



Comparative Plastid Genomics of Green-Colored Dinoflagellates Unveils Parallel Genome Compaction and RNA Editing

Eriko Matsuo^{1†}, Kounosuke Morita^{1†}, Takuro Nakayama^{1,2†}, Euki Yazaki³, Chihiro Sarai⁴, Kazuya Takahashi⁵, Mitsunori Iwataki^{5*} and Yuji Inagaki^{1,2*}

¹ Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan, ² Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan, ³ RIKEN iTHEMS, Saitama, Japan, ⁴ Graduate School of Science and Engineering, Yamagata University, Yamagata, Japan, ⁵ Asian Natural Environmental Science Center, The University of Tokyo, Tokyo, Japan

OPEN ACCESS

Edited by:

Frederik Leliaert,
Botanic Garden Meise, Belgium

Reviewed by:

Adrian Reyes-Prieto,
University of New Brunswick
Fredericton, Canada
Elisabeth Richardson,
University of Alberta, Canada

*Correspondence:

Mitsunori Iwataki
iwataki@g.ecc.u-tokyo.ac.jp
Yuji Inagaki
yuji@ccs.tsukuba.ac.jp

† These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 12 April 2022

Accepted: 22 June 2022

Published: 11 July 2022

Citation:

Matsuo E, Morita K, Nakayama T,
Yazaki E, Sarai C, Takahashi K,
Iwataki M and Inagaki Y (2022)
Comparative Plastid Genomics
of Green-Colored Dinoflagellates
Unveils Parallel Genome Compaction
and RNA Editing.
Front. Plant Sci. 13:918543.
doi: 10.3389/fpls.2022.918543

Dinoflagellates possess plastids that are diverse in both pigmentation and evolutionary background. One of the plastid types found in dinoflagellates is pigmented with chlorophylls *a* and *b* (Chl *a* + *b*) and originated from the endosymbionts belonging to a small group of green algae, Pedinophyceae. The Chl *a* + *b*-containing plastids have been found in three distantly related dinoflagellates *Lepidodinium* spp., strain MGD, and strain TGD, and were proposed to be derived from separate partnerships between a dinoflagellate (host) and a pedinophycean green alga (endosymbiont). Prior to this study, a plastid genome sequence was only available for *L. chlorophorum*, which was reported to bear the features that were not found in that of the pedinophycean green alga *Pedinomonas minor*, a putative close relative of the endosymbiont that gave rise to the current Chl *a* + *b*-containing plastid. In this study, we sequenced the plastid genomes of strains MGD and TGD to compare with those of *L. chlorophorum* as well as pedinophycean green algae. The mapping of the RNA-seq reads on the corresponding plastid genome identified RNA editing on plastid gene transcripts in the three dinoflagellates. Further, the comparative plastid genomics revealed that the plastid genomes of the three dinoflagellates achieved several features, which are not found in or much less obvious than the pedinophycean plastid genomes determined to date, in parallel.

Keywords: serial secondary endosymbiosis, peDinoflagellates, Pedinophyceae, *Lepidodinium*, RNA editing, complex plastids, plastid replacements

INTRODUCTION

Dinoflagellates are a large group of eukaryotic algae, and one of the major primary producers in the aquatic environment. Some species are infamous for causing red tides and producing deadly toxins causing shellfish poisoning (Jeffrey et al., 1975; Carty and Parrow, 2015). The vast majority of the extant dinoflagellates possess (or used to possess) the plastids containing chlorophylls *a* and *c* (Chl *a* + *c*) plus peridinin, the carotenoid uniquely found in this algal group (“peridinin plastids”;

dinoflagellates bearing peridinin plastids are termed simply as “peridinin dinoflagellates”; Jeffrey et al., 1975; Zapata et al., 2012). Peridinin plastids were one of the “secondary plastids” derived from red algal endosymbionts and are believed to be established prior to the divergence of the extant dinoflagellates (Archibald, 2009; Keeling, 2010; Sibbald and Archibald, 2020). The plastid genomes in diverse peridinin dinoflagellates comprise multiple mini-circles (Zhang et al., 1999; Barbrook et al., 2019), each of which carries a single or a few genes. Further, the transcripts from mini-circles have been known to undergo post-transcriptional base conversion (base conversion RNA editing), which converts the four nucleotides in transcripts to almost all possible others (Zauner et al., 2004; Dang and Green, 2009; Mungpakdee et al., 2014; Klinger et al., 2018). Such promiscuous base conversion editing in peridinin plastids is distinctive from the RNA editing in organelles of other eukaryotes intensively studied so far [e.g., land plants; Shikanai (2006) and Hao et al. (2021)], leaving the details of the RNA editing machinery in peridinin plastids uncertain.

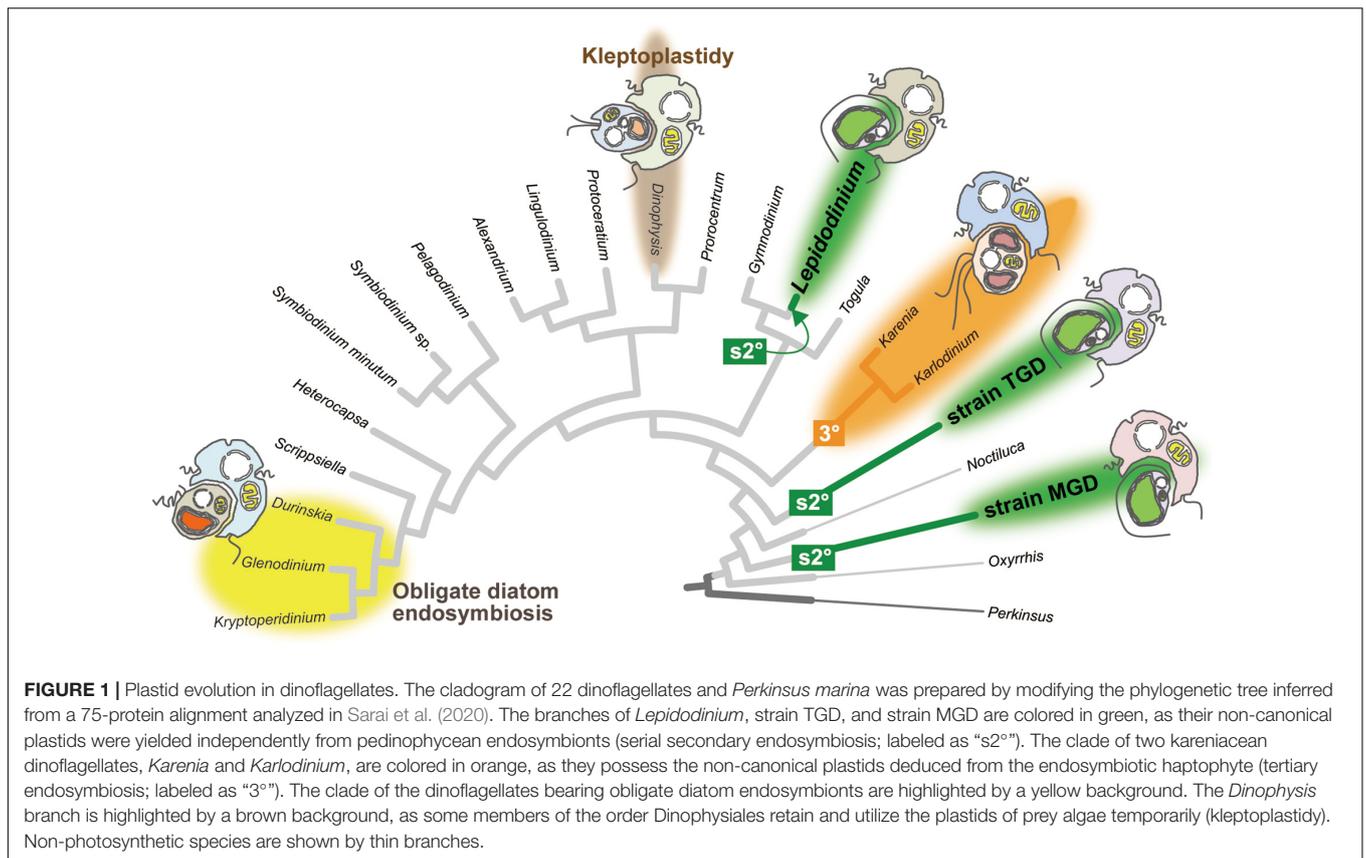
Dinoflagellates have been anticipated to provide clues to understand the evolutionary process transforming an endosymbiotic alga into the host-governed organelle (i.e., plastid), as peridinin plastids have been replaced by phylogenetically diverse algae taken up as the endosymbionts on multiple branches in the tree of dinoflagellates. **Figure 1** schematically summarizes the lineages/species bearing “non-canonical plastids” lacking peridinin (Archibald, 2009; Keeling, 2010; Sibbald and Archibald, 2020) and obligate diatom endosymbionts (Takano et al., 2008; Yamada et al., 2017). Members of the family Kareniaceae possess the non-canonical plastids containing Chl *a* + *c* plus 19'-hexanoyloxyfucoxanthin instead of peridinin (Bjørnland et al., 2003; Zapata et al., 2012). Consistent with the pigment composition, phylogenies of plastid genes suggested that the non-canonical plastids in kareniacean dinoflagellates are the product of “tertiary endosymbiosis,” in which an endosymbiotic haptophyte was reduced and integrated genetically into the dinoflagellate host as the plastid (Tengs et al., 2000; Nosenko et al., 2006; Burki et al., 2014; Bentlage et al., 2016; highlighted as “3°” in **Figure 1**). The plastid gene transcripts of kareniacean dinoflagellates were shown to receive promiscuous base conversion editing, which is similar to but more intense than that of peridinin plastids (Klinger et al., 2018).

Another type of non-canonical plastid, which contains chlorophylls *a* and *b* (Chl *a* + *b*), was found initially in members of the genus *Lepidodinium* (Watanabe et al., 1987, 1990). A phylogenetic study based on multiple plastid genes pinpointed the origin of the *Lepidodinium* plastid as a pedinophycean green alga (Kamikawa et al., 2015). The host phylogeny inferred from the nucleus-encoded ribosomal RNA (rRNA) sequences put *Lepidodinium* spp. within peridinin dinoflagellates, such as *Gymnodinium catenatum* (Saldarriaga et al., 2001; Shalchian-Tabrizi et al., 2006; Matsumoto et al., 2012), suggesting that the ancestral *Lepidodinium* cell replaced the peridinin plastid with a Chl *a* + *b*-containing plastid through the endosymbiotic partnership with a pedinophycean green alga (termed as “serial secondary endosymbiosis;” highlighted as “s2°” in **Figure 1**). Similar to the non-canonical plastids in kareniacean

dinoflagellates, the evidence for the genetic integration of the pedinophycean-derived plastid into the dinoflagellate host has been accumulated (Minge et al., 2010; Matsuo and Inagaki, 2018). More recently, some of us reported the second and third dinoflagellates bearing Chl *a* + *b*-containing plastids, strains MGD and TGD (Nakayama et al., 2020; Sarai et al., 2020). In the host phylogeny inferred from an alignment of 75 nucleus-encoded proteins, *L. chlorophorum*, strain MGD, and strain TGD were distantly related to each other, suggesting that serial secondary endosymbiosis occurred on the three separate branches in the tree of dinoflagellates (the three branches labeled with “s2°” in **Figure 1**). The plastid phylogeny inferred from plastid small subunit rRNA sequences recovered the groupings described below with high statistical support, namely (i) the monophyly of the three dinoflagellates bearing Chl *a* + *b*-containing plastids and (ii) the sister relationship between the dinoflagellate clade and the clade of *Pedinomonas minor* and *P. tuberculata* (Sarai et al., 2020). Thus, the endosymbiont alga, which gave rise to the Chl *a* + *b*-containing plastids in *L. chlorophorum*, strain MGD, and strain TGD, belong to or are closely related to the genus *Pedinomonas*. Here, we designate the three dinoflagellates bearing pedinophycean-derived plastids as “peDinoflagellates,” and their plastids as “peDinoflagellate plastids.”

We have been interested in extracting the key aspects that enabled serial secondary endosymbiosis by comparative studies of the three peDinoflagellates (Nakayama et al., 2020; Sarai et al., 2020). In the line of our research interest described above, we here evaluate how the plastid genomes of the pedinophycean endosymbionts were modified during serial secondary endosymbioses by comparative plastid genomics. Kamikawa et al. (2015) sequenced the plastid genome of *L. chlorophorum* completely and reported that the peDinoflagellate plastid genome is more compact than that of *P. minor* in terms of the repertoire of functionally assignable open reading frames (ORFs). In addition, the *L. chlorophorum* plastid genome has the features that were not found in the pedinophycean plastid genomes; (i) absence of inverted repeats (IRs), (ii) frequent ORF/gene overlapping/fusion, (iii) a deviant genetic code in which AUA codon is assigned as methionine (Met) instead of isoleucine (Ile), and (iv) pseudogenization (Kamikawa et al., 2015). Thus, it is intriguing whether the above-mentioned features found in the *L. chlorophorum* plastid genome are shared with either or both MGD and TGD plastid genomes.

In this study, we sequenced the plastid genome of strain TGD completely and that of strain MGD nearly completely (the latter genome could not be completed due to repeat sequences). The TGD plastid genome is a circular molecule of approximately 71 Kb, while strain MGD possesses the circular plastid genome of approximately 102 Kb. The current study revealed that the three peDinoflagellate plastid genomes shared many of the features that were identified by the comparison between the plastid genomes of *L. chlorophorum* and *P. minor*. In addition, we found RNA editing on plastid gene transcripts in the three peDinoflagellates. Base conversion editing on the plastid gene transcripts appeared to be common among the three peDinoflagellate plastids, while we identified a single case of base insertion editing on the *psaA*



transcript in strain TGD. Overall, the patterns of base conversion editing were similar among the dinoflagellates bearing peridinin plastids and the two types of non-canonical plastids except for that of strain MGD. If *L. chlorophorum*, strain MGD, and strain TGD truly established their current plastids separately, the modifications of plastid genomes in the pedinophycean endosymbionts occurred in a highly parallel manner during separate serial secondary endosymbioses.

MATERIALS AND METHODS

DNA Extraction From the Cultured Cells

PeDinoflagellate strains TGD and MGD established in Sarai et al. (2020) have been maintained in our laboratories and were used in this study. The culture conditions of the two strains were the same as described in Sarai et al. (2020). The algal cultures were observed by light microscopy regularly and the cells in confluent cultures were harvested by centrifugation at 1,720 *g* for 10 min. The genomic DNA was extracted from the harvested cells by the cetyl trimethyl ammonium bromide (CTAB) method. The cell pellet was dissolved in 500 μ L of CTAB extraction buffer [per 100 mL, 2 g CTAB, 10 mL of 1 M Tris-HCl (pH 8.0), 4 mL of 0.5 M ethylenediaminetetraacetic acid (pH 8.0), and 35 mL of 4 M NaCl] at 65°C for 1 h. After the cells were completely dissolved, 500 μ L of Chloroform:Isoamyl Alcohol Solution [480 μ L chloroform: 20 μ L Isoamyl Alcohol] was added to the tube. After vigorous vortexing, the aqueous solution was

saved in a fresh tube by centrifugation at 13,200 rpm for 5 min at 20°C. Isopropanol precipitation was carried out by adding 500 μ L of isopropanol to the aqueous solution, followed by centrifugation at 13,200 rpm for 20 min at 4°C. The pellet was rinsed with 1 mL of 70% ethanol, and the supernatant was discarded after centrifugation at 13,200 rpm for 5 min at 4°C. The DNA pellet was briefly dried up and then dissolved into 50 μ L of sterile distilled water.

Sequencing, Assembling, and Annotation of the Plastid Genomes

The genomic DNA sample of strain TGD was subjected to Genome-seq analysis using the Illumina Next-seq platform. Approximately 127 million of 150 base paired-end reads were generated (38.1 Gb in total). Initial reads were examined by fastQC to filter the reads containing low-quality bases (under 35) more than 20%. We trimmed the adapter sequence and excluded low-quality bases by FASTX toolkit, yielding approximately 82 million reads for the analyses described below.

We reconstructed a circular plastid genome of strain TGD by the two steps described below. As the *de novo* assembling of the 82 million reads was computationally intense, approximately 4 million reads were assembled into 116,551 contigs. TBLASTN search was carried out against the resultant contigs using the amino acid sequences of the plastid-encoded proteins of two green algae, *P. minor* and *Chlorella vulgaris*, as queries (GenBank accession numbers NC_016733 and AB001684). We retrieved

8 contigs, which were approximately 71 Kb in total length, as the tentative plastid genome of strain TGD. 37,646 reads were selected for the second assembling by mapping the 82 million reads on the tentative plastid genome contigs. We connected the resultant contigs by combining the paired-end information and PCR experiments. Finally, a circular DNA molecule of 71,225 bp was reconstructed as the plastid genome of strain TGD. SPAdes ver.3.7.1 (Prjibelski et al., 2020) and Bowtie2 (Langmead and Salzberg, 2012) were used for *de novo* assembling and mapping, respectively.

The genomic DNA sample of strain MGD was shipped to a biotech company (Hokkaido System Science Co., Ltd., Hokkaido, Japan) for Genome-seq analysis using the Illumina HiSeq 2000 platform. Approximately 224 million of 100 bp paired-end reads were generated (33.0 Gb in total). After the quality control (see above), approximately 60 million reads were assembled into 379,981 contigs using Ray (Boisvert et al., 2010). We repeated the TBLASTN search described above and identified a single contig of approximately 100 Kb in length as the plastid genome contig. The PCR experiment using a set of the primers, which were designed based on both edges of the 100-Kb contig (5'-GGGGAGAAATTTCAAGATACGG-3' and 5'-GGGAGGCAAAGGATAAACTAAACG-3'), amplified a single DNA fragment of approximately 2 Kb in length. We failed to determine the complete nucleotide sequence of the amplicon which is largely composed of the repeats of 84 bp (5'-TTA TTTAATGTCACAAAGCCAAATATATAGGCTTTCTATATGTA GAAAGACCAGTTTATTTAATAAAAAGAATAAATTTTAT GT-3'). We conclude that strain MGD has a circular plastid genome of approximately 102 Kb in length, albeit the exact length of the plastid genome remains uncertain.

We annotated the plastid genomes of strains TGD and MGD as follows. The ORFs encoding polypeptides of equal to or more than 100 amino acid residues were identified by MFannot¹. Genes encoding transfer RNAs (tRNAs) were surveyed by tRNAscan-SE² (Chan and Lowe, 2019). Ribosomal RNAs, RNase P RNA, and introns were investigated by RNAweasel³.

Detection of Possible RNA Editing

The possibility of RNA editing was explored by comparing RNA-seq reads with the plastid genome sequences. The RNA-seq data of strain TGD (GenBank accession number DRR190720), strain MGD (DRR190721), and *L. chlorophorum* (DRR124369) were downloaded from DDBJ Sequence Read Archive (Kodama et al., 2012). We also downloaded the RNA-seq data of *P. minor* SAG 1965-3 (ERR2041093) generated by the 1000 Plant Transcriptomes Initiative⁴ (Carpenter et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). After quality control with FASTP v.0.12.4 (Chen et al., 2018), all reads were mapped with HISAT2 v.2.2.1 (Kim et al., 2019) to the corresponding plastid genomes. Besides the two plastid genomes of strains MGD and TGD determined in this study, we used those of *L. chlorophorum*

and *P. minor* deposited in the GenBank database under the accession numbers NC_027093.1 and FJ968740.1, respectively. The numbers of RNA-seq short reads subjected to the mapping and those aligned with the corresponding plastid genomes are summarized in **Supplementary Table 1**.

The incongruities between the RNA-seq reads and plastid genome sequences were detected with mpileup and bcftools commands from bcftools v.1.9 (Li, 2011; Danecek et al., 2021). In this study, we set the two conditions to identify the positions that underwent RNA editing. First, quality-scores of variant calls were greater than 200. Second, the candidate positions are located in ORF/gene coding regions. Only the positions fulfilling both of the two conditions were regarded as post-transcriptional base conversions.

Phylogenetic Analyses of Plastid-Encoded Proteins

The phylogenetic relationship among green algal plastids and three peDinoflagellate plastids was examined by the maximum-likelihood (ML) phylogenetic analysis of 50 plastid-encoded proteins. The alignment generated by Kamikawa et al. (2015) was modified by adding the plastid-encoded proteins of strains TGD and MGD, as well as three free-living pedinophycean green algae, namely *P. tuberculata*, *Marsupiomonas* sp., and strain YPF701 (Lemieux et al., 2014; Jackson et al., 2018). The amino acid sequences were aligned by MAFFT v.7.490 with L-INS-I option (Katoh, 2002), following manual trimming of ambiguously aligned positions. The final “50-protein” alignment comprised 50 plastid-encoded proteins from 37 taxa (8,736 amino acid positions in total). The 50-protein alignment was subjected to the ML analysis with the LG + Γ + F + C60. Non-parametric bootstrap support values were calculated by 100-replicate ML bootstrap analysis with the LG + Γ + F + C60 + PMSF [posterior mean site frequencies; see Wang et al. (2018)] model. The ML and ML bootstrap analyses described above were repeated after excluding two out of the three peDinoflagellates considered in the original 50-protein alignment. We used IQTREE v.2.1.3 (Nguyen et al., 2015) for both ML and ML bootstrap analyses described above. The alignments generated in this study are available as a part of the **Supplementary Data**.

We also conducted Bayesian phylogenetic analysis with the CAT + GTR model by using Phylobayes v.1.8a (Lartillot and Philippe, 2004, 2006; Lartillot et al., 2007). Two Markov chain Monte Carlo (MCMC) runs ran for 10,000 cycles and the consensus tree with branch lengths and Bayesian posterior probabilities (BPPs) were calculated after the first one-fourth cycles were discarded as “burn-in.” Note that the maxdiff value stayed large (0.371194) but we found that the trees from the two MCMC chains agreed largely with the ML tree.

Comparison in Branch Length Between peDinoflagellate and Pedinophycean Green Algae

The 50-protein phylogeny indicated that the branches of the three peDinoflagellates were much longer than those of the

¹<http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>

²<http://lowelab.ucsc.edu/tRNAscan-SE/>

³<https://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl>

⁴<http://www.onekp.com/samples/list.php>

four pedinophycean green algae (see section “Results”). We examined the magnitude of “long branch-ness” of the three peDinoflagellates in the individual single-protein alignments. We prepared a 5-taxon tree comprising the four pedinophycean green algae and one of the three peDinoflagellates. The 5-taxon tree was enforced to have the sister relationship between *P. minor* and *P. tuberculata* and that between *Marsupiomonas* sp. and Pedinophyceae sp. YPF-701. The branch lengths of the 5-taxon tree were optimized over each of the 50 single-protein alignments. The branch length optimization was performed by IQTREE v.2.1.3 (Nguyen et al., 2015) with the LG + Γ + F + C60 model.

For each tree, the length of the peDinoflagellate branch was subtracted by the sum of the lengths of the rest of branches to obtain “branch length-ratios.” We split the 50 plastid-encoded proteins into two functional categories, “photosynthetic” and “non-photosynthetic.” The former category contained 30 proteins involved in photosynthesis, while the latter was composed of 17 ribosomal proteins, translation elongation factor Tu, Ycf3, and ClpP. The branch length-ratios of the 30 photosynthetic proteins were compared with those of the 20 non-photosynthetic proteins by Wilcoxon rank-sum test.

RESULTS

Overview of the Plastid Genome of the peDinoflagellate Strain TGD

We completely sequenced a single, circular DNA molecule of 71,225 bp in length as the plastid genome of the peDinoflagellate strain TGD (TGD plastid genome). The circular genome map is provided as **Figure 2** and the general features are summarized in **Table 1**. The content of guanine plus cytosine (GC content) is 34.8%. 69 ORFs, the genes for small and large subunit rRNA (*rns* and *rnl*), and the genes for 27 tRNAs were identified in the genome. 67 out of the 69 ORFs were functionally assignable. Neither BLAST nor Pfam search provided any clue to the function of *orf123* or *orf156*. By mapping the RNA-seq reads on the nucleotide sequence of the plastid genome, the incongruities of the nucleotide identity were detected between the genome and transcripts at 177 positions (0.0327% of the coding region; **Table 2**) in 40 ORFs and *rnl* (marked by stars in **Figure 2**). Thus, we concluded that base conversion editing occurred at the 177 positions in the plastid gene transcripts. The RpoB-coding region was found to be interrupted by a single stop codon in the genome. Likewise, a frameshift hinders the recovery of the continuous PsaA-coding region in the genome (a zigzag line in **Figure 2**). After referring to the RNA-seq data mapped on the two regions, we regard *rpoB* as a pseudogene while *psaA* is a functional gene. Although *rpoB* seems to be transcribed at a certain level, the stop codon remains in the corresponding transcripts. On the other hand, the single reading frame encoding the entire PsaA was recovered post-transcriptionally by the insertion of two consecutive nucleotides (**Supplementary Figure 1**). Thus, we conclude that the *psaA* transcript receives base insertion editing and is functional. The RpoC2-coding region was split into two separate ORFs (*rpoC2_N*

and *rpoC2_C*; **Figure 2**) and we detected the distinct transcripts from the two ORFs by analyzing the RNA-seq data. No intron was found in any ORFs identified. Intergenic regions occupy 17.9% of the plastid genome. Three pairs of ORFs, namely (i) *psaM* and *psbK*, (ii) *rps8* and *rpl36*, and (iii) *petL* and *petG* are fused to each other (colored in purple in **Figure 2**). We had no evidence for any post-transcriptional processing of the transcripts from the three fused ORFs, except for base conversion editing. Noteworthy, the coding region for PsaM-PsbK fused protein was found to encode the entire amino acid sequence of Ycf12 on a different reading frame (**Supplementary Figure 2**). Ten pairs of neighboring ORFs are on distinct reading frames but partially overlap each other. We found a single case of partial overlapping between a tRNA gene and an ORF. The ORF/gene overlappings described above are highlighted by red dots in **Figure 2** (see also **Supplementary Table 2** for the details). IRs were not found. One of the three Ile codons in the standard genetic code, AUA, most likely assigns Met (AUA = Met) in the TGD plastid genome, as reported in that of *L. chlorophorum* (Matsumoto et al., 2011).

Overview of the Plastid Genome of the peDinoflagellate Strain MGD

We recovered a single, circular DNA molecule of approximately 102 Kb in length as the plastid genome of the peDinoflagellate strain MGD (MGD plastid genome). See **Figure 3** and **Table 1** for the circular genome map and general features, respectively. The precise length of the plastid genome remains uncertain, as the nucleotide sequence of the region composed of the 84 bp-repeats could not be determined completely. The GC content is 34.6%. Seventy-one ORFs, two intron-encoded ORFs, *rns*, *rnl*, and the genes for 28 tRNAs were annotated in the genome. By mapping RNA-seq reads on the genome, we detected base conversion editing at 18 positions (0.0281% of the coding region; **Table 2**) in the transcripts of 12 ORFs, *rns*, and *rnl* (marked by stars in **Figure 3**). No sign of base insertion editing was detected. The functions of three out of the 71 ORFs could not be assigned (i.e., *orf158*, *orf155*, and *orf172*). We regard *ycf4* as a pseudogene ($\Psi ycf4$), as the putative N- and C-termini were found to be coded in different reading frames and few transcripts were mapped on this region (**Supplementary Figure 3**). A single group II intron, which hosts two ORFs (*orf107* and *orf355*), was found in *psbB*. The non-coding region occupies 34.6% of the plastid genome. Four pairs of ORFs, namely *rpoA* and *rps9*, *rpl5* and *rps8*, *rps19* and *rps3*, and *rps12* and *rps7* are fused to each other (colored in purple in **Figure 3**). No post-transcriptional processing was observed for the transcripts from the fused ORFs, except for base conversion editing that occurred on the *rpoA-rps9* and *rpl5-rps8* transcripts. Eight pairs of neighboring ORFs are on distinct reading frames but partially overlap each other. We found a single case of partial overlapping between a tRNA gene and an ORF. The ORF/gene overlappings described above are highlighted by red dots in **Figure 3** (see also **Supplementary Table 2** for the details). IRs were not found. The genetic code used in the MGD plastid genome appeared to be the same as those in the plastid genomes of *L. chlorophorum* and strain TGD (see above).

TABLE 1 | General features of the plastid genomes in four pedinophycean green algae and three peDinoflagellates.

	<i>Pedinomonas minor</i>	<i>Pedinomonas tuberculata</i>	<i>Marsupiomonas sp. NIES-1824</i>	<i>Pedinophyceae sp. YPF701</i>	<i>Lepidodinium chlorophorum</i>	Strain TGD	Strain MGD
GC content (%)	34.8	33.4	40.3	37.7	34.6	34.8	34.6
Genome size (bp)	98,940	126,694	94,262	91,755	66,223	71,225	~102,000
Intergenic regions (%)	25.6	33.8	24.0	27.0	13.3	17.9	37.3
Inverted repeats	+	+	+	+	–	–	–
functionally assignable ORFs	75	74	74	73	61	67	68
rRNA genes	4	6	6	6	2	2	2
tRNA genes	28	28	28	24	27	27	28
Gene containing introns	0	8	0	0	3	0	1
Pseudogenes	0	0	0	0	1	1	1
Gene fusion	0	0	0	0	2	3	4
Gene overlapping	2	1	2	1	11	12	12
Genetic code	Standard	Standard	Standard	Standard	AUA = Met	AUA = Met	AUA = Met
RNA-editing	–	Not examined	Not examined	Not examined	+	+	+
No. of the editing sites	–	–	–	–	188	177	18
% of the editing sites in the coding sites	–	–	–	–	0.327%	0.304%	0.0281%

the affinity between the peDinoflagellates and *Pedinomonas* spp. (Figure 4).

Difference in Evolutionary Tempo Between Photosynthetic and Non-photosynthetic Plastid-Encoded Proteins in the Three peDinoflagellates

The branches of the three peDinoflagellates in the 50-protein phylogeny were found to be longer than those of the pedinophycean green algae (Figure 4). The ratios of the branch length for a peDinoflagellate to those of the four pedinophycean green algae (branch length-ratios) were calculated for each protein considered in the alignment, aiming to identify the plastid-encoded proteins that contributed to the long branch-ness observed in the 50-protein phylogeny (Figure 5A). The peDinoflagellate branches were generally long in the vast majority of the single-protein trees, regardless of the peDinoflagellate included (see Supplementary Material). Nevertheless, the branch length-ratios calculated from the non-photosynthetic proteins tend to be larger than those from the photosynthetic proteins (Figure 5B). The Wilcoxon rank-sum test rejected the null hypothesis of no difference between the median values of the branch length-ratios calculated from the two categories at the 1% level ($p = 7.50 \times 10^{-3}$, 3.49×10^{-5} , and 2.14×10^{-3} in the comparisons considering *L. chlorophorum*, strain MGD, and strain TGD, respectively). These results suggest that the overall substitution rates in non-photosynthetic plastid-encoded proteins are higher than those in photosynthetic plastid-encoded proteins in the three peDinoflagellates. The significant difference in substitution rate between photosynthetic and non-photosynthetic plastid-encoded proteins may have been achieved separately in *L. chlorophorum*, strain MGD, and strain TGD.

Features Shared Among the peDinoflagellate Plastid Genomes but Not Found in the Pedinophycean Plastid Genomes

Prior to the comparison between the plastid genomes of the three peDinoflagellates and those of pedinophycean green algae, we briefly review the four pedinophycean plastid genomes sequenced completely to date (Table 1). As do phylogenetically diverse algae, the four pedinophycean green algae have the plastid genomes with low GC content (33.4–40.3%) and IRs (Table 1). Although the sizes of the four plastid genomes vary from 91.8 to 126.7 Kb, they appeared to carry similar numbers of functionally assignable ORFs (73–75; henceforth here designated simply as “ORFs”). The plastid genome of *Pedinomonas tuberculata* (126.7 Kb) appeared to be larger than the rest of the three plastid genomes (91.8–98.3 Kb) compared here, partially due to 7 introns in the *P. tuberculata* plastid genome but none in others (Table 1).

We revealed 7 of the plastid genome features that were shared among the three peDinoflagellates but were not found or much less obvious in the pedinophycean green algae. (i) The peDinoflagellate plastid genomes commonly lack IRs

TABLE 2 | Patterns of base-conversion editing in the plastid gene transcripts of the three peDinoflagellates, *Symbiodinium minutum*, and *Karlodinium veneficum*.

	<i>Lepidodinium chlorophorum</i>	Strain TGD	Strain MGD	<i>Symbiodinium minutum</i>	<i>Karlodinium veneficum</i>
A→U	0	0	1	+	+
A→G	118	88	15	+++	+++
A→C	0	0	0	+	+
U→A	0	1	0	–	+
U→G	0	0	0	+	+
U→C	59	82	0	++	++
G→A	5	0	1	+	+
G→U	0	0	0	–	–
G→C	5	0	0	+	+
C→A	0	0	0	–	+
C→U	1	6	1	++	+
C→G	0	0	0	–	+
Total	188	177	18	389	1,087

For *S. minutum* and *K. veneficum*, the frequency of each base-conversion type is shown by symbols as follows: +++, ≥50%; ++, ≥10%; +, <10%; –, not detected. The corresponding data were coopted from **Tables 1, 2** in Klinger et al. (2018). For the three peDinoflagellates, the number of each editing type is shown.

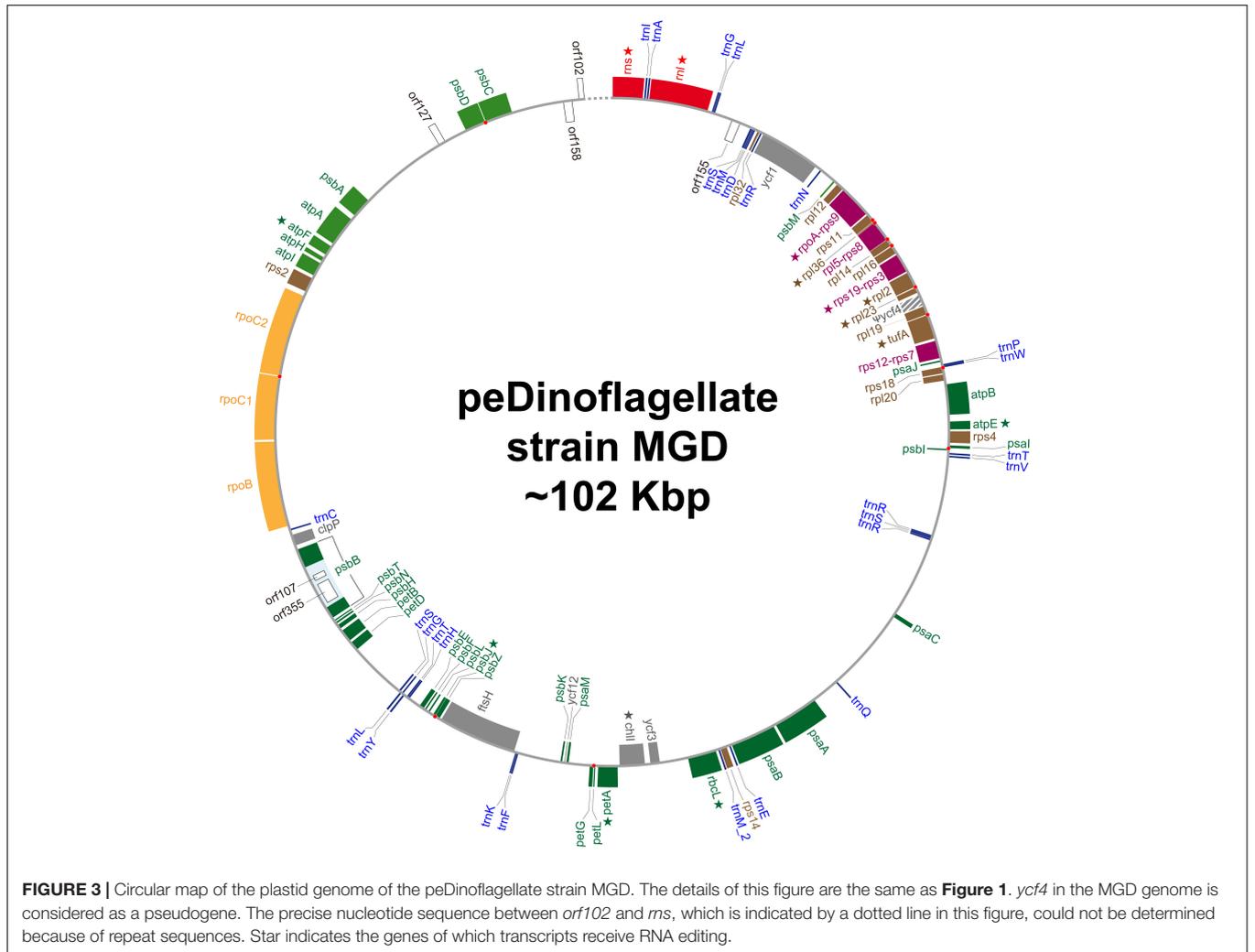
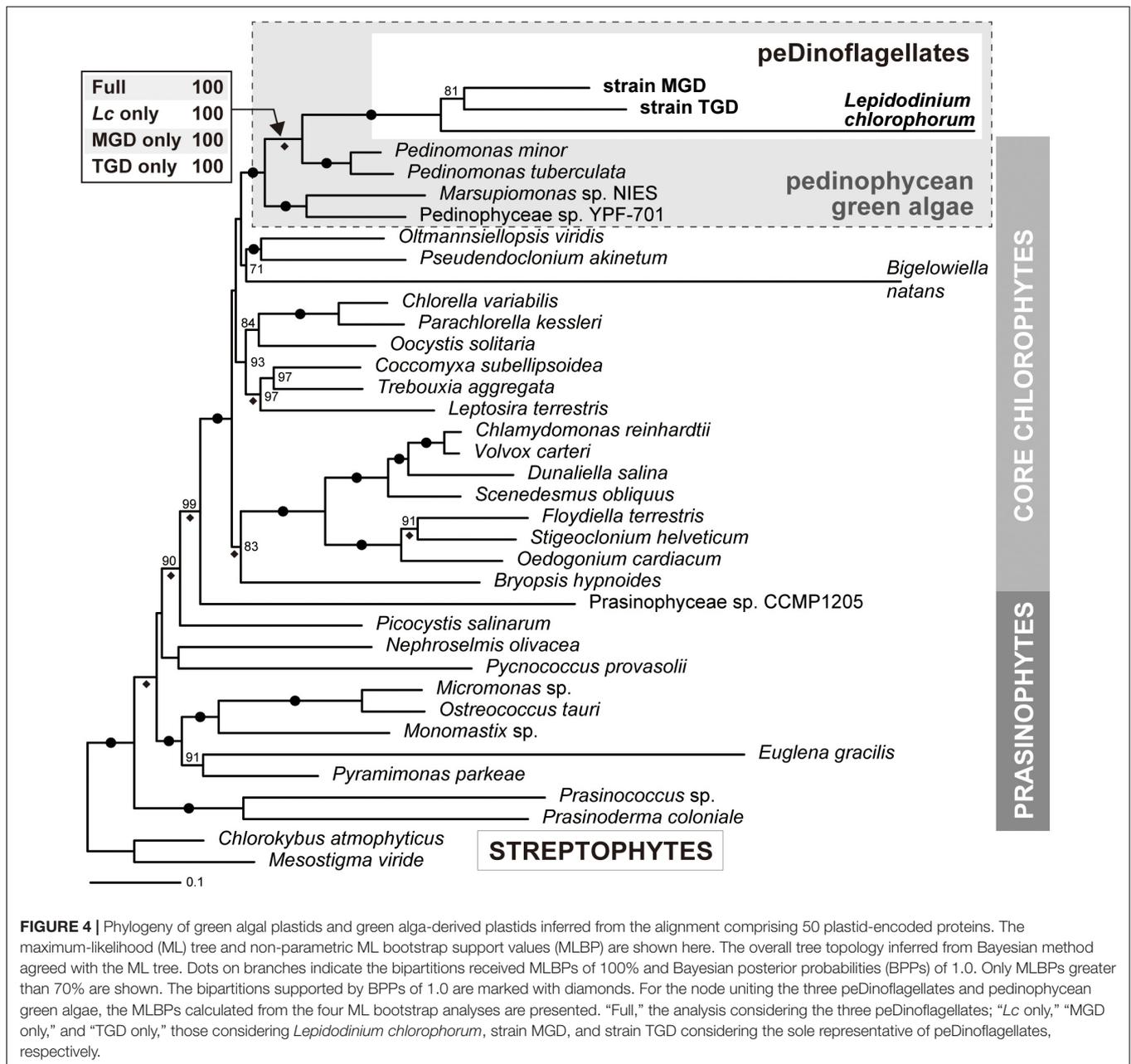
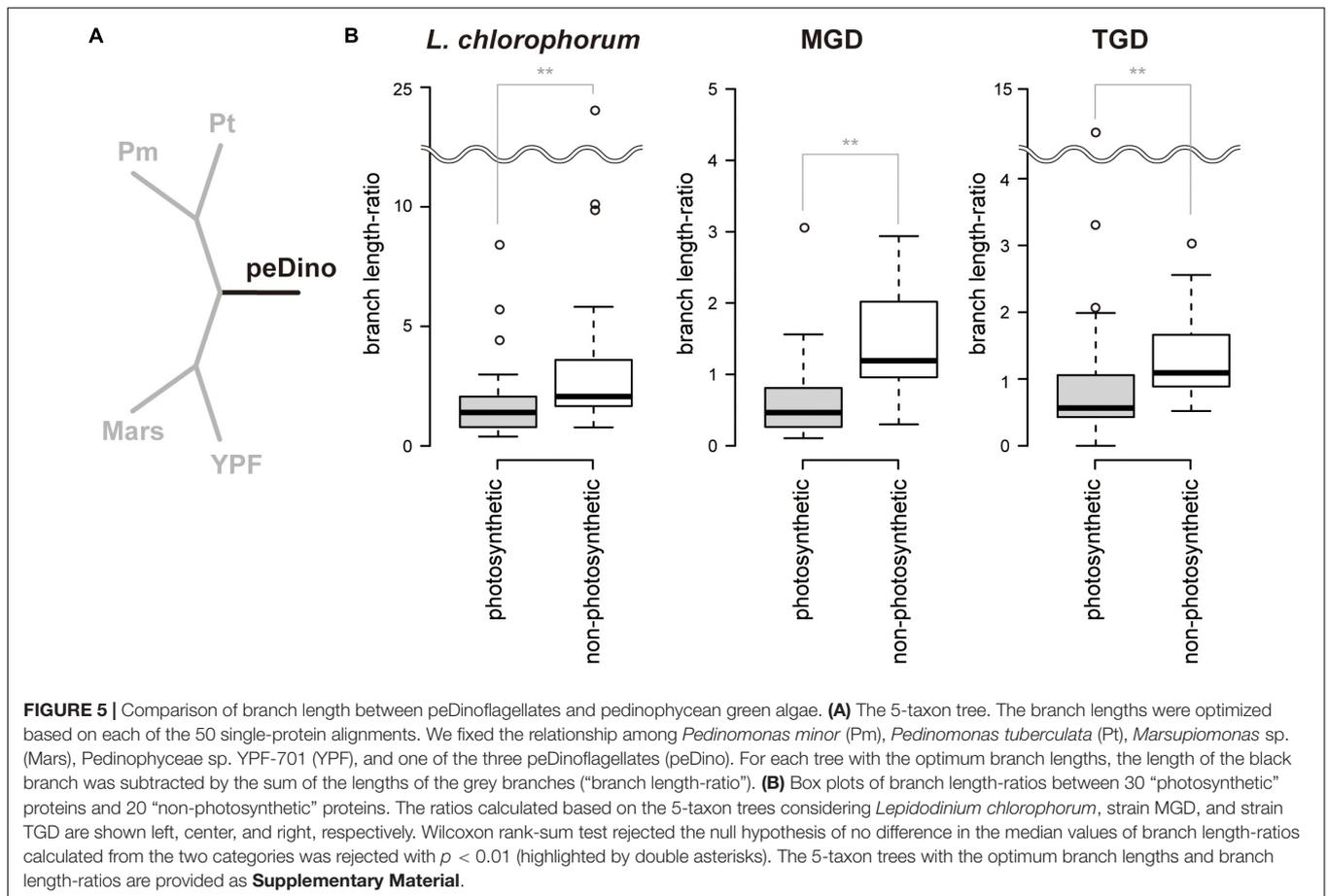


FIGURE 3 | Circular map of the plastid genome of the peDinoflagellate strain MGD. The details of this figure are the same as **Figure 1**. *ycf4* in the MGD genome is considered as a pseudogene. The precise nucleotide sequence between *orf102* and *ms*, which is indicated by a dotted line in this figure, could not be determined because of repeat sequences. Star indicates the genes of which transcripts receive RNA editing.



(Figures 2, 3, and Supplementary Figure 3), although the four pedinophycean plastid genomes bear IRs (Table 1). (ii) We identified more than 10 cases of ORF/gene overlapping/fusion in the peDinoflagellate plastid genomes (Table 1). In contrast, only a few cases of ORF overlapping and no ORF fusion were detected in the four pedinophycean plastid genomes. (iii) The peDinoflagellate plastid genomes appeared to carry fewer numbers of ORFs (i.e., 61–68) than 73–75 ORFs found in those of the four pedinophycean green algae (Table 1). The ORF repertoires of the three peDinoflagellate plastid genomes appeared to be similar to each other but the TGD and MGD plastid genomes retain 6–8 ORFs that are absent from the *L. chlorophorum* plastid genome (Figure 5). (iv) The

peDinoflagellate plastid genomes commonly use a deviant genetic code, in which AUA is assigned as methionine, not as isoleucine as the standard genetic code (Table 1). (v) Certain levels of pseudogenization seemingly operated on the peDinoflagellate plastid genomes. A single non-functional ORF was found in the individual peDinoflagellate plastid genomes, albeit none has been reported in the pedinophycean plastid genomes. (vi) Base conversion type RNA editing appeared to occur to the plastid gene transcripts in the three peDinoflagellates (see below). Importantly, we found no evidence for RNA editing on the transcripts from the *P. minor* plastid genome. Finally, (vii) the significant difference in substitution rate between the plastid-encoded proteins in “photosynthetic” category and those



in “non-photosynthetic” category appeared to be shared among the three peDinoflagellates (Figure 5B).

Patterns and Frequencies in Base Conversion RNA Editing on the Transcripts From the Three peDinoflagellate Plastid Genomes

Base conversion editing on plastid gene transcripts was observed in the three peDinoflagellates (the details of the sites received base conversion editing are summarized in a spreadsheet in **Supplementary Data**). However, the pattern and frequency of RNA editing appeared to vary among the three peDinoflagellate plastids. We observed similar numbers of editing positions in the *L. chlorophorum* and TGD plastids (188 and 177). In the two peDinoflagellate plastids, more than 90% of the identified editings converted A to G or U to C (Table 2). In the *L. chlorophorum* plastid, three minor types of base conversion, namely conversion from G to A, that from G to C, and that from C to U, were observed. We observed a single case of the conversion from U to A and 6 cases of the conversion of C to U in the TGD plastid.

The trend of the editing in the MGD plastid appeared to be distinct from those in the *L. chlorophorum* and TGD plastids described above in two aspects. First, only 18 editing positions were detected in the MGD plastid (Table 2). Second,

TABLE 3 | Frequencies of base conversion-type RNA editing at the codon positions in the three peDinoflagellates.

	<i>Lepidodinium chlorophorum</i>	Strain TGD	Strain MGD
Codon first position	86	81	10
Codon second position	60	80	2
Codon third position	5	12	0
Regions overlapped by two ORFs	3	2	1
Transfer and ribosomal RNAs	34	3	5

the conversion from A to G occupied 15 out of the 18 observed editing positions but no case of the conversion from U to C was found (Table 2). The rest of the editings converted A to U, G to A, and C to U.

The majority of the base conversion editing in the transcripts from protein-coding regions in the three peDinoflagellate plastids were found at codon first and second positions. In *L. chlorophorum*, 86, 60, and five cases of the editing were found at codon first, second, and third positions, respectively, after exclusion of those occurred in tRNAs, rRNAs, and the regions overlapped by two ORFs (Table 3). A similar strong bias of the editing toward codon first/second positions over third positions was observed in the TGD and MGD plastids (Table 3).

A single editing position in the *petA* transcript appeared to be shared between the *L. chlorophorum* and TGD plastids. A-to-G conversions occurred at the homologous positions in the *L. chlorophorum* and TGD *petA* transcripts (**Supplementary Figure 5**). These editings introduced the amino acid change from lysine to arginine and that from aspartic acid to glycine in the *L. chlorophorum* and TGD proteins, respectively. All editings except the case described above were found to occur at unique positions across the three peDinoflagellate plastids.

DISCUSSION

Reconfirmation of the Pedinophycean Origin of the *L. chlorophorum*, MGD, and TGD Plastids

The plastid phylogeny inferred from plastid small subunit rRNA genes demonstrated the pedinophycean green algal origin of the *L. chlorophorum*, TGD, and MGD plastids (Sarai et al., 2020). In the current study, we strengthened the pedinophycean green algal origin of the three peDinoflagellate plastids by analyzing the 50-protein alignment (**Figure 4**). The *L. chlorophorum*, TGD, and MGD plastids were grouped together and connected to the *P. minor* and *P. tuberculata* plastids with full statistical support. At the face value, the 50-protein phylogeny indicates that the three dinoflagellates took up the same pedinophycean green alga as the endosymbionts. However, we had a suspicion of the three peDinoflagellate plastids being grouped together artifactually and misplaced in the tree of green algal plastids, as the proteins encoded in the three peDinoflagellate plastids appeared to evolve much more rapidly than the orthologs in the green algal plastids.

To examine the above-mentioned possibility, we reanalyzed the 50-protein alignment after excluding two out of the three peDinoflagellate plastids alternatively. Significantly, the 50-protein phylogeny constantly grouped the *Pedinomonas* plastids and one of the three peDinoflagellate plastids together, arguing against their intimate phylogenetic affinity being an artifact in the tree reconstruction. The conclusion deduced from the 50-protein analyses agrees well with the discussion in Sarai et al. (2020)—*L. chlorophorum*, strains MGD, and strain TGD separately transformed green algal endosymbionts belonging to the genus *Pedinomonas* or those closely related to *Pedinomonas*. Nevertheless, there is a large possibility for the endosymbionts in the three peDinoflagellates being closely related but different species belonging to the genus *Pedinomonas*, as the genuine diversity of *Pedinomonas* (and their close relatives) is underrepresented by *P. minor* and *P. tuberculata*, for which plastid genome data are currently available. The phylogenetic relationship among the three peDinoflagellate plastids should be reexamined after we obtain the plastid genome data from the species that sufficiently represent the diversity of pedinophycean green algae in the natural environment.

In order to address why the pedinophycean endosymbiosis was repeated in distantly related branches in the tree of dinoflagellates, we need to accumulate both environmental and physiological data of pedinophycean green algae and

peDinoflagellates. First, the three peDinoflagellates are of marine and most likely took up marine pedinophycean algae as the endosymbionts. Nevertheless, *P. minor* and *P. tuberculata* were isolated originally from fresh water and soil environments, respectively. Thus, future studies may explore the diversity and distribution of *Pedinomonas* and their close relatives in the marine environment. Secondly, the repeated pedinophycean endosymbiosis imply the potential merits for the host cells to bear the pedinophycean-derived plastids in their natural habitats. Thus, it is necessary to examine the key factors for the emergence of peDinoflagellates in the marine environments and the physiological characteristics that differentiate peDinoflagellates from other dinoflagellate species.

Unique Features Shared Among the Three peDinoflagellate Plastid Genomes: Parallel Gain or Vertical Inheritance?

The comparative plastid genomics identified the features that are common in the three peDinoflagellate plastids but absent in the pedinophycean plastids (see section “Results”). We here propose that some of the above-mentioned features were achieved in the three peDinoflagellate plastids in parallel.

Secondary Loss of Inverted Repeats

The plastid genomes lacking IRs have been documented in diverse land plants (Palmer and Thompson, 1981, 1982; Guisinger et al., 2011; Wu et al., 2011; Li et al., 2016; Ruhlman et al., 2017; Choi et al., 2019; Cauz-Santos et al., 2020; Jin et al., 2020), green algae (Turmel et al., 2009a; Cai et al., 2017), haptophytes (Baurain et al., 2010; Gabrielsen et al., 2011), euglenids (Hallick et al., 1993; Gockel and Hachtel, 2000; Karnkowska et al., 2018), and cryptophytes (Donaher et al., 2009; Tanifuji et al., 2020). As the four pedinophycean plastid genomes determined to date bear IRs, we proposed that the plastid genomes in the pedinophycean endosymbionts taken up by *L. chlorophorum*, strain MGD, and strain TGD used to have IRs but the parallel losses of one of the two copies occurred during serial secondary endosymbioses.

Reduced Open Reading Frame Repertory, Pseudogenization, and Difference in Evolutionary Tempo Between Photosynthetic and Non-photosynthetic Proteins

Uthanumallian et al. (2022) recently demonstrated the parallel reduction of ORF repertory in the plastid genomes in chlorarachniophytes, euglenids, and *L. chlorophorum*, all of which established their current plastids throughout the reductions of green algal endosymbionts. By expanding the discussion in the pioneering work, we propose the parallel reduction of ORF repertoires in the *L. chlorophorum*, TGD, and MGD plastid genomes. The convergence of ORF repertoires among the three peDinoflagellate plastid genomes (**Figure 6**) can be reconciled if the pressure for discarding the genes encoding the proteins not involved in the core plastid functions, such as photosynthesis, translation, and transcription (Uthanumallian et al., 2022), was common across serial secondary endosymbioses. Pseudogenization can be regarded as

codon, the complex interplay of the change in codon usage and the evolution of tRNAs is required and only a few cases of deviant genetic codes in plastid genomes have been reported prior to this study (Matsumoto et al., 2011; Su et al., 2019; Ceriotti et al., 2021). Considering the number of the plastid genomes completely sequenced to date (6,661 genomes labeled as “apicoplast,” “chloroplast,” or “cyanelle” in the GenBank Genome database as of December 2021), we can regard the reassignment of a codon (or codons) as rare events in the evolution of plastid genomes. We here propose that as-yet-unstudied pedinophycean species possess the plastid genomes, in which AUA codon is assigned as Met, and were the origins of the current peDinoflagellate plastids. In other words, we anticipate the AUA assignment in the plastid genome as the probe to pinpoint the origins of the pedinophycean green algae that gave rise to the three peDinoflagellate plastids.

RNA Editing

The original works reported the four pedinophycean plastid genomes did not examine the presence/absence of RNA editing experimentally (Turmel et al., 2009b; Jackson et al., 2018), most likely because the plastid gene sequences and corresponding amino acid sequences lacked any sign of RNA editing. Indeed, the mapping of RNA-seq reads on the plastid genome found no incongruence between the genome and transcript sequences in the *P. minor* plastid by our criterion (see above). Thus, until a future study finds a clear case of RNA editing in an as-yet-unstudied pedinophycean plastid, we here propose that (i) the pedinophycean plastids are primarily free from RNA editing and (ii) the three peDinoflagellates configured and started executing the RNA editing on the plastid gene transcripts in parallel.

One may argue that the machinery required for RNA editing was too complex to evolve *de novo* in the three separate occasions in the evolution of dinoflagellates. It is worthy to note that RNA editing has been documented in peridinin plastids of diverse dinoflagellates, suggesting that this molecular trait can be traced back to an early dinoflagellate species bearing peridinin plastid (Zauner et al., 2004; Dang and Green, 2009; Dorrell and Howe, 2012; Mungpakdee et al., 2014; Klinger et al., 2018). Thus, the ancestral species, which gave rise to *L. chlorophorum*, strain MGD, and strain TGD, used to bear peridinin plastids and most likely operated the RNA editing on plastid gene transcripts. We propose that *L. chlorophorum* and strain TGD transplanted the RNA editing machinery, which originally worked in peridinin plastids, to the plastids derived from the pedinophycean endosymbionts. The principal reason for the above proposal is the similarity between the pattern of base conversion editing between *L. chlorophorum*/TGD and peridinin dinoflagellates. In both peridinin and two peDinoflagellate plastids, base conversion occurred in diverse directions, but the cases of A-to-G and U-to-C conversions appeared to predominate over other types of base conversion (Table 2). The above proposal is not totally unexpected, as the putative co-option of the RNA editing machinery in peridinin plastid was proposed for the non-canonical plastids derived from the haptophyte endosymbiont in

the ancestral karenian dinoflagellate for the same reasoning (Dorrell and Howe, 2012; Jackson et al., 2013; Klinger et al., 2018).

The pattern of the base conversion editing distinguishes the MGD plastid from other dinoflagellate plastids including those of *L. chlorophorum* and strain TGD. Due to the absence of U-to-C conversion in the 18 editing positions identified in the MGD plastid (Table 2), we have no solid ground to assume that base conversion editing in the MGD plastid is performed by the machinery that existed prior to serial secondary endosymbiosis. One possibility is that strain MGD has retained the RNA editing machinery beyond serial secondary endosymbiosis but discarded the molecular components that were required for U-to-C conversion.

We here explore why the pattern and frequency of base conversion editing are different between strain MGD and *L. chlorophorum*/strain TGD, besides the potential difference in the RNA editing machinery (see above). Intriguingly, the plastid-encoded RNA polymerases in both *L. chlorophorum* and strain TGD, of which plastid gene transcripts receive base conversion editing at higher frequencies than those of strain MGD, are potentially deficient. Both *rpoC1* in *L. chlorophorum* and *rpoB* in strain TGD can be regarded as non-functional (Supplementary Figure 4 and Figure 2). In addition, *RpoC2* in strain TGD are encoded in two separate ORFs (Figure 2) and it is difficult to exclude absolutely the possibility of this unusual *RpoC2* being dysfunctional/non-functional. In contrast, the genes encoding RNA polymerase subunits in the MGD plastid genome seem to be intact (Figure 3). Thus, the potential deficiency in the plastid-encoded RNA polymerase might connect with the frequency of base conversion editing on the corresponding plastid gene transcripts, albeit we can provide no molecular background for the above hypothesis. Curiously, *Karldinium veneficum*, of which *rpoB* and *rpoC2* were found to be disrupted by frameshifts, operates base conversion editing on plastid gene transcripts at a much higher frequency than *L. chlorophorum* or strain TGD (Gabrielsen et al., 2011; Klinger et al., 2018). There is a large room for the “connection” between the deficiency in plastid-encoded RNA polymerase and frequency of base conversion editing being coincident but worthy to being revisited when the data of genomes and RNA editing are accumulated from additional non-canonical-types of dinoflagellate plastids.

Finally, we found a single case of base insertion editing in the *psaA* transcript in strain TGD (Supplementary Figure 1). This type of RNA editing has not been documented in any dinoflagellate plastids, suggesting that strain TGD has developed this type of RNA editing after serial secondary endosymbiosis. There are three possibilities for the origin of the putative molecular machinery for base insertion editing found in the TGD plastid. First, strain TGD invented the machinery *de novo*. Second, strain TGD laterally acquired the machinery from a distantly related organism. Third, strain TGD modified the machinery for base conversion editing to operating base insertion editing. To evaluate the three scenarios described above, we first need to identify the enzymes that operate the base insertion in the *psaA* transcript in the TGD plastid.

CONCLUSION

In the current study, we reported the plastid genomes of peDinoflagellate strains MGD and TGD and unveiled the plastid genome features, many of which were predicted to have emerged separately after serial secondary endosymbioses involved in pedinophycean green algae. Among the features found in the peDinoflagellate plastid genomes, RNA editing is intriguing. The plastid gene transcripts of *L. chlorophorum* and strain TGD appeared to share the pattern of base conversion editing with those of peridinin dinoflagellates, suggesting that the RNA editing machinery was inherited in the two peDinoflagellates beyond serial secondary endosymbioses. On the other hand, we could provide no plausible idea of how strain MGD established the RNA editing in the plastid, as the pattern of base conversion editing on the plastid gene transcripts was distinct between strain MGD and other dinoflagellates including *L. chlorophorum* and strain TGD.

DATA AVAILABILITY STATEMENT

The nucleotide sequences of the MGD and TGD plastid genomes are available under the GenBank/DDBJ/EMBL accession numbers LC716140 and LC716139, respectively.

AUTHOR CONTRIBUTIONS

CS, KT, and MI conducted the experiments related to the algal cultures. EM and KM conducted the molecular biological

experiments required for sequencing the plastid genomes. EM, KM, and TN reconstructed the plastid genome sequences and identified RNA editing by mapping RNA-seq reads. EM, EY, and YI prepared alignments and carried out phylogenetic analyses. EM, KM, TN, MI, and YI wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the grants from the Japan Society for Promotion of Science awarded to YI (numbers 18KK0203 and 19H03280) and TN (number 20H03305).

ACKNOWLEDGMENTS

The phylogenetic analyses conducted in this work have been carried out under the “Interdisciplinary Computational Science Program” in the Center for Computational Sciences, University of Tsukuba. Computations in this study were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.918543/full#supplementary-material>

REFERENCES

- Archibald, J. M. (2009). The puzzle of plastid evolution. *Curr. Biol.* 19, R81–R88. doi: 10.1016/j.cub.2008.11.067
- Barbrook, A. C., Howe, C. J., and Nisbet, R. E. R. (2019). Breaking up is hard to do: the complexity of the dinoflagellate chloroplast genome. *Perspect. Phycol.* 6, 31–37. doi: 10.1127/pip/2018/0084
- Baurain, D., Brinkmann, H., Petersen, J., Rodriguez-Ezpeleta, N., Stechmann, A., Demoulin, V., et al. (2010). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol. Biol. Evol.* 27, 1698–1709. doi: 10.1093/molbev/msq059
- Bentlage, B., Rogers, T. S., Bachvaroff, T. R., and Delwiche, C. F. (2016). Complex ancestries of isoprenoid synthesis in dinoflagellates. *J. Eukaryot. Microbiol.* 63, 123–137. doi: 10.1111/jeu.12261
- Bjørnland, T., Haxo, F. T., and Liaaen-Jensen, S. (2003). Carotenoids of the Florida red tide dinoflagellate *Karenia brevis*. *Biochem. Syst. Ecol.* 31, 1147–1162. doi: 10.1016/S0305-1978(03)00044-9
- Boisvert, S., Lavolette, F., and Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* 17, 1519–1533. doi: 10.1089/cmb.2009.0238
- Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S., and Keeling, P. J. (2014). Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot. Cell* 13, 246–255. doi: 10.1128/EC.00299-13
- Cai, C., Wang, L., Zhou, L., He, P., and Jiao, B. (2017). Complete chloroplast genome of green tide algae *Ulva flexuosa* (Ulvophyceae, Chlorophyta) with comparative analysis. *PLoS One* 12:e0184196. doi: 10.1371/journal.pone.0184196
- Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., et al. (2019). Access to RNA-sequencing data from 1,173 plant species: the 1000 Plant transcriptomes initiative (1KP). *GigaScience* 8:giz126. doi: 10.1093/gigascience/giz126
- Carty, S., and Parrow, M. W. (2015). “Dinoflagellates,” in *Freshwater Algae of North America*, ed. J. D. Wehr (Amsterdam: Elsevier), 773–807.
- Cauz-Santos, L. A., da Costa, Z. P., Callot, C., Cauet, S., Zucchi, M. I., Bergès, H., et al. (2020). A repertoire of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. *Genome Biol. Evol.* 12, 1841–1857. doi: 10.1093/gbe/evaa155
- Ceriotti, L. F., Roulet, M. E., and Sanchez-Puerta, M. V. (2021). Plastomes in the holoparasitic family Balanophoraceae: extremely high AT content, severe gene content reduction, and two independent genetic code changes. *Mol. Phylogenet. Evol.* 162:107208. doi: 10.1016/j.ympev.2021.107208
- Chan, P. P., and Lowe, T. M. (2019). “tRNAscan-SE: searching for tRNA genes in genomic sequences,” in *Gene Prediction, Methods in Molecular Biology*, ed. M. Kollmar (New York, NY: Springer), 1–14.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Choi, I.-S., Jansen, R., and Ruhlman, T. (2019). Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biol. Evol.* 11, 1321–1333. doi: 10.1093/gbe/evz076
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. doi: 10.1093/gigascience/giab008
- Dang, Y., and Green, B. R. (2009). Substitutional editing of *Heterocapsa triquetra* chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA. *Gene* 442, 73–80. doi: 10.1016/j.gene.2009.04.006
- 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 nonphotosynthetic alga *Cryptomonas paramecium*: reduction, compaction, and

- accelerated evolutionary rate. *Genome Biol. Evol.* 1, 439–448. doi: 10.1093/gbe/evp047
- Dorrell, R. G., and Howe, C. J. (2012). Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18879–18884. doi: 10.1073/pnas.1212270109
- Gabrielsen, T. M., Minge, M. A., Espelund, M., Tooming-Klunderud, A., Patil, V., Nederbragt, A. J., et al. (2011). Genome evolution of a tertiary dinoflagellate plastid. *PLoS One* 6:e19132. doi: 10.1371/journal.pone.0019132
- Gockel, G., and Hachtel, W. (2000). Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist* 151, 347–351. doi: 10.1078/S1434-4610(04)70033-4
- Guisinger, M. M., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the Angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* 28, 583–600. doi: 10.1093/molbev/msq229
- Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., et al. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21, 3537–3544. doi: 10.1093/nar/21.15.3537
- Hao, W., Liu, G., Wang, W., Shen, W., Zhao, Y., Sun, J., et al. (2021). RNA editing and its roles in plant organelles. *Front. Genet.* 12:757109. doi: 10.3389/fgene.2021.757109
- Jackson, C., Knoll, A. H., Chan, C. X., and Verbruggen, H. (2018). Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Sci. Rep.* 8:1523. doi: 10.1038/s41598-017-18805-w
- Jackson, C. J., Gornik, S. G., and Waller, R. F. (2013). A tertiary plastid gains RNA editing in its new host. *Mol. Biol. Evol.* 30, 788–792. doi: 10.1093/molbev/mss270
- Jeffrey, S. W., Sielicki, M., and Haxo, F. T. (1975). Chloroplast pigment patterns in dinoflagellates. *J. Phycol.* 11, 374–384. doi: 10.1111/j.1529-8817.1975.tb02799.x
- Jin, D.-M., Wicke, S., Gan, L., Yang, J.-B., Jin, J.-J., and Yi, T.-S. (2020). The loss of the inverted repeat in the Putranjivoid clade of Malpighiales. *Front. Plant Sci.* 11:942. doi: 10.3389/fpls.2020.00942
- Kamikawa, R., Tanifuji, G., Kawachi, M., Miyashita, H., Hashimoto, T., and Inagaki, Y. (2015). Plastid genome-based phylogeny pinpointed the origin of the green-colored plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Genome Biol. Evol.* 7, 1133–1140. doi: 10.1093/gbe/evv060
- Karnkowska, A., Bennett, M. S., and Triemer, R. E. (2018). Dynamic evolution of inverted repeats in Euglenophyta plastid genomes. *Sci. Rep.* 8:16071. doi: 10.1038/s41598-018-34457-w
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 729–748. doi: 10.1098/rstb.2009.0103
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4
- Klinger, C. M., Paoli, L., Newby, R. J., Wang, M. Y.-W., Carroll, H. D., Leblond, J. D., et al. (2018). Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. *Genome Biol. Evol.* 10, 1019–1038. doi: 10.1093/gbe/evy057
- Kodama, Y., Shumway, M., Leinonen, R., and International Nucleotide Sequence Database Collaboration (2012). The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40, D54–D56. doi: 10.1093/nar/gkr854
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* 7:S4. doi: 10.1186/1471-2148-7-S1-S4
- Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109. doi: 10.1093/molbev/msh112
- Lartillot, N., and Philippe, H. (2006). Computing bayes factors using thermodynamic integration. *Syst. Biol.* 55, 195–207. doi: 10.1080/10635150500433722
- Lemieux, C., Otis, C., and Turmel, M. (2014). Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol. Biol.* 14:211. doi: 10.1186/s12862-014-0211-2
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, J., Gao, L., Chen, S., Tao, K., Su, Y., and Wang, T. (2016). Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of Sciadopityaceae. *Sci. Rep.* 6:20934. doi: 10.1038/srep20934
- Matsumoto, T., Ishikawa, S. A., Hashimoto, T., and Inagaki, Y. (2011). A deviant genetic code in the green alga-derived plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Mol. Phylogenet. Evol.* 60, 68–72. doi: 10.1016/j.ympev.2011.04.010
- Matsumoto, T., Kawachi, M., Miyashita, H., and Inagaki, Y. (2012). Prasinonxanthin is absent in the green-colored dinoflagellate *Lepidodinium chlorophorum* strain NIES-1868: pigment composition and 18S rRNA phylogeny. *J. Plant Res.* 125, 705–711. doi: 10.1007/s10265-012-0486-6
- Matsuo, E., and Inagaki, Y. (2018). Patterns in evolutionary origins of heme, chlorophyll *a* and isopentenyl diphosphate biosynthetic pathways suggest non-photosynthetic periods prior to plastid replacements in dinoflagellates. *PeerJ* 6:e5345. doi: 10.7717/peerj.5345
- Minge, M. A., Shalchian-Tabrizi, K., Tørresen, O. K., Takishita, K., Probert, I., Inagaki, Y., et al. (2010). A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evol. Biol.* 10:191. doi: 10.1186/1471-2148-10-191
- Mühlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H., and Kollmar, M. (2016). A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Res.* 26, 945–955. doi: 10.1101/gr.200931.115
- Mungpakdee, S., Shinzato, C., Takeuchi, T., Kawashima, T., Koyanagi, R., Hisata, K., et al. (2014). Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome. *Genome Biol. Evol.* 6, 1408–1422. doi: 10.1093/gbe/evu109
- Nakayama, T., Takahashi, K., Kamikawa, R., Iwataki, M., Inagaki, Y., and Tanifuji, G. (2020). Putative genome features of relic green alga-derived nuclei in dinoflagellates and future perspectives as model organisms. *Commun. Integr. Biol.* 13, 84–88. doi: 10.1080/19420889.2020.1776568
- 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
- Nosenko, T., Lidie, K. L., Van Dolah, F. M., Lindquist, E., Cheng, J.-F., US Department of Energy–Joint Genome Institute, et al. (2006). Chimeric plastid proteome in the Florida “red tide” dinoflagellate *Karenia brevis*. *Mol. Biol. Evol.* 23, 2026–2038. doi: 10.1093/molbev/msl074
- One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2
- Osawa, S., and Jukes, T. H. (1989). Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* 28, 271–278. doi: 10.1007/BF02103422
- Palmer, J. D., and Thompson, W. F. (1981). Rearrangements in the chloroplast genomes of mung bean and pea. *Proc. Natl. Acad. Sci. U.S.A.* 78, 5533–5537. doi: 10.1073/pnas.78.9.5533
- Palmer, J. D., and Thompson, W. F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29, 537–550. doi: 10.1016/0092-8674(82)90170-2
- Prijbelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Curr. Protoc. Bioinform.* 70:102. doi: 10.1002/cpbi.102
- Ruhlman, T. A., Zhang, J., Blazier, J. C., Sabir, J. S. M., and Jansen, R. K. (2017). Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *Am. J. Bot.* 104, 559–572. doi: 10.3732/ajb.1600453
- Saldarriaga, J. F., Taylor, F. J. R., Keeling, P. J., and Cavalier-Smith, T. (2001). Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J. Mol. Evol.* 53, 204–213. doi: 10.1007/s002390010210

- Santos, M. A. S., Cheesman, C., Costa, V., Moradas-Ferreira, P., and Tuite, M. F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp. *Mol. Microbiol.* 31, 937–947. doi: 10.1046/j.1365-2958.1999.01233.x
- Sarai, C., Tanifuji, G., Nakayama, T., Kamikawa, R., Takahashi, K., Yazaki, E., et al. (2020). Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellenogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 117, 5364–5375. doi: 10.1073/pnas.1911884117
- Schultz, D. W., and Yarus, M. (1994). Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.* 235, 1377–1380. doi: 10.1006/jmbi.1994.1094
- Sengupta, S., and Higgs, P. G. (2005). A unified model of codon reassignment in alternative genetic codes. *Genetics* 170, 831–840. doi: 10.1534/genetics.104.037887
- Shalchian-Tabrizi, K., Minge, M. A., Cavalier-Smith, T., Nedreklepp, J. M., Klaveness, D., and Jakobsen, K. S. (2006). Combined heat shock protein 90 and ribosomal RNA sequence phylogeny supports multiple replacements of dinoflagellate plastids. *J. Eukaryot. Microbiol.* 53, 217–224. doi: 10.1111/j.1550-7408.2006.00098.x
- Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell. Mol. Life Sci.* 63, 698–708. doi: 10.1007/s00018-005-5449-9
- Sibbald, S. J., and Archibald, J. M. (2020). Genomic insights into plastid evolution. *Genome Biol. Evol.* 12, 978–990. doi: 10.1093/gbe/evaa096
- Su, H.-J., Barkman, T. J., Hao, W., Jones, S. S., Naumann, J., Skippington, E., et al. (2019). Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 934–943. doi: 10.1073/pnas.1816822116
- Takano, Y., Hansen, G., Fujita, D., and Horiguchi, T. (2008). Serial replacement of diatom endosymbionts in two freshwater dinoflagellates, *Peridiniopsis* spp. (Peridinales, Dinophyceae). *Phycologia* 47, 41–53. doi: 10.2216/07-36.1
- 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 genomic responses to loss of photosynthesis. *Genome Biol. Evol.* 12, 3926–3937. doi: 10.1093/gbe/evaa001
- Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F., et al. (2000). Phylogenetic analyses indicate that the 19' hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol.* 17, 718–729. doi: 10.1093/oxfordjournals.molbev.a026350
- Turmel, M., Gagnon, M.-C., O'Kelly, C. J., Otis, C., and Lemieux, C. (2009a). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* 26, 631–648. doi: 10.1093/molbev/msn285
- Turmel, M., Otis, C., and Lemieux, C. (2009b). The chloroplast genomes of the green algae *Pedinomonas minor*, *Parachlorella kessleri*, and *Oocystis solitaria* reveal a shared ancestry between the Pedinomonadales and Chlorellales. *Mol. Biol. Evol.* 26, 2317–2331. doi: 10.1093/molbev/msp138
- Uthanumallian, K., Iha, C., Repetti, S. I., Chan, C. X., Bhattacharya, D., Duchene, S., et al. (2022). Tightly constrained genome reduction and relaxation of purifying selection during secondary plastid endosymbiosis. *Mol. Biol. Evol.* 39:msab295. doi: 10.1093/molbev/msab295
- Wang, H.-C., Minh, B. Q., Susko, E., and Roger, A. J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235. doi: 10.1093/sysbio/syx068
- Watanabe, M. M., Suda, S., Inouya, I., Sawaguchi, T., and Chihara, M. (1990). *Lepidodinium viride* gen. et sp. nov. (Gymnodinaiales, Dinophyta), a green dinoflagellate with a chlorophyll *a*- and *b*-containing endosymbiont. *J. Phycol.* 26, 741–751. doi: 10.1111/j.0022-3646.1990.00741.x
- Watanabe, M. M., Takeda, Y., Sasa, T., Inouye, I., Suda, S., Sawaguchi, T., et al. (1987). A green dinoflagellate with chlorophylls *a* and *b*: morphology, fine structure of the chloroplast and chlorophyll composition. *J. Phycol.* 23, 382–389. doi: 10.1111/j.1529-8817.1987.tb04148.x
- Wu, C.-S., Wang, Y.-N., Hsu, C.-Y., Lin, C.-P., and Chaw, S.-M. (2011). Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of Gymnosperm phylogeny. *Genome Biol. Evol.* 3, 1284–1295. doi: 10.1093/gbe/evr095
- Yamada, N., Sym, S. D., and Horiguchi, T. (2017). Identification of highly divergent diatom-derived chloroplasts in dinoflagellates, including a description of *Durinskia kwazulunatalensis* sp. nov. (Peridinales, Dinophyceae). *Mol. Biol. Evol.* 34, 1335–1351. doi: 10.1093/molbev/msx054
- Zapata, M., Fraga, S., Rodriguez, F., and Garrido, J. (2012). Pigment-based chloroplast types in dinoflagellates. *Mar. Ecol. Prog. Ser.* 465, 33–52. doi: 10.3354/meps09879
- Zauner, S., Greilinger, D., Laatsch, T., Kowallik, K. V., and Maier, U.-G. (2004). Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate *Ceratium horridum*. *FEBS Lett.* 577, 535–538. doi: 10.1016/j.febslet.2004.10.060
- Zhang, Z., Green, B. R., and Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400, 155–159. doi: 10.1038/22099

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Matsuo, Morita, Nakayama, Yazaki, Sarai, Takahashi, Iwataki and Inagaki. 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.