIAGEN DNEasy Blood Mini Kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. A paired-end library was prepared using the NexteraXT protocol (Illumina) according to the manufacturer's protocol. Whole genome sequencing was performed using the Illumina MiSeq platform to generate paired-end 2 × 300 bp sequencing reads. More than 2 Gb of raw data were generated for each species.

Genome Assembly and Annotation of Plastid Genomes

Sequence data were trimmed (i.e., base = 80 bp, error threshold = 0.05, n ambiguities = 2) using Trimmomatic 0.36 (Bolger et al., 2014) for *de novo* assembly with the default option (automatic bubble size, minimum contig length = 1,000 bp). The trimmed reads were assembled using the SPAdes 3.7 assembler using *k*-mer size 127 (Bankevich et al., 2012) and then mapped using Geneious Pro 10.2.2 (Kearse et al., 2012) to assemble the contigs (similarity = 95%, length fraction = 75%), excluding contigs < 1,000 bp. The assembled 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 significant hits for these contigs (Jung et al., 2017) and (2) the predicted gene contents were similar to the previously published 160 kbp plastid genome of the chrysophytes *Ochromonas* sp. (KJ877675) and *Synura petersenii* (MH795128).

To aid in gene annotation, we created a database of protein-coding, rRNA, and tRNA genes using data from previously sequenced chrysophyte plastid genomes. Preliminary annotation of protein coding genes was performed using AGORA (Jung et al., 2018) and GeneMarkS (Besemer et al., 2001). The final annotation file was checked in Geneious Pro 10.2.2 (Kearse et al., 2012) using the ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) with genetic code 11 (Bacterial, Archaeal, and Plant plastid code). The predicted open reading frames (ORFs) were checked manually, and the corresponding ORFs (and predicted functional domains) in the genome sequence were annotated.

The tRNA genes were identified using tRNAscan-SE version 1.21 (Lowe and Chan, 2016) with the default settings using the “Mito/Chloroplast” model. To help identify rRNA gene sequences, a set of known plastid rRNA sequences from the public database was used as a query sequence to search our new genome data using BLASTn. Physical maps were visualized with OrganellarGenomeDRAW 1.3.1 (Greiner et al., 2019). For structural and synteny comparisons, the genomes were aligned using GeneCo (Jung et al., 2019) with default settings.

Phylogenomic and Phylogenetic Analyses

Phylogenomic analysis was carried out on amino acid sequence datasets created by combining 40 protein coding genes (8,267 amino acids) from 98 plastid genomes of stramenopiles. The plastid genome of *Hydrurus foetidus* was analyzed and annotated based on publicly available data on scaffold UYFQ01001012 (Bråte et al., 2019). The sequences of six haptophyte species were used as outgroup taxa for rooting purposes. For the phylogeny of the nuclear-encoded SSU rDNA gene (1,591 nucleotides) from 177 chrysophyte taxa and *Leukarachnion* sp., *Nannochloropsis limnetica*, and *Synchroma grande* were used as outgroup taxa. The dataset was aligned using MUSCLE 8.0 in the program MacGDE2.6 (Smith et al., 1994; Edgar, 2004).

Bayesian analyses were run using MrBayes 3.2.7 (Ronquist et al., 2012) with a random starting tree, two simultaneous runs (nruns = 2) and four Metropolis-coupled Markov chain Monte Carlo (MC3) algorithms for 2×10^7 generations, with one tree retained every 1,000 generations. The burn-in point was identified graphically by tracking the likelihood values using TRACER v. 1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>). ML phylogenetic analyses of individual protein alignments and concatenated alignments were conducted using IQ-TREE v.1.5.2 (Nguyen et al., 2015) with 1,000 bootstrap replicates. ML phylogenetic analysis of the nuclear-encoded SSU rDNA gene was performed using RAXML 8.1.20 (Stamatakis, 2014) with the general time reversible plus Gamma (GTR + GAMMA) model. We used 1,000 independent tree inferences using the `-#` option of the program to identify the best tree. The best evolutionary model for each tree was selected using the posterior mean site frequency (PMSF) model (the LG+F+G tree

as the guide tree) incorporated in IQ-TREE. Trees were visualized using FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

RESULTS AND DISCUSSION

General Features of Non-Photosynthetic Chrysophyte Plastid Genomes

Three new plastid genomes were sequenced from the non-photosynthetic chrysophyte genera *Spumella*, *Pedospumella* and *Poteriospumella* (Table 1). The structures and coding capacities of these genomes were compared to those of the published genomes of the related photosynthetic chrysophyte *Ochromonas* sp. CCMP1393 and synuralean species (Ševčíková et al., 2015; Kim et al., 2019) and the non-photosynthetic “*Spumella*” sp. NIES-1846 (Dorrell et al., 2019). Most obvious is the fact that the plastid genomes of non-photosynthetic chrysophyte *Spumella*-like flagellates have lost all of the photosynthesis-related genes found in their photosynthetic relatives (Figure 1, see below). The plastid genome sizes of the four non-photosynthetic chrysophyte taxa range from ~53.2 kbp (“*Spumella*” sp. NIES-1846) to ~62.9 kbp (*Poteriospumella lacustris*), and the overall GC content ranged from 21.5% (“*Spumella*” sp. NIES-1846) to 38.9% (*Spumella* sp. Baekdong012018B8) (Table 1). These taxa share a core set of 45 protein-coding genes, two rRNA operons, and 32–34 tRNAs. All of the plastid genomes of *Spumella*-like flagellates analyzed herein have typical IR regions, short single-copy (SSC) regions and long single-copy (LSC) regions. Putative lateral gene transfers from other organisms were not detected in any of the newly sequenced genomes.

Photosynthesis-Related Genes

Not surprisingly, the bulk of the gene loss in the plastid genomes of *Spumella*-like flagellates occurred in relation to photosynthesis (Table 1, Figures 1 and 2). This includes loss of genes for photosystem I, photosystem II, the cytochrome *b6/f* complex and ATP synthase subunits (Figure 1). Neither the genes for carbon fixation (RuBisCO) and chlorophyll biosynthesis (*chlI*) nor those encoding cytochrome *c* biogenesis proteins (*ccs1* and *ccsA*) were found to be present in any of the non-photosynthetic chrysophyte plastid genomes.

The *psa* and *psb* gene families encode protein subunits of photosystem I and photosystem II, respectively. Although 10 *psa* genes and two photosystem I assembly protein genes (*ycf3* and *ycf4*) occur in photosynthetic *Ochromonas* sp. CCMP1393 and synuralean species (Ševčíková et al., 2015; Kim et al., 2019), none are present in the plastid genomes of any *Spumella*-like flagellates. Similarly, the colorless cryptophyte plastid genome has lost all 11 *psa* genes found in the genomes of photosynthetic cryptophyte species (Donaher et al., 2009; Kim et al., 2017; Tanifuji et al., 2020), and the plastid genomes of non-photosynthetic diatoms (colorless *Nitzschia* spp.) have also lost all ten *psa* genes found in photosynthetic diatoms (Kamikawa et al., 2018). The photosynthetic chrysophyte plastid genomes possess a total of 16 *psb* genes, but all of these genes were absent in the four *Spumella*-like flagellates. The entire set of 18 *psb* genes

TABLE 1 | Characteristics of chrysophyte plastid genomes analyzed in this study.

General characteristics	<i>Pedospumella</i> sp. Jangsampo120217C5	<i>Poteriospumella</i> <i>lacustris</i> Yongseonkyo072317C3	<i>Spumella</i> sp. Baekdong012018B8	" <i>Spumella</i> " sp. NIES- 1846	<i>Ochromonas</i> sp. CCMP1393	<i>Synura</i> <i>petersenii</i> S114.C7, CZ	<i>Synura</i> <i>sphagnicola</i> CNUKRJ2	<i>Synura uvella</i> CNUKR	<i>Mallomonas</i> <i>splendens</i> CCMP1872	<i>Neotessella</i> <i>volvocina</i> CCMP1871
Nutrition	Heterotroph	Heterotroph	Heterotroph	Heterotroph	Mixotroph	Phototroph	Phototroph	Phototroph	Phototroph	Phototroph
Type of plastid	Leucoplast	Leucoplast	Leucoplast	Leucoplast	Plastid	Plastid	Plastid	Plastid	Plastid	Plastid
Key characteristics for genus classification	Small, colorless, naked single cell, independent clade in nuclear SSU rDNA phylogeny	Small, colorless, naked single cell, independent clade in nuclear SSU rDNA phylogeny	Small, colorless, naked single cell, independent clade in nuclear SSU rDNA phylogeny	Small, colorless, naked cell, independent clade in nuclear SSU rDNA phylogeny	naked single cell	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
Size (bp)	58,577	62,910	55,641	53,209	126,746	133,059	129,699	133,257	146,918	130,705
Inverted repeat (IR)	8,531	8,078	8,283	5,603	22,906	23,151	22,505	23,691	31,611	24,064
Small single-copy region	153	2,144	1,548	1,297	805	1,135	1,191	2,939	711	2,432
Large single-copy region	41,362	44,610	37,527	40,706	80,129	85,622	83,498	82,936	82,985	80,145
G+C (%)	35	32.2	38.9	21.5	30.9	37.89	38.76	38.19	42.39	37.54
Total gene (include RNAs)	97	99	97	82	183	182	181	189	187	186
No. of protein-coding genes	60	63	59	45	144	144	144	151	150	149
tRNAs	33	32	34	33	33	34	33	34	33	33
rRNA operons	2	2	2	2	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	1	2	3	3	7	8	3	9	11	9
Pseudogene	–	–	–	–	–	–	–	<i>trnR</i>	<i>trnR</i>	<i>trnE</i>
Partial copied gene	<i>dnaK</i>	<i>secA</i>	<i>rps14</i>	–	–	–	<i>dnaK</i>	–	–	<i>dnaK</i>
Specific encoded genes	<i>petF</i>	<i>petF</i>	<i>petF</i>	<i>petF</i>	–	<i>dnaB/cemA</i>	<i>dnaB/cemA</i>	<i>dnaB/cemA</i>	<i>dnaB/cemA</i>	<i>cemA</i>
Missing gene	<i>psa-</i> , <i>psb-</i> , <i>pet-</i> , <i>acpP</i> , <i>acsF</i> , <i>cbxB</i> , <i>ccs1</i> , <i>ccsA</i> , <i>chlI</i> , <i>chlL</i> , <i>ftfB</i> , <i>ftfH</i> , <i>groEL</i> , <i>ivvB</i> , <i>ivvH</i> , <i>rbcL</i> , <i>rbcS</i> , <i>rpl24</i> , <i>secA</i> , <i>secG</i> , <i>secY</i> , <i>syfB</i> , <i>tatC</i> ,	<i>psa-</i> , <i>psb-</i> , <i>pet-</i> , <i>acsF</i> , <i>cbxB</i> , <i>ccs1</i> , <i>ccsA</i> , <i>chlI</i> , <i>chlL</i> , <i>groEL</i> , <i>ftfB</i> , <i>ftfH</i> , <i>ivvB</i> , <i>ivvH</i> , <i>rbcL</i> , <i>rbcS</i> , <i>secG</i> , <i>secY</i> , <i>syfB</i> , <i>tatC</i> , <i>tsf</i> ,	<i>psa-</i> , <i>psb-</i> , <i>pet-</i> , <i>acsF</i> , <i>acpP</i> , <i>cbxB</i> , <i>ccs1</i> , <i>ccsA</i> , <i>chlI</i> , <i>dnaK</i> , <i>ftfB</i> , <i>ftfH</i> , <i>groEL</i> , <i>ivvB</i> , <i>ivvH</i> , <i>rbcL</i> , <i>rbcS</i> , <i>rpl24</i> , <i>secA</i> , <i>secG</i> , <i>secY</i> , <i>syfB</i> , <i>tatC</i> , <i>tsf</i> , <i>tufA</i>	<i>psa-</i> , <i>psb-</i> , <i>pet-</i> , <i>acsF</i> , <i>acpP</i> , <i>cbxB</i> , <i>chlI</i> , <i>ccs1</i> , <i>ccsA</i> , <i>clpN</i> , <i>dnaK</i> , <i>ftfB</i> , <i>ftfH</i> , <i>groEL</i> , <i>ivvB</i> , <i>ivvH</i> , <i>rbcL</i> , <i>rbcS</i> , <i>rpl18</i> , <i>rpl23</i> , <i>rpl24</i> , <i>rpl29</i> , <i>rpl31</i> , <i>rpl34</i> , <i>rpl35</i> , <i>rpl36</i> , <i>rps18</i> , <i>rps20</i> , <i>secA</i> , <i>secG</i> , <i>secY</i> , <i>syfB</i> , <i>tatC</i> , <i>tsf</i>	–	<i>syfB</i>	<i>syfB</i>	–	–	–
GenBank accession	MN935477	MN935478	MN935479	AP019363	KJ877675	MH795128	MH795129	NC_040134	NC_040135	MH795132

Bold type indicates gene loss in specific lineages among non-photosynthetic chrysophytes.

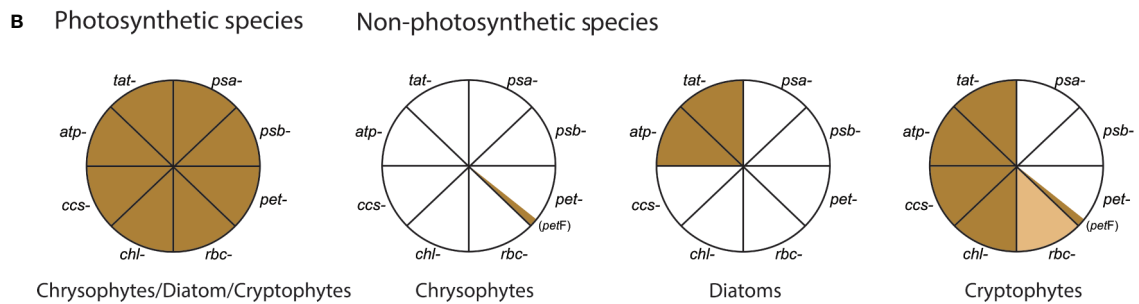
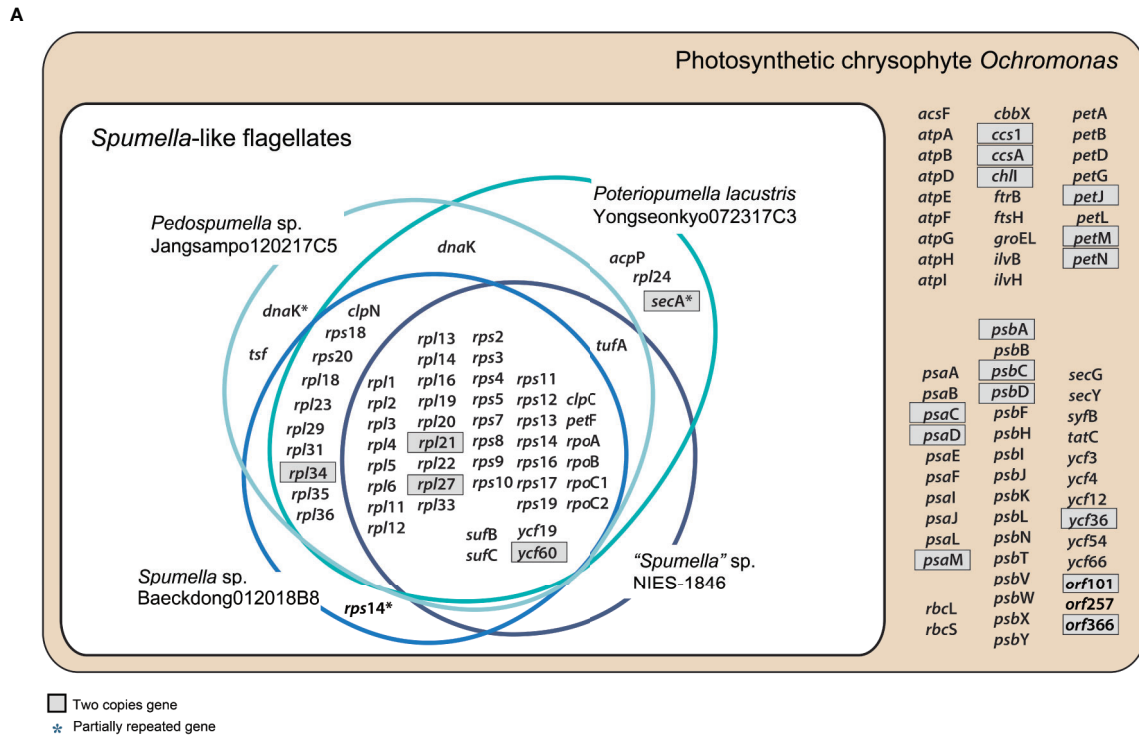


FIGURE 1 | (A) Chrysophyte plastid genome content. Genes with two copies are shown in gray boxes. **(B)** Presence or absence of genes for photosystem I (PSI, *psa*-), photosystem II (PSII, *psb*-), the cytochrome *b6/f* complex (*pet*-), carbon fixation (*rbc*-), chlorophyll biosynthesis (*chl*-), cytochrome *c* biogenesis proteins (*ccs*-), the ATP synthase subunits (*atp*-), and the TAT system (*tat*-) are shown for three distantly related lineages, i.e., photosynthetic and non-photosynthetic chrysophytes, diatoms, and cryptophytes. Genes present/absent in plastid genomes are shown in brown or white, respectively. The *rbc*- gene present only in certain species is colored light brown. The data were derived from previously published studies (Donaher et al., 2009; Kamikawa et al., 2015; Kim et al., 2017; Dorrell et al., 2019; Kim et al., 2019; Tanifuji et al., 2020; this study).

has also been completely lost in non-photosynthetic cryptophytes and diatoms (Figure 1B).

The *pet* (photosynthetic electron transport) gene is almost completely absent from the plastid genomes of all non-photosynthetic *Spumella*-like flagellates, with the curious exception of *petF* (Figures 1A, B). In photosynthetic organisms, *pet* proteins create a complex required for oxygenic photosynthesis, in particular the non-cyclic electron flow mediated by the cytochrome *b6f* complex at the thylakoid membrane. The

petF (ferredoxin) coding region of non-photosynthetic chrysophyte *Spumella*-like flagellates, as well as non-photosynthetic cryptophytes, appears to be intact and the predicted protein is surprisingly well conserved (Figure 1B). The gene may be involved in various redox reactions, given that transcriptome data from *Spumella*-like flagellates (Dorrell et al., 2019) show that genes for the proteins glutamine synthetase and glutamate synthase are expressed, although they do not contain obvious plastid-targeting sequences. In other secondarily non-

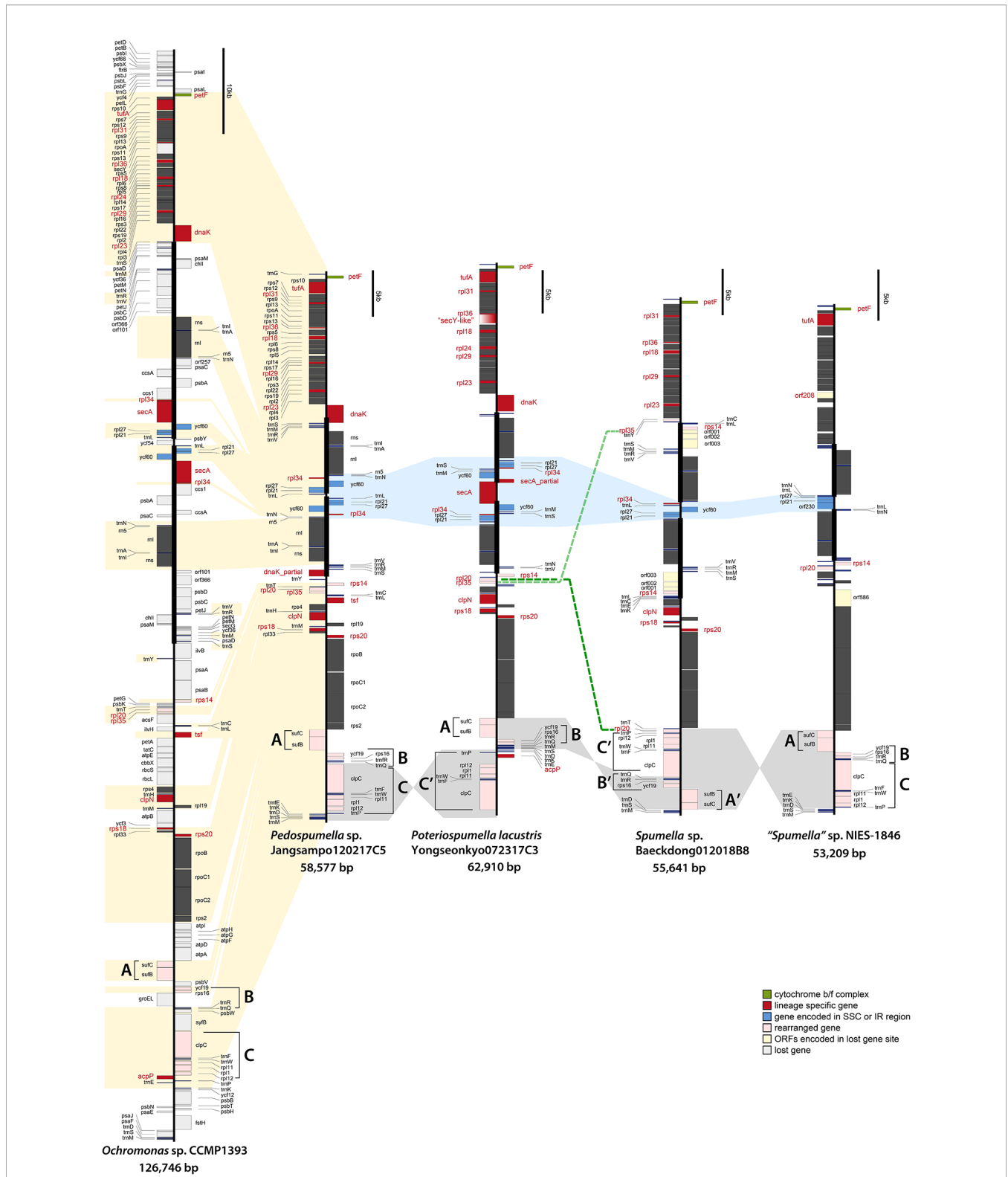


FIGURE 2 | Linearized maps of the plastid genomes of chrysophytes. The structures and coding capacities of the non-photosynthetic chrysophyte plastid genomes examined herein are almost identical to those of the photosynthetic chrysophyte *Ochromonas* sp. CCMP1393. The protein-coding, rRNA and tRNA genes are labeled left or right of the line (genes on the left are transcribed bottom to top, those on the right top to bottom). Inverted repeats are highlighted with thick vertical lines (single-copy regions have thin lines). The gene clusters (A–C) are related to gene rearrangement.

photosynthetic organisms such as the colorless euglenoid *Euglena longa*, the *pet* genes are all missing (Gockel and Hachtel, 2000). In addition to *petF*, the *Spumella*-like flagellates retain *sufB/C* (involved in Fe-S cluster biogenesis) in their plastid genomes.

Plastid genes involved in carbon fixation (RuBisCO subunit, *rbcL* and *rbcS*) and its regulation (*cbbX*), chlorophyll biosynthesis (*chlI*), and cytochrome *c* biogenesis proteins (*ccs1* and *ccsA*) were absent in the *Spumella*-like flagellates but present in the plastids of photosynthetic chrysophytes. The cryptophytes represent an interesting point of comparison in that the *cbbX*, *rbcL* and *rbcS* genes are found in the genome of some colorless *Cryptomonas* species (e.g., *C. paramecium*) but not others (*Cryptomonas* spp. SAG9772f and CCAP1634B), while colorless diatoms appear to uniformly lack such genes (Figure 1B).

In contrast to photosynthetic chrysophytes, non-photosynthetic cryptophytes and diatoms, *Spumella*-like flagellates lack a full complement of plastid ATP synthase subunit genes, which are typically associated with the electron transport chain of photosynthesis (Figure 1B). The plastid genomes of non-photosynthetic cryptophytes contain eight ATP genes (*atpA*, B, D, E, F, G, H, and I), with the exception of an *atpF* pseudogene in *Cryptomonas paramecium* (Donaher et al., 2009; Tanifuji et al., 2020). Non-photosynthetic diatoms retain a near-complete set of ATP synthase genes in their plastid genomes, with *atpE* and *atpF* being present or absent in a species-specific fashion (Kamikawa et al., 2018). The loss of ATP synthase subunit genes in response to the loss of photosynthesis is a recurring theme in land plants and diatoms (Barrett et al., 2014; Kamikawa et al., 2018).

The ATP synthase complex is involved in generating proton gradients for thylakoid protein transporters, mediated by the twin arginine translocator (TAT) system, which depends on proton gradients to perform protein translocation across the thylakoid membrane. The *tatC* gene, the core protein of the TAT system, is absent from the genomes of the four *Spumella*-like flagellates compared here, whereas it is present in photosynthetic chrysophytes and non-photosynthetic cryptophytes and diatoms. With respect to plastid ultrastructure and ATP synthase complex genes, the *Spumella*-like flagellates we analyzed appear to have lost both thylakoid membranes and all of the ATP synthase complex genes that were present in their photosynthetic ancestors (Figure 1B). Interestingly, another *Spumella*-like flagellate, *Cornospumella fuschlensis* AR4D6, retains visible thylakoid structures (Grossmann et al., 2016). The colorless diatom *Nitzschia* sp. retains a few reduced thylakoid membranes in the plastid stroma and ATP synthase complex and TAT genes (Kamikawa et al., 2015), whereas the colorless cryptophyte *Cryptomonas paramecium* does not appear to possess any thylakoid membranes (Sepsenwol, 1973) but has retained many of its ATP synthase complex genes (Donaher et al., 2009; Tanifuji et al., 2020).

Housekeeping Genes in Non-Photosynthetic Chrysophytes

The ribosomal protein operons in the four analyzed *Spumella*-like flagellate plastid genomes were found to be almost identical in terms of gene content and structure relative to each other and to those of the photosynthetic chrysophytes. 25 *rpl* genes for

50S ribosomal subunit proteins are found in photosynthetic chrysophytes, almost all of which are also present in the *Spumella*-like flagellate plastid genomes. The single exception is *rpl24*, which is absent in other *Spumella*-like flagellates but is present in *Poteriospumella lacustris* (Figure 2). The *rpl24* gene has also been described as missing in the plastid genomes of Eustigmatophyceae, a stramenopile lineage that is closely related to chrysophytes based on plastid genome data (Ševčíková et al., 2019).

In terms of 30S ribosomal protein genes, seventeen are present in the plastid genomes of *Spumella*-like flagellates and photosynthetic chrysophyte species, with the exception of *rps8* and *rps20*, which are missing in “*Spumella*” sp. NIES-1846. The *rps14* gene was found to be duplicated and translocated in the genome of *Spumella* sp. Baeckdong012018B8 sequenced herein, whereas the previously published genome of “*Spumella*” sp. NIES-1846 is quite different from the three genomes we sequenced in terms of its *rpl* gene repertoire. No fewer than ten ribosomal protein genes were lost in the NIES-1846 strain: *rpl18*, *rpl23*, *rpl24*, *rpl29*, *rpl31*, *rpl34*, *rpl35*, *rpl36*, *rps18*, and *rps20* (Figures 2 and 3). Six of these *rpl* gene losses (*rpl18*, *rpl23*, *rpl24*, *rpl29*, *rpl31*, and *rpl36*) map to the conserved ribosomal protein operon.

The set of ribosomal protein genes is generally conserved amongst the plastid genomes of diverse species, speaking to their importance in the assembly of functional organellar ribosomes. The smallest known number of plastid ribosomal protein genes occurs in apicomplexan apicoplasts: between 15 and 17 in total (Wilson and Williamson, 1997). In other non-photosynthetic organisms, the number of plastid-encoded ribosomal protein genes are only slightly reduced relative to phototrophs, such as in *Epifagus virginiana* (Wolfe et al., 1992) and the parasitic green alga species *Helicosporidium* sp. (de Koning and Keeling, 2004), and in non-photosynthetic diatoms, chrysophytes, cryptophytes, and parasitic red algae (Figure 3 and Supplementary Table S1). For reasons that are unclear, the loss of ribosomal protein genes goes hand in hand with plastid genome reduction in non-photosynthetic species.

The *tufA* gene, which encodes the plastid protein synthesis elongation factor Tu (EF-Tu), is present in almost all sequenced plastid genomes. In the green line, *tufA* is present in the plastid genomes of most green algae, including the non-photosynthetic *Helicosporidium* and *Polytoma* species, but has been transferred to the nucleus in embryophytes and the charophycean *Zygnema circumcarinatum* (Martin et al., 1998; de Vries et al., 2013). In red alga-derived secondary plastids, *tufA* is present in all sequenced genomes, even in apicomplexans and the non-photosynthetic diatom *Nitzschia* species, the only apparent exception being peridinin-pigmented dinoflagellates, which have extremely reduced plastid genomes (Howe et al., 2008). With the exception of *Spumella* sp. Baeckdong012018B8, all sequenced chrysophyte plastid genomes harbor the *tufA* gene (Figure 2). While the *tufA* gene was found to be missing in the plastid genome of *Spumella* sp. Baeckdong012018B8, gene transcripts for a putatively plastid-targeted EF-Tu protein were detected in transcriptome data from members of the core *Spumella* clade (Dorrell et al., 2019).

et al., 2020). The non-photosynthetic chrysophyte *Spumella*-like flagellate plastid genomes have uniformly lost the *groEL* gene as well (Table 1), while the *dnaK* gene shows a complex pattern of presence and absence in the four genomes we analyzed. The *dnaK* gene is located in the IR region and has been partially copied in both IR regions in *Pedospumella* sp. Jangsampo120217C5. The partially copied *dnaK* gene has also been detected in photosynthetic synuralean algae in a species-specific fashion (Kim et al., 2019). Here we found that the *dnaK* gene is present in *Poteriospumella lacustris* Yongseonkyo072317C3 but is absent in *Spumella* sp. Baekdong012018B8 and the “*Spumella*” sp. NIES-1846 (Figure 2).

The *secA* gene encodes a protein translocase subunit involved in the hydrolysis of ATP to transfer proteins into the thylakoid lumen. The *secA* gene is known to be absent in the glaucophytes, land plants and green algae, but is present in rhodophytes and the red alga-derived plastid genomes of cryptophytes, haptophytes, and stramenopiles. Colorless cryptophytes and diatoms (*Nitzschia* spp.) possess the gene (Dorrell et al., 2019; Tanifuji et al., 2020), but “*Spumella*” sp. NIES-1846 has lost it. In phototrophic chrysophytes, the *secA* gene is present in the SSC region of the plastid genomes of *Ochromonas* sp. CCMP1393 and synuralean species (Kim et al., 2019). In the non-photosynthetic chrysophytes, we found the *secA* gene to be present in *Poteriospumella lacustris* but absent in the plastid genomes of other *Spumella*-like flagellates; this gene absence correlates with extensive gene reduction in the SSC region (Figure 2). Interestingly, a *secY*-like gene was found between *rpl36* and *rpl18* in *Poteriospumella lacustris* (Figure 2, gradient red box). The Sec translocon subunits SecA and SecY, may function together for protein translocation across the thylakoid membranes. The existence of a divergent *secY* gene in *Poteriospumella lacustris* is consistent with the idea of ongoing plastid genome reduction in these organisms.

Previous studies have shown that the *tsf* gene, which encodes elongation factor Ts, is present in primary red algal plastids (Grzebyk et al., 2003). While photosynthetic chrysophytes also have a *tsf* gene in their plastid genomes, among the heterotrophic *Spumella*-like flagellates the gene was found to be present only in *Pedospumella* sp. Jangsampo120217C5; it does not reside in the plastid genomes of the other *Spumella*-like species examined in our study (Figure 2). The *tsf* gene was also not detected in transcriptome data from *Poteriospumella lacustris* (strains JBC07 and JBM10), *Pedospumella encystans* and *Pedospumella sinomuralis* (Dorrell et al., 2019).

Finally, the *acpP* gene, which encodes an acyl carrier protein, thus far appears to be present only in the plastid genome of *Poteriospumella lacustris* and photosynthetic chrysophytes (Ševčíková et al., 2015; Kim et al., 2019) (Figure 2). The acyl carrier protein is involved in the fatty acid biosynthesis pathway and is variably present and absent in the plastid genomes of other algae across the eukaryotic tree of life (Grzebyk et al., 2003).

Plastid Genome Reduction and Rearrangement in Non-Photosynthetic Chrysophytes

A genome-wide comparison of synteny shows that plastid gene order in the *Spumella*-like flagellates is almost identical to that of

photosynthetic chrysophytes (Figure 2), with the complete absence of the photosynthesis-related genes accounting for the bulk of the differences in genome size. We also carried out comparative genomic studies of plastid housekeeping genes in non-photosynthetic species among the red alga-derived plastids, including the “apicoplast” genome of the malaria parasite *Plasmodium falciparum* (Wilson et al., 1996; Waller and McFadden, 2005), colorless cryptophyte *Cryptomonas* species (Donaher et al., 2009; Tanifuji et al., 2020), and the colorless diatom *Nitzschia* sp. (Kamikawa et al., 2018). Overall, the presence of shared plastid genes in chrysophyte lineages reveals non-random retention of genes in the plastid genome despite the independent loss of photosynthesis in the four non-photosynthetic chrysophytes examined herein. Our data support the idea that independently evolved non-photosynthetic plastids show similar genome structure and gene loss patterns to those in the non-photosynthetic chrysophyte plastid genome.

The non-photosynthetic chrysophyte plastid genomes exhibit slightly different gene contents and structures in their inverted repeat (IR) regions (Figure 2). The IRs, which ranged in length from 8.08 kbp (*Poteriospumella lacustris*) to 8.53 kbp (*Pedospumella* sp. Jangsampo120217C5), contained three ribosomal protein genes (*rpl21*, *rpl27*, and *rpl34*), *ycf60*, three rRNAs and eight tRNAs. Twelve protein-coding genes in the repeat region were lost in the plastid genome of *Spumella*-like flagellates relative to the photosynthetic chrysophyte *Ochromonas* sp. CCMP1393 and synuralean plastid genomes (Figure 2).

The plastid genomes of the *Spumella*-like flagellates exhibit different gene orders and gene loss patterns among the three ribosomal proteins (*rpl21*, *rpl27*, and *rpl34*) and *ycf60* in the IR/SSC junctions (Figure 2, marked in blue). The gene order in this region of the *Pedospumella* sp. Jangsampo120217C5 genome was found to be most similar to that of the photosynthetic chrysophyte *Ochromonas* sp. CCMP1393 and the synuralean species (i.e., loss of the *secA* gene and the following gene order: *trnN-rpl34-ycf60-rpl27-rpl21-trnL*; Figure 2). The genes in this region of the *Poteriospumella lacustris* genome were inverted (*trnL-rpl21-rpl27-rpl34-trnS-trnM-ycf60-partial secA*) and also contained a *secA* gene in the SSC region. The *Spumella* sp. Baekdong012018B8 genome was significantly reduced in this area, with the *rpl34-trnL-ycf60-rpl27-rpl21* genes located in the SSC region. Finally, in the plastid genome of “*Spumella*” sp. NIES-1846, only the *trnN-trnL* genes were encoded in the IR region, with the *rpl27-rpl21-orf230* genes being located in the SSC portion of the genome (and, as noted above, the *rpl34* gene was lost).

Previous studies have shown that expansions and contractions are common in the IR boundaries of diatom and green algal genomes (Goulding et al., 1996; Wang et al., 2008; Jansen and Ruhlman, 2012; Sabir et al., 2014; Turmel et al., 2015). Here we have observed that instances of SSC/IR expansion and contraction have also occurred during the evolutionary history of non-photosynthetic *Spumella*-like flagellates, leading to changes in genome structure and coding capacity. Such events can alter gene order through inversion, transposition, gene loss, and/or contraction of the IR region. Such dynamics have also been

observed in the LSC/IR junction of the plastid genomes of photosynthetic synuraleans, leading to gene rearrangements, expansion/contraction of the IR region, and gene loss events (Kim et al., 2019). The contraction of the IR region caused by gene loss is also likely one of the factors contributing to plastid genome reduction in non-photosynthetic chrysophytes.

Gene Rearrangements

The *Spumella*-like flagellate plastid genomes are generally highly syntenic, with the exception of variations in the IR/SSC boundary regions and relative positions of 13 genes located in three clusters designated A, B and C (Figure 2). These clusters are as follows: (A) *sufC-sufB*, (B) *ycf19-rps16-trnR-trnQ*, and (C) *clpC-trnF-trnW-rpl11-rpl1-rpl12-trnP*. Three different cluster patterns were inferred. The first pattern, observed in *Pedospumella* sp. Jangsampo120217C5 and “*Spumella*” sp. NIES-1846, involves the clusters in the order A-B-C, and is shared with photosynthetic *Ochromonas* sp. CCMP1393 and synuralean species. The second pattern, A-B-C', is observed in *Poteriospumella lacustris*, and the third pattern is seen in *Spumella* sp. Baeckdong012018B8; in this case the cluster order is C'-B'-A' and is inverted relative to *Pedospumella* sp. Jangsampo120217C5 and “*Spumella*” sp. NIES-1846. The genes in these three clusters may be translocated and rearranged as a result of the loss of genes flanking the clusters in each species.

Phylogenetic Relationships Among Chrysophytes

Phylogenomic analysis using 40 plastid-encoded proteins showed a monophyletic assemblage of chrysophytes within stramenopiles with maximal support (posterior probability (PP)=1.00, maximum likelihood (ML)=100%, **Supplementary Table S2**). The chrysophyte assemblage formed a sister relationship with Eustigmatophyceae (Figure 4A and **Supplementary Figure S1**), consistent with recently published studies suggesting that eustigmatophyte plastids are closely related to those of chrysophytes (Ševčíková et al., 2015; Han et al., 2019; Kim et al., 2019; Ševčíková et al., 2019). Our plastid phylogenomic investigations showed that the mixotrophic species *Ochromonas* sp. CCMP1393 and the putatively facultative photoautotrophic species *Hydrurus foetidus* formed a strongly supported sister relationship with the exclusively photosynthetic synuralean assemblage (Figure 4A), which is consistent with previous multigene phylogenetic studies based on plastid-encoded proteins (Kim et al., 2019).

We also carried out a phylogenetic analysis of nuclear SSU rDNA sequences from a wide range of chrysophyte species (Figure 4B) to serve as a reference point for interpreting the plastid genome tree and, more generally, to better understand changes in nutritional modes during the evolution of the group. The photosynthetic synuralean lineage was found to form a monophyletic assemblage. We also found that sequences from non-photosynthetic chrysophytes were intermixed with those of mixotrophic lineages (Grossmann et al., 2016; Kristiansen and

Škaloud, 2017; Dorrell et al., 2019; this study). Collectively, the combined results of plastid and nuclear gene phylogenies demonstrate that photosynthesis has been lost multiple times during the evolutionary history of chrysophytes.

The Biology and Evolution of Chrysophytes and Their Plastid Genomes

Photosynthetic chrysophytes are generally mixotrophs that rely on the ingestion of organic nutrients, with the exception of the phototrophic synuralean lineage (Graupner et al., 2018; Lie et al., 2018). The chrysophyte *Hydrurus foetidus* forms filamentous branches within a soft polysaccharide coat and is a putatively facultative phototrophic species (Klaveness and Lindstrøm, 2011; Lavoie et al., 2018). Genome size and cell size differ significantly among species exhibiting different nutritional modes (Olefeld et al., 2018). Where known, heterotrophic chrysophytes have the smallest genomes and cell volumes, phototrophs have the largest, and mixotrophs are generally in between the two. The extreme reduction in the plastid genome of the non-photosynthetic *Spumella*-like flagellates examined herein surely results from the evolution of a heterotrophic lifestyle concomitant with the loss of genes for photosynthesis that are no longer needed. According to Olefeld et al., 2018, the loss of photosynthesis in chrysophytes may be due to energetic factors. Chrysophytes generally do not have a carbon-concentrating mechanism (CCM), indicating that they do not use bicarbonate (HCO_3^-) as a carbon source in diverse water environments (Maberly et al., 2009) and that they are subject to carbon limitation as a selection pressure. Therefore, many chrysophyte lineages may have independently changed their nutritional mode to that of obligate heterotrophs.

The phylogenetic analyses carried out in our study and elsewhere (e.g., Dorrell et al., 2019) provide clear evidence for parallel evolution of plastid genomes in response to a shift to heterotrophy. How many times this occurred is still unclear. Nevertheless, from the data in hand we propose a modified model of chrysophyte plastid evolution, as follows (Figure 5). The red alga-type plastid of photosynthetic chrysophytes stems from a secondary (or possibly) tertiary endosymbiotic event in an ancestor shared with other plastid-bearing stramenopiles (see Sibbald and Archibald, 2020 and references therein; Figure 5, ①). Much later, after the diversification of chrysophytes, *Spumella*-like flagellates lost their photosynthesis-related genes on multiple occasions (Figure 5, ②, ③). Plastid genome reduction occurred as a result of gene loss from one side of the IR/SSC boundary region (Figure 5, ④), and the expansion of the SSC region led to the presence of remnant genes on the other side of the IR/SSC region (Figure 5, ⑤). The loss of much larger numbers of genes brought about complete loss of plastid genomes (Figure 5, ⑥), as observed in the paraphysomonad clade (Dorrell et al., 2019). The loss of the *secA*, *secY*, *tufA*, *rpl*, and *rps* genes in a species-specific fashion and contraction of the IR/SSC boundary region observed in our study provide examples of genome reduction “in action,” providing evidence consistent with this model.

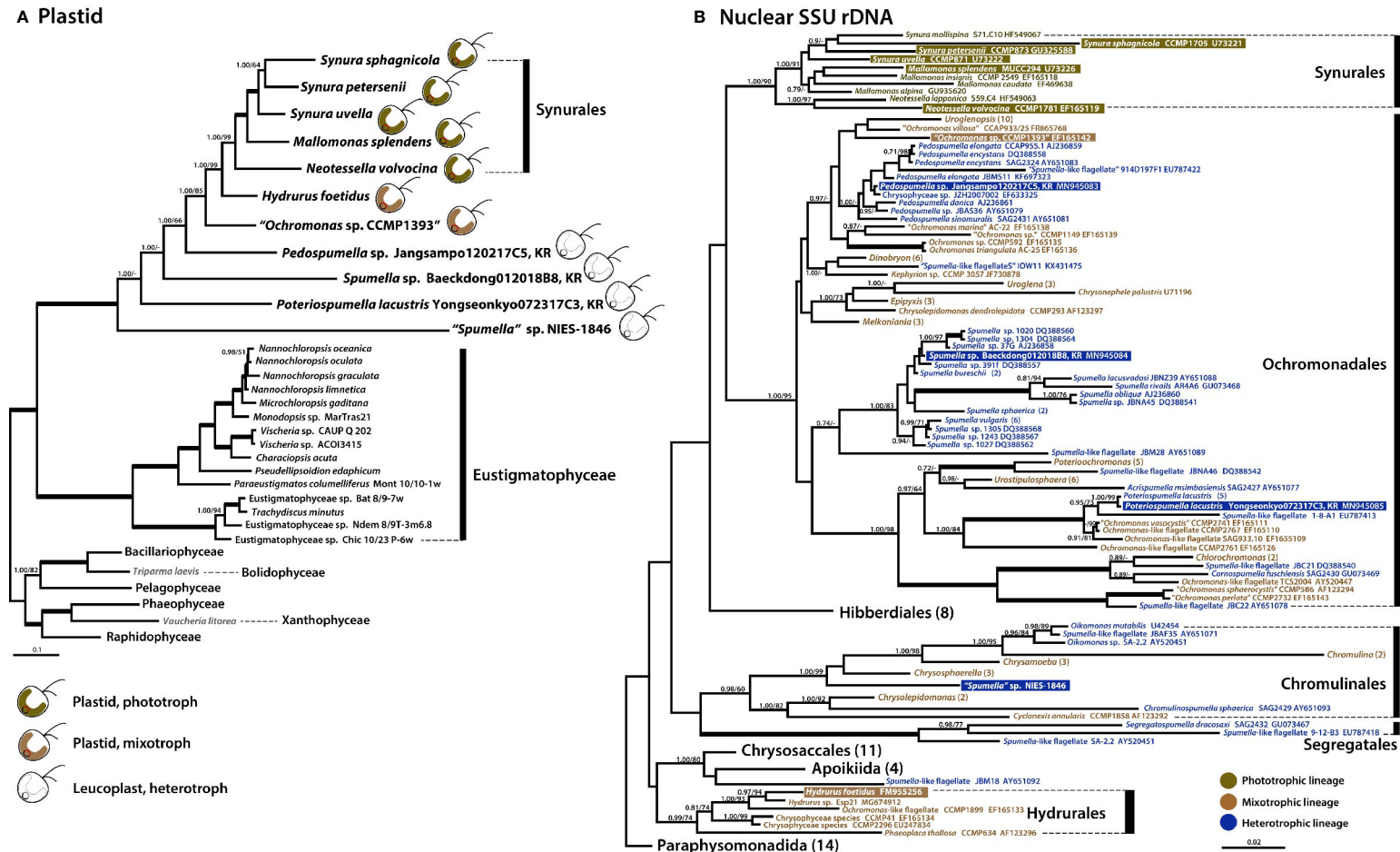
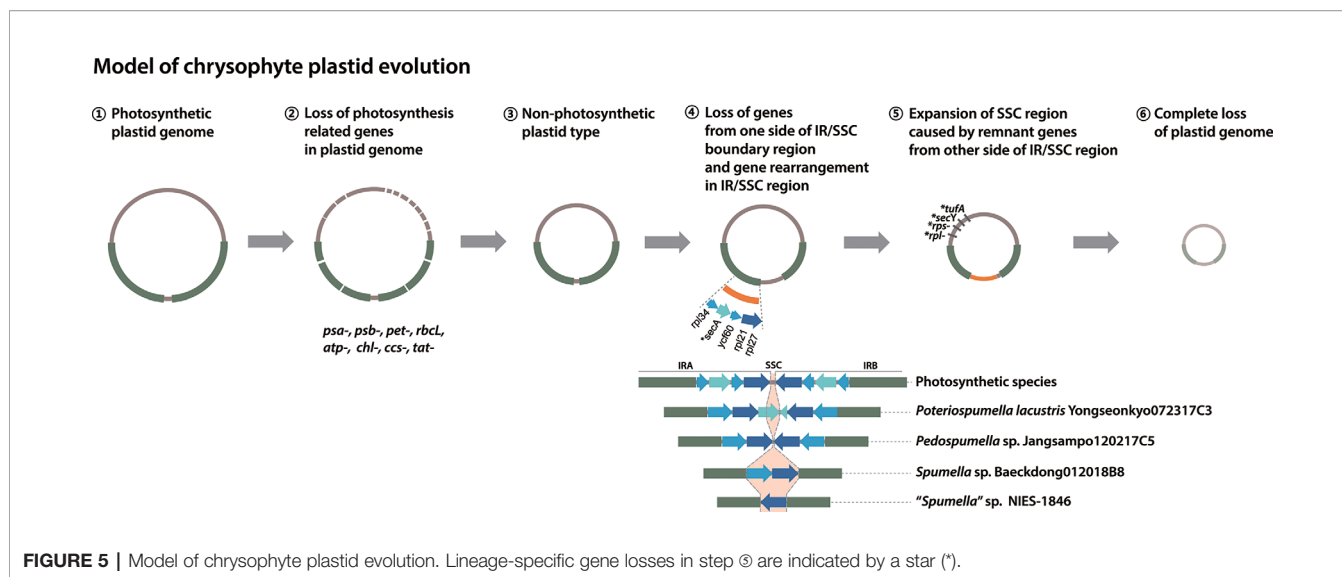


FIGURE 4 | Phylogenies of phototrophic, mixotrophic, and heterotrophic chrysophytes. **(A)** Plastid genome-based Bayesian tree of chrysophytes and other stramenopile taxa. The topology of this tree (inferred from an alignment of 40 proteins and 8,267 amino acids) is consistent with a single acquisition of photosynthetic ability from a red alga-derived secondary plastid in a chrysophyte ancestor. The numbers on each node represent posterior probabilities (left) and ultrafast bootstrap approximation (UFBoot) values calculated using IQ-Tree (right). Thick branches indicate fully supported nodes (PP = 1.00/ML = 100). This tree shows the phylogenetic relationships of chrysophytes and other stramenopiles based on a subset of taxa; for phylogenies inferred using a fully expanded dataset, refer to **Supplementary Figures S1 and S3**. **(B)** Nuclear SSU rDNA tree of chrysophytes showing the putative relationships of the phototrophic chrysophyte lineage in the context of non-photosynthetic lineages. Sequences from the strains whose plastid genomes were sequenced in this study are highlighted. Together, the plastid and nuclear gene tree topologies suggest parallel evolution of chrysophyte plastid genomes in response to shifts to heterotrophy. The numbers on each node represent posterior probabilities (left) and maximum-likelihood (ML) bootstrap support values calculated using RAxML (right). Support values (PP < 0.70/ML < 70) are shown on each node. Bold branches indicates fully supported values (PP = 1.00/ML = 100). The numbers in () are indicate the number of taxa in the species, genus, or order. The species name with “ ” indicates uncertain taxonomic status. An expanded phylogeny based on a much larger dataset is provided in **Supplementary Figure S2**. The scale bars indicate the number of substitutions/site.



CONCLUSIONS

Analysis of three newly sequenced plastid genomes from *Spumella*-like flagellates has provided insight into the fine-scale dynamics of genome reduction in chrysophytes. Heterotrophic lineages appear to have evolved from photosynthetic and mixotrophic ones multiple times independently during chrysophyte evolution. Our results are consistent with previous suggestions that heterotrophic chrysophytes are polyphyletic and their plastid genomes have undergone extreme reduction due to the loss of photosynthesis-related genes. The almost identical gene content and structure of the genomes of *Spumella*-like flagellates analyzed herein suggests that non-photosynthetic chrysophytes have experienced parallel gene losses as they independently transitioned from phototrophy to heterotrophy.

DATA AVAILABILITY STATEMENT

The plastid genome sequences were deposited in the NCBI GenBank database under the following accession numbers: MN935477 (*Pedospumella* sp. Jangsampo120217C5), MN935478 (*Poterospumella lacustris*) and MN935479 (*Spumella* sp. Baeckdong012018B8). The nuclear SSU rDNA sequences of three *Spumella*-like flagellates were deposited in the NCBI GenBank database under the following accession numbers: MN945085 (*Pedospumella* sp. Jangsampo120217C5), MN945083 (*Poterospumella lacustris* Yongseonkyo072317C3), and MN945084 (*Spumella* sp. Baeckdong012018B8).

AUTHOR CONTRIBUTIONS

JIK and WS conceived and designed the experiments. JIK and MJ performed the experiments and analyzed the data. JIK, JMA, and WS interpreted the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by funds from the National Research Foundation (NRF) of Korea (NRF-2015R1D1A1A01057899 and 2018R1D1A1B07050727) provided to JIK, a Natural Sciences and Engineering Research Council of Canada grant (RGPIN-2014) awarded to JMA, and the Collaborative Genome Program funded by the Ministry of Oceans and Fisheries (M01201820180430) and the NRF (2019R1I1A2A01063159 and 2015M1A5A1041808) to WS. These funding organizations were not involved in the design of the study, the collection, analysis or interpretation of the data, or the writing of the manuscript.

ACKNOWLEDGMENTS

The authors thank Drs. Gangman Yi and Jaehee Jung for bioinformatic assistance and Dr. Richard Dorrell for sharing the SSU rDNA sequence of "*Spumella*" sp. NIES-1846. The two reviewers are also thanked for their helpful comments on an earlier version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.572703/full#supplementary-material>

ADDITIONAL FILE 1: SUPPLEMENTARY FIGURE S1 | Phylogenetic tree of chrysophyte plastid-encoded proteins and those of other photosynthetic stramenopiles. This tree was constructed using a dataset of 40 concatenated protein-coding genes (8,297 amino acids) selected with a main focus on the leucoplasts of non-photosynthetic chrysophyte *Spumella*-like flagellates. The numbers on each node represent posterior probabilities (left) and ultrafast bootstrap approximation (UFBoot) values calculated using IQ-Tree (right). The bold branch indicates strongly supported values (PP = 1.00/ML = 100). The scale bar indicates the number of substitutions/site.

ADDITIONAL FILE 2: SUPPLEMENTARY FIGURE S2 | The nuclear encoded SSU rDNA tree of 177 chrysophytes and 3 outgroup taxa showing the putative relationships of the non-photosynthetic chrysophyte lineages relative to the photosynthetic ones. The NCBI accession numbers are provided with taxon names. The numbers on each node represent posterior probabilities using Bayesian analysis. The bold branch indicates strongly supported values (PP = 1.00). The scale bar indicates the number of substitutions/site.

ADDITIONAL FILE 3: SUPPLEMENTARY FIGURE S3 | Phylogenetic tree of chrysophyte plastids and those of other photosynthetic stramenopiles. This tree was constructed using a dataset of 40 concatenated protein-coding genes selected with a main focus on the leucoplasts of non-photosynthetic chrysophyte *Spumella*-like flagellates (8,297 amino acids). The tree was generated using the PMSF model (the LG+F+G tree as guide tree) and ultrafast bootstrap approximation (UFBoot) values calculated using IQ-Tree.

REFERENCES

- Andersen, R. A., Berges, J. A., Harrison, P. J., and Watanabe, M. M. (2005). "Appendix A. Recipes for freshwater and seawater media" in *Algal culturing techniques*. Ed. R. A. Andersen (Elsevier Academic Press), 429–538. doi: 10.1016/B978-012088426-1/50027-5
- Andersen, R. A., Graf, L., Malakhov, Y., and Yoon, H. S. (2017). Rediscovery of the *Ochromonas* type species *Ochromonas triangulata* (Chrysophyceae) from its type locality (Lake Veysove, Donetsk region, Ukraine). *Phycologia* 56, 591–604. doi: 10.2216/17-15.1
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., et al. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* 31, 3095–3112. doi: 10.1093/molbev/msu252
- Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607
- Boenigk, J., and Stadler, P. (2004). Potential toxicity of chrysophytes affiliated with *Poteroiochromonas* and related 'Spumella-like' flagellates. *J. Plankton Res.* 26, 1507–1514. doi: 10.1093/plankt/fbh139
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Borza, T., Popescu, C. E., and Lee, R. W. (2005). Multiple metabolic roles for the non-photosynthetic plastid of the green alga *Prototheca wickerhamii*. *Eukaryot. Cell* 4, 253–261. doi: 10.1128/EC.4.2.253-261.2005
- Bråte, J., Fuss, J., Mehrota, S., Jakobsen, K. S., and Klaveness, D. (2019). Draft genome assembly and transcriptome sequencing of the golden alga *Hydrurus foetidus* (Chrysophyceae) [version 3; peer review: 2 approved]. *F1000Research* 8, 401. doi: 10.12688/f1000research.16734.3
- de Koning, A. P., and Keeling, P. J. (2004). Nucleus-encoded genes for plastid-targeted proteins in *Helicosporidium*: functional diversity of a cryptic plastid in a parasitic alga. *Eukaryot. Cell* 3, 1198–1205. doi: 10.1128/ec.3.5.1198-1205.2004
- de Pamphilis, C. W., and Palmer, J. D. (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348, 337–339. doi: 10.1038/348337a0
- de Vries, J., Habicht, J., Woehle, C., Huang, C., Christa, G., Wägele, H., et al. (2013). Is *ftsH* the key to plastid longevity in sacoglossan slugs? *Genome Biol. Evol.* 5, 2540–2548. doi: 10.1093/gbe/evt205
- Donaher, N., Tanifuji, G., Onodera, N. T., Malfatti, S. A., Chain, P. S. G., Hara, Y., et al. (2009). The complete plastid genome sequence of the secondarily non-photosynthetic alga *Cryptomonas paramecium*: reduction, compaction, and accelerated evolutionary rate. *Genome Biol. Evol.* 1, 439–448. doi: 10.1093/gbe/evp047
- Dorrell, R. G., Azuma, T., Nomura, M., de Kerdrel, G. A., Yang, S., Bowler, C., et al. (2019). Principles of plastid reductive evolution illuminated by non-photosynthetic chrysophytes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 6914–6923. doi: 10.1073/pnas.1819976116
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Figuerola-Martinez, F., Nedelcu, A. M., Smith, D. R., and Reyes-Prieto, A. (2017). The plastid genome of *Polytoma uvella* is the largest known among colorless algae and plants and reflects contrasting evolutionary paths to non-photosynthetic lifestyles. *Plant Physiol.* 173, 932–943. doi: 10.1104/pp.16.01628
- Findenig, B. M., Chatzinotas, A., and Boenigk, J. (2010). Taxonomic and ecological characterization of stomatocysts of *Spumella*-like flagellates (Chrysophyceae). *J. Phycol.* 46, 868–881. doi: 10.1111/j.1529-8817.2010.00892.x
- Gockel, G., and Hachtel, W. (2000). Complete gene map of the plastid genome of the non-photosynthetic euglenoid flagellate *Astasia longa*. *Protist* 151, 347–351. doi: 10.1078/S1434-4610(04)70033-4
- Goulding, S. E., Olmstead, R. G., Morden, C. W., and Wolfe, K. H. (1996). Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* 252, 195–206. doi: 10.1007/BF02173220
- Graupner, N., Jensen, M., Bock, C., Marks, S., Rahmann, S., Beisser, D., et al. (2018). Evolution of heterotrophy in chrysophytes as reflected by comparative transcriptomics. *FEMS Microbiol. Ecol.* 94. doi: 10.1093/femsec/fiy039
- Green, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* 66, 34–44. doi: 10.1111/j.1365-3113X.2011.04541.x
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Grossmann, L., Bock, C., Schweikert, M., and Boenigk, J. (2016). Small but manifold - hidden diversity in "Spumella-like flagellates". *J. Eukaryot. Microbiol.* 63, 419–439. doi: 10.1111/jeu.12287
- Grzebyk, D., Schofield, O., Vetriani, C., and Falkowski, P. G. (2003). The mesozoic radiation of eukaryotic algae: The portable plastid hypothesis. *J. Phycol.* 39, 259–267. doi: 10.1046/j.1529-8817.2003.02082.x
- Hadariová, L., Vesteg, M., Hampl, V., and Krajčovič, J. (2018). Reproductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr. Gen.* 64, 365–387. doi: 10.1007/s00294-017-0761-0
- Han, K. Y., Maciszewski, K., Graf, L., Yang, J. H., Andersen, R. A., Karnkowska, A., et al. (2019). Dictyochophyceae plastid genomes reveal unusual variability in their organization. *J. Phycol.* 55, 1166–1180. doi: 10.1111/jpy.12904
- Howe, C. J., Nisbet, R. E., and Barbrook, A. C. (2008). The remarkable chloroplast genome of dinoflagellates. *J. Exp. Bot.* 59, 1035–1045. doi: 10.1093/jxb/erm292
- Jansen, R. K., and Ruhlman, T. A. (2012). "Plastid genomes of seed plants" in *Genomics of chloroplasts and mitochondria*. Eds. R. Bock and V. Knoop (Springer), 103–126. doi: 10.1007/978-94-007-2920-9_5
- Jung, J., Kim, J.II, Jeong, Y.-S., and Yi, G. (2017). A robust method for finding the automated best matched genes based on grouping similar fragments of large-scale references for genome assembly. *Symmetry* 9, 192. doi: 10.3390/sym9090192
- Jung, J., Kim, J.II, Jeong, Y.-S., and Yi, G. (2018). AGORA: organellar genome annotation from the amino acid and nucleotide references. *Bioinformatics* 34, 2661–2663. doi: 10.1093/bioinformatics/bty196
- Jung, J., Kim, J.II, and Yi, G. (2019). geneCo: a visualized comparative genomic method to analyze multiple genome structures. *Bioinformatics* 35, 5303–5305. doi: 10.1093/bioinformatics/btz596
- Kamikawa, R., Tanifuji, G., Ishikawa, S., Ishii, K.-I., Matsuno, Y., Onodera, N. T., et al. (2015). Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from non-photosynthetic plastid genomes. [Corrected]. *Mol. Biol. Evol.* 32, 2598–2604. doi: 10.1093/molbev/msv134
- Kamikawa, R., Azuma, T., Ishii, K.II, Matsuno, Y., and Miyashita, H. (2018). Diversity of organellar genomes in non-photosynthetic diatoms. *Protist* 169, 351–361. doi: 10.1016/j.protis.2018.04.009
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

- Kim, E., and Archibald, J. M. (2009). “Diversity and evolution of plastids and their genomes,” in *The chloroplast-interactions with the environment*. Eds. H. Aronsson and A. S. Sandelius (Springer-Verlag), 1–39. doi: 10.1007/978-3-540-68696-5
- Kim, J. I., Moore, C. E., Archibald, J. M., Bhattacharya, D., Yi, G., Yoon, H. S., et al. (2017). Evolutionary dynamics of cryptophyte plastid genomes. *Genome Biol. Evol.* 9, 1859–1872. doi: 10.1093/gbe/evx123
- Kim, J. I., Shin, H., Škaloud, P., Jung, J., Yoon, H. S., Archibald, J. M., et al. (2019). Comparative plastid genomics of Synurophyceae: inverted repeat dynamics and gene content variation. *BMC Evol. Biol.* 19, 20. doi: 10.1186/s12862-018-1316-9
- Klavness, D., and Lindström, E.-A. (2011). *Hydrurus foetidus* (Chromista, Chrysophyceae): a large freshwater chromophyte alga in laboratory culture. *Phycol. Res.* 59, 105–112. doi: 10.1111/j.1440-1835.2010.00606.x
- Knauf, U., and Hachtel, W. (2002). The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Mol. Genet. Genomics* 267, 492–497. doi: 10.1007/s00438-002-0681-6
- Kristiansen, K., and Škaloud, P. (2017). “Chrysophyta” in *Handbook of the Protists*. Eds. J. Archibald, A. Simpson and C. Slamovits (Springer), 331–366. doi: 10.1007/978-3-319-28149-0_43
- Lavoie, M., Duval, J. F. L., Raven, J. A., Maps, F., Bejaoui, B., Kieber, D. J., et al. (2018). Carbonate disequilibrium in the external boundary layer of freshwater chrysophytes: implications for contaminant uptake. *Environ. Sci. Technol.* 52, 9403–9411. doi: 10.1021/acs.est.8b00843
- Lie, A. A. Y., Liu, Z., Terrado, R., Tatters, A. O., Heidelberg, K. B., and Caron, D. A. (2018). A tale of two mixotrophic chrysophytes: insights into the metabolisms of two *Ochromonas* species (Chrysophyceae) through a comparison of gene expression. *PLoS One* 13, e0192439. doi: 10.1371/journal.pone.0192439
- Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413
- Maberly, S. C., Ball, L. A., Raven, J. A., and Sultemeyer, D. (2009). Inorganic carbon acquisition by chrysophytes (1). *J. Phycol.* 45, 1052–1061. doi: 10.1111/j.1529-8817.2009.00734.x
- Maciszewski, K., and Karnkowska, A. (2019). Should I stay or should I go? Retention and loss of components in vestigial endosymbiotic organelles. *Curr. Opin. Genet. Dev.* 59, 33–39. doi: 10.1016/j.gde.2019.07.013
- Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., and Kowallik, K. V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162–165. doi: 10.1038/30234
- Nedelcu, A. M. (2001). Complex patterns of plastid 16S rRNA gene evolution in non-photosynthetic green algae. *J. Mol. Evol.* 53, 670–679. doi: 10.1007/s002390010254
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Olefeld, J. L., Majda, S., Albach, D. C., Marks, S., and Boenigk, J. (2018). Genome size of chrysophytes varies with cell size and nutritional mode. *Org. Divers. Evol.* 18, 163–173. doi: 10.1007/s13127-018-0365-7
- Ronquist, F., Teslenki, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Ruck, E. C., Nakov, T., Jansen, R. K., Theriot, E. C., and Alverson, A. J. (2014). Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. *Genome Biol. Evol.* 6, 644–654. doi: 10.1093/gbe/evu039
- Sabir, J. S. M., Yu, M., Ashworth, M. P., Baeshen, N. A., Baeshen, M. N., Bahieldin, A., et al. (2014). Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *PLoS One* 9, e107854. doi: 10.1371/journal.pone.0107854
- Sepsenwol, S. (1973). Leucoplast of the cryptomonad *Chilomonas paramecium*. *Exp. Cell Res.* 76, 395–409. doi: 10.1016/0014-4827(73)90392-3
- Ševčíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., et al. (2015). Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* 5:10134. doi: 10.1038/srep10134
- Ševčíková, T., Yurchenko, T., Fawley, K. P., Amaral, R., Strnad, H., Santos, L. M. A., et al. (2019). Plastid genomes and proteins illuminate the evolution of eustigmatophyte algae and their bacterial endosymbionts. *Genome Biol. Evol.* 11, 362–379. doi: 10.1093/gbe/evz004
- Sibbald, S., and Archibald, J. M. (2020). Genomic insights into plastid evolution. *Genome Biol. Evol.* 12, 978–990. doi: 10.1093/gbe/evaa096
- Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W., and Gillevet, P. M. (1994). The genetic data environment an expandable GUI for multiple sequence analysis. *Comput. Appl. Biosci.* 10, 671–675. doi: 10.1093/bioinformatics/10.6.671
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Suzuki, S., Endoh, R., Manabe, R.II, Ohkuma, M., and Hirakawa, Y. (2018). Multiple losses of photosynthesis and convergent reductive genome evolution in the colourless green algae *Prototheca*. *Sci. Rep.* 8, 940. doi: 10.1038/s41598-017-18378-8
- Tanifuji, G., Kamikawa, R., Moore, C. E., Mills, T., Onodera, N. T., Kashiyama, Y., et al. (2020). Comparative plastid genomics of *Cryptomonas* species reveals fine-scale genome responses to loss of photosynthesis. *Genome Biol. Evol.* 12, 3926–3937. doi: 10.1093/gbe/evaa001
- Turmel, M., Otis, C., and Lemieux, C. (2015). Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. *Genome Biol. Evol.* 7, 2062–2082. doi: 10.1093/gbe/evv130
- Waller, R. F., and McFadden, G.II (2005). The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr. Issues Mol. Biol.* 7, 57–79. doi: 10.21775/cimb.007.057
- Wang, R. J., Cheng, C.-L., Chang, C.-C., Wu, C.-L., Su, T.-M., and Chaw, S.-M. (2008). Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* 8, 36. doi: 10.1186/1471-2148-8-36
- Wilson, R. J., and Williamson, D. H. (1997). Extrachromosomal DNA in the Apicomplexa. *Microbiol. Mol. Biol. Rev.* 61, 1–16. doi: 10.1128/61.1.1-16.1997
- Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., et al. (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 261, 155–172. doi: 10.1006/jmbi.1996.0449
- Wolfe, K. H., Morden, C. W., and Palmer, J. D. (1992). Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10648–10652. doi: 10.1073/pnas.89.22.10648

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kim, Jeong, Archibald and Shin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.