Comparative Plastid Genomics of *Cryptomonas* Species Reveals Fine-Scale Genomic Responses to Loss of Photosynthesis

Goro Tanifuji (ID) [1,*], Ryoma Kamikawa[2], Christa E. Moore[3], Tyler Mills[3], Naoko T. Onodera[3], Yuichiro Kashiyama (ID) [4], John M. Archibald (ID) [3], Yuji Inagaki[5,6], and Tetsuo Hashimoto[6]

[1]Department of Zoology, National Museum of Nature and Science, Ibaraki, Japan

[2]Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan

[3]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

[4]Department of Applied Chemistry and Food Science, Fukui University of Technology, Fukui, Japan

[5]Center for Computational Sciences, University of Tsukuba, Ibaraki, Japan

[6]Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan

*Corresponding author: E-mail: gorot@kahaku.go.jp.

Abstract

Loss of photosynthesis is a recurring theme in eukaryotic evolution. In organisms that have lost the ability to photosynthesize, nonphotosynthetic plastids are retained because they play essential roles in processes other than photosynthesis. The unicellular algal genus *Cryptomonas* contains both photosynthetic and nonphotosynthetic members, the latter having lost the ability to photosynthesize on at least three separate occasions. To elucidate the evolutionary processes underlying the loss of photosynthesis, we sequenced the plastid genomes of two nonphotosynthetic strains, *Cryptomonas* sp. CCAC1634B and SAG977-2f, as well as the genome of the phototroph *Cryptomonas curvata* CCAP979/52. These three genome sequences were compared with the previously sequenced plastid genome of the nonphotosynthetic species *Cryptomonas paramecium* CCAP977/2a as well as photosynthetic members of the Cryptomonadales, including *C. curvata* FBCC300012D. Intraspecies comparison between the two *C. curvata* strains showed that although their genome structures are stable, the substitution rates of their genes are relatively high. Although most photosynthesis-related genes, such as the *psa* and *psb* gene families, were found to have disappeared from the nonphotosynthetic strains, at least ten pseudogenes are retained in SAG977-2f. Although gene order is roughly shared among the plastid genomes of photosynthetic Cryptomonadales, genome rearrangements are seen more frequently in the smaller genomes of the nonphotosynthetic strains. Intriguingly, the light-independent protochlorophyllide reductase comprising *chlB*, *L*, and *N* is retained in nonphotosynthetic SAG977-2f and CCAC1634B. On the other hand, whereas CCAP977/2a retains ribulose-1,5-bisphosphate carboxylase/oxygenase-related genes, including *rbcL*, *rbcS*, and *cbbX*, the plastid genomes of the other two nonphotosynthetic strains have lost the ribulose-1,5-bisphosphate carboxylase/oxygenase protein-coding genes.

Key words: nonphotosynthetic plastid, *Cryptomonas*, loss of photosynthesis, genome reduction.

Introduction

Eukaryotic photosynthesis has evolved on multiple occasions through endosymbioses between nonphotosynthetic eukaryotes and photosynthetic partners. The progenitor of the primary plastid of Archaeplastida was very clearly a cyanobacterial endosymbiont, and subsequent eukaryote–eukaryote endosymbioses involving both red and green algal endosymbionts gave rise to various extant photosynthetic eukaryotes (Zimorski et al. 2014; Archibald 2015; Nowack and Weber 2018).

Plastids are the eukaryotic organelles that house the photosynthetic machinery that generates organic carbon in plant and algal cells. However, plastids are also the location of various other biochemical processes other than photosynthesis. Some plants and protists have secondarily lost the ability to photosynthesize but still retain a plastid, which generally retains a genome (a notable exception is the colorless green alga *Polytomella*) (Smith and Lee 2014; Hadariová et al. 2018). Such nonphotosynthetic plastids are known to maintain essential functions such as the biosynthesis of fatty acids, isoprenoids, and amino acids. The genomes of nonphotosynthetic plastids tend to still encode some essential proteins for plastid metabolism, despite reductions in size and the number of protein-coding genes. The protein-coding gene repertoire of the plastid genomes of nonphotosynthetic organisms varies, as does the suite of biological functions taking place in these derived organelles. For example, the plastid genomes of nonphotosynthetic *Cryptomonas paramecium* (Cryptophyceae), *Euglena longa* (Euglenophyceae), and *Aneura mirabilis* (parasitic plant) contain the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Gockel and Hachtel 2000; Wickett et al. 2008; Donaher et al. 2009), although other lineages, such as *Nitzschia* sp. (a diatom), *Plasmodium* spp. (Apicomplexa), *Epifagus virginiana* (a parasitic plant), and chrysophytes (Stramenopila) have lost RuBisCO and proteins related to its function (Wolfe et al. 1992; Wilson et al. 1996; Kamikawa, Tanifuji, et al. 2015; Dorrell et al. 2019). The retention of ATP synthase in some nonphotosynthetic plastids such as *Nitzschia* sp. (a diatom), *C. paramecium* (Cryptista), and several parasitic plants is also intriguing, but the exact roles of the ATP synthase complex in nonphotosynthetic plastids are debated (Wickett et al. 2008; Donaher et al. 2009; Kamikawa, Tanifuji, et al. 2015; Suzuki et al. 2018).

Cryptista is a unicellular eukaryotic group comprising plastid-lacking (i.e., Cyathomonadacea, Kathablepharidacea, and *Palpitomonas*) and plastid-bearing (i.e., Cryptomonadales) organisms (Yabuki et al. 2014; Hoef-Emden and Archibald 2016; Tanifuji and Onodera 2017; Cenci et al. 2018; Adl et al. 2019) (fig.1). Cryptomonadales acquired their photosynthetic capabilities by engulfing a red algal endosymbiont and retain the endosymbiont-derived plastid and nucleus (the latter called the nucleomorph) (Moore and Archibald 2009; Tanifuji and Archibald 2014; Tanifuji and Onodera 2017). Although nucleomorph genomes possess fewer than 500 protein-coding genes, only a small proportion of them are associated with plastid functions (31 and 18 in photosynthetic and nonphotosynthetic strains, respectively; Moore and Archibald 2009; Tanifuji and Archibald 2014; Tanifuji and Onodera 2017); the majority of plastid proteins in Cryptomonadales are encoded in the nuclear and plastid genomes (Curtis et al. 2012; Hempel et al. 2014; Gould et al. 2015). To date, eight cryptomonad plastid genomes have been sequenced
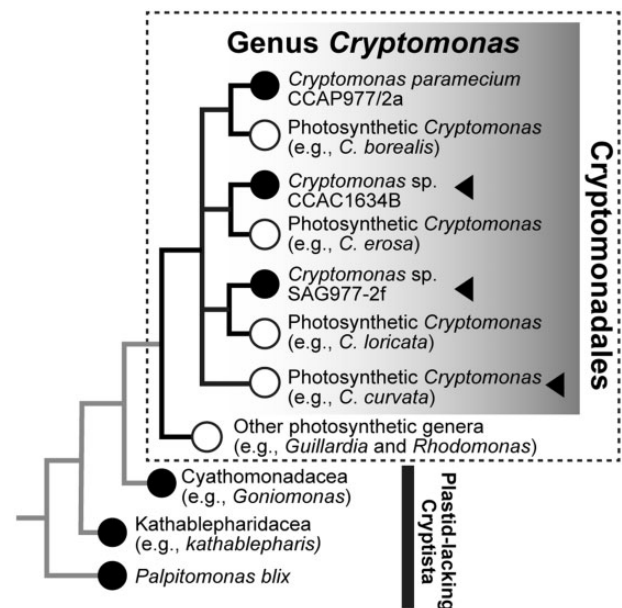

Fig. 1.—Schematic tree of Cryptista. Tree topology and classification were adopted from Adl et al. (2019), Hoef-Emden (2005, 2007), Hoef-Emden et al. (2005), and Yabuki et al. (2014). White and black circles indicate photosynthetic and nonphotosynthetic organisms, respectively. The plastid genomes sequenced in this study are shown by black arrowheads. The independent origins of nonphotosynthetic *Cryptomonas* were robustly suggested by the phylogenetic analyses based on nuclear internal transcribed spacer 2, nuclear LSU rRNA, and nucleomorph SSU rRNA sequences (Hoef-Emden 2005, 2007; Hoef-Emden et al. 2005).

(Douglas and Penny 1999; Khan et al. 2007; Donaher et al. 2009; Kim et al. 2015, 2017). Within the genus *Cryptomonas*, the plastid genomes of photosynthetic species *C. curvata* FBCC300012D and the secondarily nonphotosynthetic *C. paramecium* CCAP977/2a were sequenced (Donaher et al. 2009; Kim et al. 2017). However, phylogenetic surveys have revealed that three nonphotosynthetic *Cryptomonas* lineages are closely related to different photosynthetic species (Hoef-Emden 2005, 2007; Hoef-Emden et al. 2005), suggesting that members of the genus *Cryptomonas* have lost the ability to photosynthesize on at least three separate occasions. This provides an opportunity to study the dynamics of plastid genome structure and coding capacity as they relate to the loss of photosynthesis over short evolutionary timescales. Unfortunately, genomic sampling is presently sparse; only one plastid genome from a nonphotosynthetic *Cryptomonas* species has been sequenced, that of *C. paramecium* CCAP977/2a (Donaher et al. 2009).

To rectify this situation, we sequenced two plastid genomes of nonphotosynthetic *Cryptomonas* species, *Cryptomonas* sp. SAG977-2f and CCAC1634B, which independently lost the ability to photosynthesize and are distantly related to *C. paramecium* CCAP977/2a. We also sequenced the plastid genome of the photosynthetic species *C. curvata* CCAP979/52 and compared it to that of another strain,

FBCC300012D, to provide the first insight into intraspecies diversity of plastid genomes in *Cryptomonas*. Comparative analyses of the plastid genomes of photosynthetic and non-photosynthetic members of the genus *Cryptomonas* provide snapshots of the loss of photosynthesis, including the pseudogenization of photosynthesis-related genes and DNA elimination steps. Our results also support the idea that genome rearrangements are accelerated by the loss of inverted repeats (IRs) of ribosomal RNA operons (e.g., Palmer and Thompson 1982; Strauss et al. 1988; Wei et al. 2013; Vieira Ldo et al. 2014). Furthermore, unexpected variability in the repertoires of genes for photosynthetic functions was observed among the plastid genomes of nonphotosynthetic *Cryptomonas*.

Materials and Methods

Cell Culturing, DNA Extraction, and Genome Sequencing

Cryptomonas curvata CCAP979/52, *Cryptomonas* sp. SAG977-2f, and *Cryptomonas* sp. CCAC1634B were obtained from the Culture Collection of Algae and Protozoa (CCAP), the Sammlung von Algenkulturen der Universität Göttingen (SAG) and the Culture Collection of Algae at the University of Cologne (CCAC), respectively. *Cryptomonas curvata* CCAP979/52 was cultured in MWC medium under 12:12 light/dark conditions at 21 °C (Guillard and Lorenzen 1972). *Cryptomonas* sp. CCAC1634B was maintained in Warris-H medium (McFadden and Melkonian 1986) with 1% volume of bacteria standard medium under dark conditions at 18 °C. DNA extractions for *C. curvata* CCAP979/52 and *Cryptomonas* sp. CCAC1634B were done using the standard SDS–phenol/chloroform extraction method. For *C. curvata* CCAP979/52, total DNA was separated into nuclear and organelle DNA-enriched fractions (i.e., plastid, mitochondrion, and nucleomorph) by Hoechst dye (Sigma-Aldrich)–cesium chloride density gradient centrifugation at 35,000 rpm for 65 h at 4 °C, as described previously (Tanifuji et al. 2011, 2014; Moore et al. 2012). An organelle DNA-enriched fraction of *C. curvata* CCAP979/52 was subjected to sequencing library construction using the Nextera XT DNA Library Preparation Kit (Illumina), and DNA sequencing was carried out using a Mi-Seq instrument (Illumina). Because *Cryptomonas* sp. SAG977-2f did not grow well in the laboratory, cells received directly from SAG were immediately subjected to genome amplification using the REPLI-g Mini Kit (Qiagen) following the manufacturer's instructions. The amplified genomic DNA of *Cryptomonas* sp. SAG977-2f and the total DNA of *Cryptomonas* sp. CCAC1634B were sent to Hokkaido System Science (Hokkaido, Japan) for TruSeq library construction and Illumina HiSeq sequencing.

Genome Assembly

For *C. curvata* CCAP979/52, a data set containing 7.75 million paired-end reads 300 base-pairs (bp)-long were obtained.

After sequence quality check using FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc, last accessed January 21, 2020), positions with average quality scores <30 were trimmed down and 271 (forward) and 221 (reverse)-base reads were retained. Paired-end reads (101 base-long; 88.75 million and 154.46 million) were obtained for *Cryptomonas* sp. SAG977-2f and *Cryptomonas* sp. CCAC1634B, respectively. The first and last four bases of raw reads were trimmed due to their low sequence quality. The remaining 94-bp reads were subjected to additional quality control steps. Trimmed sequence reads in which >80% of the bases had quality scores of >20 were extracted using FASTAX-Toolkit (Ver. 0.0.14) (http://hannonlab.cshl.edu/fastx_toolkit/, last accessed January 21, 2020). Preliminary assembly was carried out using SPAdes ver. 3.5.0 (Bankevich et al. 2012) with the remaining reads for each strain. The putative plastid genome sequences were retrieved from the resulting scaffolds by BlastN similarity searches using the plastid genome sequences of *C. paramecium* CCAP977/2a and *Guillardia theta* as queries (GenBank accession numbers: GQ358203 and AF041468 [Douglas and Penny 1999; Donaher et al. 2009]). Because the read depth coverages of the plastid genome scaffolds in the preliminary assemblies were "too high" for optimal assembly (>1,200×) in *Cryptomonas* sp. CCAC1634B, 10% of the raw reads were subjected to a second-round assembly (Tanifuji et al. 2014). The CCAC1634B plastid genome contig that resulted from this second-round assembly was nearly identical to that of the first round, differing only at the boundary regions. The assembly of data from *Cryptomonas* sp. SAG977-2f was performed using the whole raw reads with the single cell mode of SPAdes (Bankevich et al. 2012). Putative plastid genome scaffolds were then assembled into two and one scaffolds for *Cryptomonas* sp. SAG977-2f and *Cryptomonas* sp. CCAC1634B, respectively. The gaps between scaffolds were filled by PCR and Sanger sequencing.

Genome Annotation and Comparative Analyses

Protein and tRNA gene predictions were performed by MFANNOT (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl, last accessed January 21, 2020) and tRNAscan-SE (Lowe and Eddy 1997). The ribosomal RNA operons were predicted by BlastN searches using known plastid RNA operon sequences as queries. Individual gene models were confirmed manually using Artemis software (Rutherford et al. 2000), and unannotated nucleotide sequences were subjected to BlastX searches against the NCBI nr database to find putative pseudogenes and protein-coding regions with weak similarity to known proteins/genes. Genome statistics such as GC content were also generated using Artemis.

In the *C. curvata* CCAP979/52 genome, a putative *psbN*-associated group II intron was predicted by MFANNOT and

MFOLD (Zuker 2003), and the full intron structure and associated sequence elements were manually curated. This included identification of a reverse transcriptase coding region located 230-bp upstream of "domain V," the splicing motif (5'-GUGYG…AY-3') consensus sequence, and the *psbN* open reading frame itself. The predicted secondary structure of the intron was concordant with known group II introns in other Cryptomonadales (Khan and Archibald 2008 ; Kim et al. 2017).

Genome synteny analyses were carried out using the Mauve multiple genome aligner (Darling et al. 2004). Detection and analysis of sequence repeats was carried out using the mreps software (Kolpakov et al. 2003).

The plastid genome sequences and gene annotations determined in this study were deposited under DDBJ accession numbers LC484192 (*C. curvata* CCAP979/52), LC484193 (*Cryptomonas* sp. CCAC1634B), and LC484194 (*Cryptomonas* sp. SAG977-2f).

Synonymous and Nonsynonymous Substitution Rates Test

For estimation of synonymous (d_S) and nonsynonymous (d_N) substitution rates in the *C. curvata* CCAC979/52 and FBCC300012D plastid genomes, 146 genes from each genome were individually aligned by mafft (Katoh et al. 2005). Codon alignments were generated by pal2Nal.pl (Suyama et al. 2006) with the "no gap" option. The alignments were then concatenated resulting in a supermatrix containing 96,171 bp. To collect the *rbcL* gene sequences of Cryptomonadales, the *rbcL* sequence of *C. curvata* CCAP979/52 was used as a blast query against the NCBI nt database. Seventy-one *rbcL* sequences of Cryptomonadales were obtained and aligned by mafft (Katoh et al. 2005). Short sequences and gap sites were removed and a multiple alignment consisting of 64 Cryptomonadales sequences and 852 bp was generated. Sequences from redundant and unidentified species were removed. During this process, *Cryptomonas*, *Rhodomonas*, and *Chroomonas* were retained for the purposes of interspecific comparison. Substitution rates for all data sets were estimated using the codeml programs of PAML4.8 with the F3 x 4 codon model (Yang, 2007).

Results and Discussion

Intraspecific Plastid Genome Diversity in *C. curvata*

We determined the plastid genome sequence of the photosynthetic species *C. curvata* CCAP979/52 and compared it with that of the previously sequenced *C. curvata* FBCC300012D (fig. 2 and table 1). Together these two genomes serve as important reference points for investigating the loss of photosynthesis in *Cryptomonas* spp. In terms of structure and coding capacity, the two *C. curvata* plastid genomes are generally similar to those of other photosynthetic Cryptomonadales (table 1) (Douglas and Penny 1999;

Khan et al. 2007; Kim et al. 2015, 2017). The plastid genome of the CCAP979/52 strain is 1,689-bp longer and has a lower GC content compared with FBCC300012D (i.e., 129,974 and 128,285 bp in size and 34.6% and 35.3% GC content in the CCAP979/52 and FBCC300012D strains, respectively) (table 1; Kim et al. 2017). The number of predicted protein-coding genes in *C. curvata* CCAP979/52 and FBCC300012D is 148 and 146, respectively (table 1 and supplementary table S1, Supplementary Material online). The total amount of noncoding DNA in the CCAP979/52 plastid genome is 19,325 bp, 1,117 bp less than that of strain FBCC300012D (table 1). Two genes, *orf264* (annotated as *ycf55* in *Chroomonas*) and *ycf20*, both showing significant sequence similarity to homologs in *Chroomonas mesostigmatica*, were detected in the plastid genome of the CCAP979/52 strain. However, the FBCC300012D plastid genome encodes only partial protein sequences of *orf264* (*ycf55*) and *ycf20*, from which the N-terminal regions, including a viable initiation methionine, are missing. Although the *ycf20* protein is encoded in the plastid genomes of *Guillardia*, *Storeatula*, and *Chroomonas*, it is missing in *Rhodomonas* and *Teleaulax* (Douglas and Penny 1999; Khan et al. 2007; Kim et al. 2015, 2017). Therefore, these two genes appear to be dispensable for plastid function and cell viability in Cryptomonadales. The gene order, including the position of a *psbN* intron, is identical between the two *C. curvata* plastid genomes, with the exception of the presence/absence of *orf264* and *ycf20* (figs. 2 and 3) (see below for more discussion).

Pairwise comparison between the two *C. curvata* plastid genomes revealed an overall sequence identity level of 79% with 27,843 substitutions and 3,929 bp of gap sequences. The sequence identity of rRNA regions between the two *C. curvata* strains was determined to be 1,468/1,490 bp (99%) in the small subunit and 2,736/2,825 bp (97%) in the large subunit. The observed intraspecific genetic diversity seen in *C. curvata* is higher than that typically seen in other algae. For example, an analysis of 13 strains of the green alga *Ostreococcus tauri* revealed only 314 single nucleotide polymorphisms sites (15–156 substitutions against consensus sequence) in ~72 kb of plastid genome sequence (Blanc-Mathieu et al. 2013). To obtain further insight into the genetic diversity between the two *C. curvata* strains, we measured synonymous (d_S) and nonsynonymous (d_N) substitution rates in a pairwise comparison of 146 plastid protein-coding genes as well as those in a concatenated data set (supplementary table S2, Supplementary Material online). The d_S and d_N values between the two *C. curvata* strains are 7.4467 (d_S) and 0.0716 (d_N) in an average of the genes and 2.6798 (d_S) and 0.0609 (d_N) in the concatenated data set. These substitution rates are higher than that observed in a multistrain examination of another Cryptomonadales species, *Hemiselmis andersenii* ($d_S = 0.0017$, $d_N = 0.0007$) (Grisdale et al. 2019).

Hoef-Emden (2005) and Hoef-Emden et al. (2005) hypothesized that an accelerated evolutionary rate in a
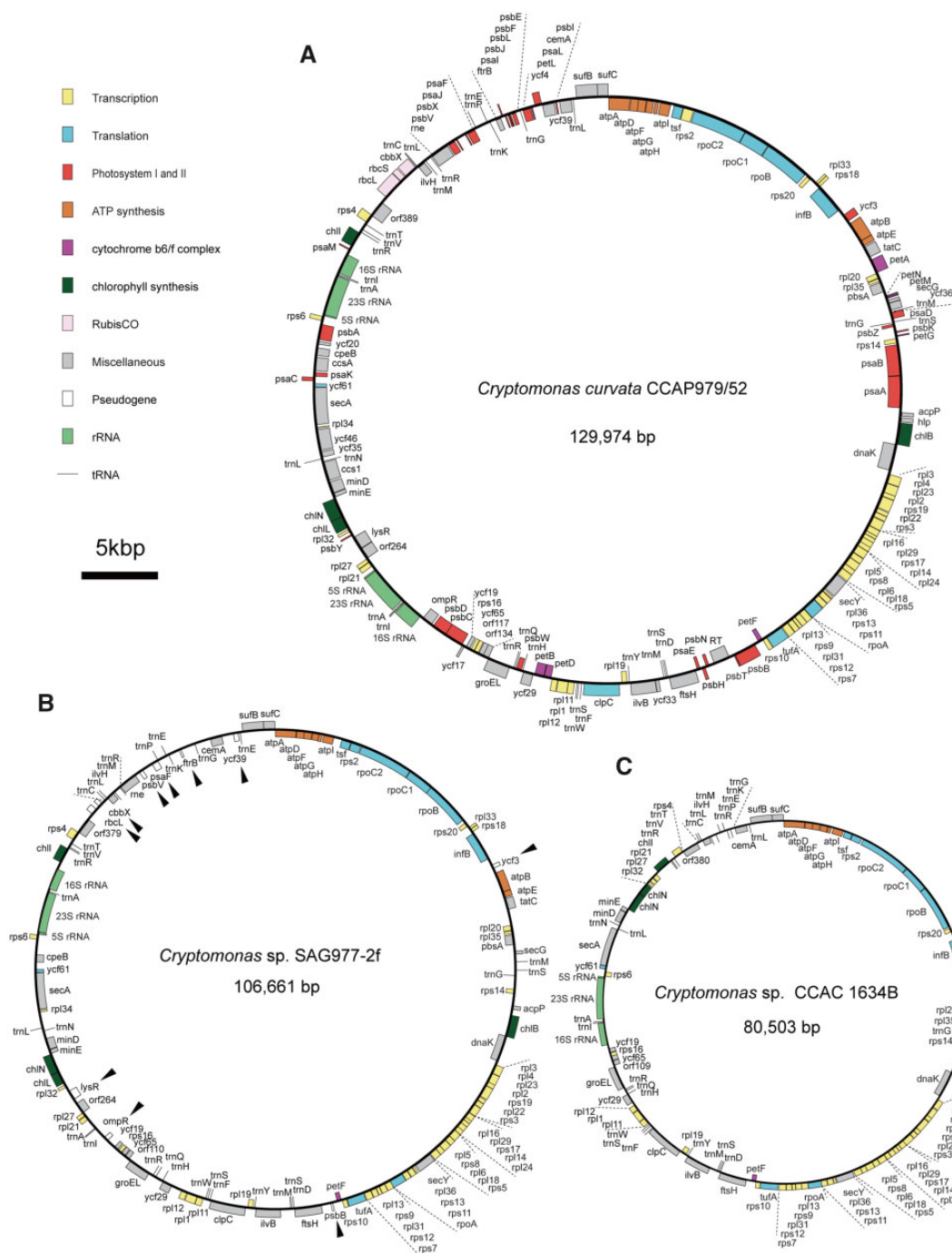
Fig. 2.—Circular physical maps of the plastid genome of (*A*) *Cryptomonas curvata* CCAP979/52, (*B*) *Cryptomonas* sp. SAG977-2f, and (*C*) *Cryptomonas* sp. CCAP1634B. Genes shown on the outside the circle are transcribed clockwise. Annotated genes are colored according to the functional categories shown in the center. Recognizable pseudogenes are emphasized by arrowheads.

photosynthetic ancestor of *Cryptomonas* triggered loss of photosynthesis in some present-day strains (Hoef-Emden 2005; Hoef-Emden et al. 2005). Indeed, evolutionary rates of the plastid *rbcL* and nucleomorph SSU rRNA genes in particular *Cryptomonas* species, including both nonphotosynthetic and photosynthetic species (designated as "long

branching"), were found to be accelerated relative to other *Cryptomonas* species (Hoef-Emden 2005; Hoef-Emden et al. 2005). That said, genetic diversity between different genera within Cryptomonadales has not been investigated. Here, we measured d_S and d_N rates of available *rbcL* genes in Cryptomonadales (supplementary table S3, Supplementary

Table 1:

Overview of Cryptomonadales Plastid Genomes

	The Photosynthetic Plastid Genomes in Other Genera of Cryptomonadales[a]	*Cryptomonas curvata* FBCC300012D[a]	*Cryptomonas curvata* CCAP979/52	*Cryptomonas* sp. SAG977-2f	*Cryptomonas* sp. CCAC1634B	*Cryptomonas paramecium* CCAP977/2a[a]
Genome size (bp)	121,524–140,953	128,285	129,974	106,661	80,503	77,717
No. of protein genes	143–149	147	148	90	84	82
No. of tRNAs	30–32	31	31	30	29	29
GC contents (%)	32–36	35.3	34.59	34.6	39.7	38.1
Amount of noncoding DNA (%)[b]	14,792–27,796 bp (12.2–20.3%)	20,442 bp (15.9%)	19,325 bp (14.9%)	32,281 bp (30.2%)	10,704 bp (13.3%)	9,917 bp (12.8%)

[a]Data adopted from Donaher et al. (2009), Douglas and Penny (1999), Khan et al. (2007), and Kim et al. (2015, 2017).
[b]Pseudogenes were not considered.

Material online). The averages of d_S and d_N rates of *rbcL* genes among 12 *Cryptomonas* species were found to be 1.5286 (d_S) and 0.0134 (d_N) (0.53 and 0.0045 between the two *C. curvata* strains), which is higher than that seen for six *Rhodomonas* species (0.30 [d_S] and 0.0052 [d_N]) and six *Hemiselmis* species (0.17 [d_S] and 0.0095 [d_N]) (supplementary table S2, Supplementary Material online). Thus, in addition to the species within the long-branching lineage (Hoef-Emden 2005; Hoef-Emden et al. 2005), our data suggest that the evolutionary rate across the whole *Cryptomonas* genus is higher than that of other genera.

Common Features of Plastid Genomes in Nonphotosynthetic *Cryptomonas* Species

Complete circular mapping plastid genome sequences were successfully obtained for two nonphotosynthetic *Cryptomonas* species. The newly sequenced plastid genomes of *Cryptomonas* sp. SAG977-2f and *Cryptomonas* sp. CCAC1634B were estimated to be 106,661 and 80,503 bp in size, respectively. Other features, including GC content and amount of noncoding DNA, are shown in table 1. *Cryptomonas* sp. SAG977-2f and *Cryptomonas* sp. CCAC1634B contain 90 and 84 protein-coding genes in their plastid genomes, respectively (fig. 2, table 1, and supplementary table S1, Supplementary Material online). Relative to the photosynthetic Cryptomonadales, including those discussed above (Douglas and Penny 1999; Khan et al. 2007; Donaher et al. 2009; Kim et al. 2015, 2017), ~60 protein-coding genes are missing from the *Cryptomonas* sp. SAG977-2f and *Cryptomonas* sp. CCAC1634B genomes, including photosystems I and II (i.e., the *psa* and *psb* gene families) and the cytochrome b_6/f complex (i.e., the *pet* gene family) (table 1 and supplementary table S1, Supplementary Material online), or have become pseudogenized (see section below), as seen in the previously sequenced *C. paramecium* CCAP977/2a genome (Douglas and Penny 1999; Khan et al. 2007; Donaher et al. 2009; Kim et al. 2015, 2017). The loss of photosynthesis-related genes is reported in other

lineages of photosynthesis-lacking organisms (Hadariová et al. 2018). The number of tRNA gene copies was slightly reduced (by 1 or 2 genes) in the plastid genomes of nonphotosynthetic species relative to the photosynthetic *C. curvata* FBCC300012D (Kim et al. 2017). Although two copies of isoleucyl-tRNA with anticodon GAU (trnI [gau]) were found in the *C. curvata* CCAP977/52 and FBCC300012D plastid genomes, the nonphotosynthetic strains possessed a single copy of trnI (gau). In addition, alanyl-tRNA with anticodon UGC (trnA [ugc]) was present in a single copy in both the *Cryptomonas* sp. CCAC1634B and *C. paramecium* CCAP977/2a plastid genomes, whereas two strains of *C. curvata* and *Cryptomonas* sp. SAG977-2f contained two copies.

Out of ~90 proteins encoded in the plastid genomes of three nonphotosynthetic *Cryptomonas*, more than half were associated with housekeeping functions such as translation and transcription (e.g., ribosomal proteins, initiation factors, and RNA polymerase) (fig. 2 and supplementary table S1, Supplementary Material online). The second largest proportion of retained genes are those involved in the ATP synthase complex. Genes for eight subunits of the ATP synthase complex were found, namely, *atpA*, *atpB*, *atpD*, *atpE*, *atpF*, *atpG*, *atpH*, and *atpI*, although *atpF* in *C. paramecium* CCAP977/2a was found to be a pseudogene (Donaher et al. 2009). Many (but not all) secondarily nonphotosynthetic eukaryotes that are distantly related to Cryptista, such as certain diatoms and parasitic plants, retain plastid ATP synthase complex genes in their plastid genomes (Wicke et al. 2013; Kamikawa, Tanifuji, et al. 2015; Kamikawa et al. 2018; Suzuki et al. 2018), indicating a constraint against the loss of those genes from the plastid genome even after the loss of photosynthesis. Kamikawa, Tanifuji, et al. (2015) hypothesized that in the plastid of nonphotosynthetic organisms the ATP synthase acts to maintain a proton gradient between the thylakoid lumen and stroma, which is required for the twin arginine translocator (Tat) system (Kamikawa, Tanifuji, et al. 2015). In addition to ATP synthase complex genes, a core protein of the Tat system (TatC) was also maintained in the

plastid genomes of the three independently evolved nonphotosynthetic *Cryptomonas* lineages investigated here. Although nonphotosynthetic plastids generally lack conspicuous thylakoid structures, "thylakoid-like structures" were observed in nonphotosynthetic *Nitzschia* species (diatoms) (Kamikawa, Yubuki, et al. 2015), supporting the importance of, and evolutionary constraint for, retention of Tat system-mediated protein translocation into the thylakoid lumen in some nonphotosynthetic plastids. On the other hand, given that some nonphotosynthetic green algae, such as *Prototheca* and *Helicosporidium*, possess ATP synthase complex genes but lack the Tat system (Suzuki et al. 2018), the exact role(s) of the ATP synthase complex in nonphotosynthetic plastids is still unclear. As suggested previously (Kamikawa et al. 2017), the ATP synthase complex might also contribute to maintenance of a higher stromal pH so that certain plastid proteins can be activated under alkaline conditions. There may in fact be different reasons for the retention of the ATP synthase in plastid genomes of independently evolved nonphotosynthetic organisms.

We found genes encoding iron–sulfur cluster assembly proteins (*sufB* and *sufC*), acetolactate synthase subunits (*ilvB* and *ilvH*), and an acyl carrier protein (*acpP*) in the three plastid genomes of nonphotosynthetic *Cryptomonas* strains; this suggests that the functions of iron–sulfur cluster assembly, amino acid biosynthesis, and lipid metabolism are functional. In addition, genes for a few secretory proteins (*secA* and *secY*), the chaperonin *groEL*, a protease (*clpC*), ferredoxin (*petF*), and plastid membrane protein *cemA* were commonly retained. In summary, the plastid genomes that have independently become nonphotosynthetic have lost the protein-coding genes directly related to photosynthesis (i.e., photosystems I and II and cytochrome b_6/f complex), but retained part of the ATP synthase complex, protein translocators, iron–sulfur cluster assembly, amino acid synthesis, lipid metabolism, and genes involved in their expression and maintenance (e.g., housekeeping genes and chaperonins).

Genome Reduction Steps: Pseudogenization and DNA Elimination

Among the three nonphotosynthetic species examined, we found that the plastid genome of SAG977-2f is the largest, being ~26 kb larger than that of the other *Cryptomonas* spp. The main reason for this size variation is the larger intergenic spacers in the SAG977-2f genome. Whereas 12–20% of the plastid genomes in other Cryptomonadales are intergenic, regardless of whether or not they were photosynthetic, 30.2% of the SAG977-2f plastid genome is noncoding (table 1 and fig. 2). The total amount of intergenic DNA in the *Cryptomonas* sp. SAG977-2f, CCAC1634B, and *C. paramecium* CCAP977/2a plastid genomes is 32,281, 10,704, and 9, 917 bp (table 1), respectively. The *Cryptomonas* sp. SAG977-2f plastid genome contains 17

short (10–18 bp) repeats, similar to, or slightly lower than, the amount of obviously repetitive DNA in the plastid genomes of photosynthetic relatives (e.g., 10–20 and 18–23 bp repeats in *C. curvata* FBCC300012D and *G. theta*, respectively). Given that long stretches of repetitive sequence are conspicuously absent in the SAG977-2f plastid genome, it seems unlikely that their longer intergenic regions are the product of genome expansion.

A remarkable feature of the SAG977-2f plastid genome is the presence of pseudogenes. Although SAG977-2f has lost almost all of its plastid genes related to photosynthesis, as have strains CCAC1634B and CCAP977/2a, partial fragments homologous to at least ten genes (i.e., *lysR*, *ftrB*, *psaF*, *psbB*, *psbV*, *ompR*, *ycf3*, *ycf39*, *rbcL*, and *cbbX*) were identified in the plastid genome (fig. 2 and supplementary fig. S1, Supplementary Material online). The most striking example is *lysR*, where the 751-bp intergenic region between *orf264* and *rpl32* showed significant similarity to *lysR* of photosynthetic relatives. The pairwise alignment of the most significant high scoring pairs against the homolog of *C. curvata* CCAP979/52 is shown in supplementary figure S1, Supplementary Material online, which includes six in-frame stop codons, clearly indicating that this is a pseudogene of *lysR*. The alignment of another pseudogene, that of *ycf3*, is also shown in supplementary figure S1, Supplementary Material online. The intergenic region between *infB* and *atpB* showed similarity to *ycf3* of photosynthetic relatives, and two high scoring pairs showed positive alignments at 39/59 amino acids (63%) and 24/37 amino acids (65%) in blast searches, respectively (supplementary fig. S1, Supplementary Material online). These ten pseudogenes are clearly nonfunctional due to their fragmentation, the presence of in-frame stop codons, and/or the absence of an initiation methionine codon.

Pseudogenes in nonphotosynthetic plastid genomes have also been seen in other lineages. For example, the parasitic plants *Aneura mirabilis* and *Epifagus virginiana* possess pseudogenes of photosystem proteins in their plastid genomes (Wolfe et al. 1992; Wickett et al. 2008). Intriguingly, despite the pseudogenization of some photosystem I and II proteins, open reading frames for some other photosystem subunits appeared to be intact, suggesting that photosynthesis was lost in these organisms very recently (Wickett et al. 2008). The existence of multiple pseudogenes in SAG977-2f, combined with the lack of detectable pseudogene counterparts in the other nonphotosynthetic *Cryptomonas* plastid genomes, suggests that the loss of photosynthesis in SAG977-2f occurred more recently in the lineage to which SAG977-2f belongs compared with the other nonphotosynthetic *Cryptomonas*. Interestingly, the total length of the ten pseudogenes in the SAG977-2f plastid genome is 3,609 bp, which does not account for the ~20 kb of "extra" sequence compared with the plastid genomes of other nonphotosynthetic *Cryptomonas* (table 1). This suggests that at least some of
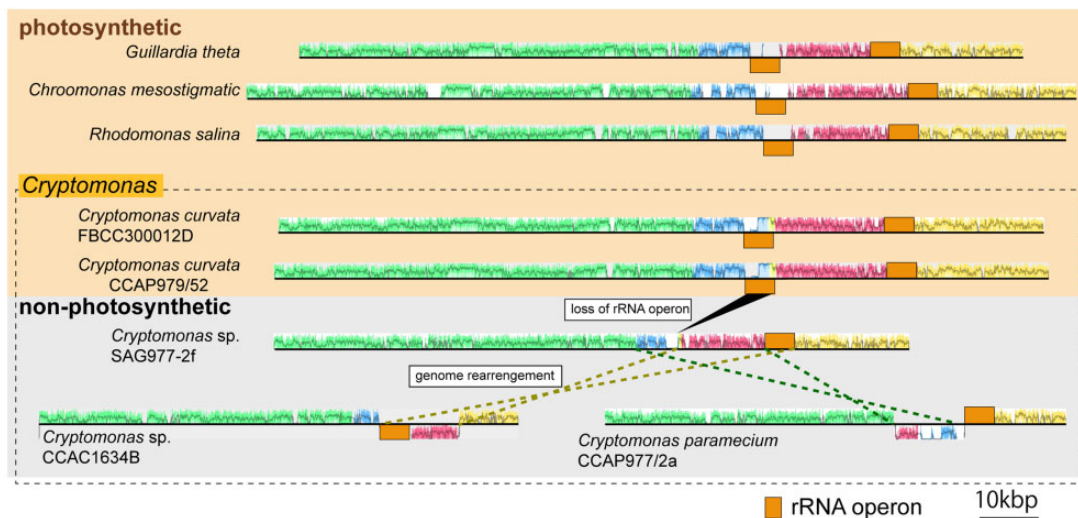
FIG. 3.—Alignment of plastid genomes of photosynthetic and nonphotosynthetic Cryptomonadales. Images were generated using the Mauve genome alignment tool (Darling et al. 2004). Green, blue, red, and yellow boxes indicate corresponding locally collinear blocks (LCBs), conserved segments among genomes identified by Mauve (Darling et al. 2004). The upper three lines correspond to the plastid genomes of three select photosynthetic species (*Guillardia theta* [Douglas and Penny 1999], *Chroomonas mesostigmatica* [Kim et al. 2017], and *Rhodomonas salina* [Khan et al. 2007]). The middle section shows *Cryptomonas curvata* FBCC300012D and CCAP9779/52. The lower lines indicate the genomes of *Cryptomonas* sp. SAG977-2f, *Cryptomonas* sp. CCAP1634B, and *Cryptomonas paramecium* CCAP977/2a. The rRNA operons, including SSU and LSU rRNA regions shown in orange boxes, were added manually. Dotted lines show rearranged regions in CCAC1634B (dark yellow) and CCAP977/2a (dark green), respectively.

these longer intergenic regions are in fact highly degenerate—and at present bioinformatically undetectable—pseudogenes that have yet to be purged from the genome. All things considered, given that the *Cryptomonas* strains considered here lost their photosynthetic capacities independently, this three-way comparison provides snapshots in the process of the loss of photosynthesis and accompanying genome reduction in eukaryotes.

The Loss of IRs of Ribosomal RNA Operons May Accelerate Genome Rearrangement

A notable structural difference between the plastid genomes of photosynthetic and nonphotosynthetic Cryptomonadales is the presence or absence of the IRs of ribosomal RNA operons. All of the photosynthetic Cryptomonadales plastid genomes from eight strains among six genera contain IRs (Douglas and Penny 1999; Khan et al. 2007; Kim et al. 2015, 2017). In contrast, one part of the IR disappeared from the plastid genomes of the nonphotosynthetic members (figs. 2 and 3). It is possible that the loss of IRs is related to the genome rearrangements observed in *C. paramecium* CCAP977/2a and *Cryptomonas* sp. CCAC1634B (fig. 3), given that the genomes of all photosynthetic Cryptomonadales, including *C. curvata* FBCC300012D and CCAC979/52, maintain genome synteny (i.e., conserved gene order), despite lineage-specific gene and intron losses (fig. 3) (Douglas and Penny 1999; Khan et al. 2007; Kim et al. 2015, 2017). In contrast, in the plastid genome of SAG977-2f, the gene order is the

same as in the photosynthetic plastid genomes, with the exception of simple gene losses (figs. 2 and 3). Thus, the apparent lack of genome rearrangements observed in the SAG977-2f plastid genome might also reflect the relatively recent loss of photosynthesis in the organism (fig. 3).

In this context, it is suggested that SAG977-2f is maintaining genome synteny, indicating that the loss of photosynthesis in *Cryptomonas* is not strictly associated with genome rearrangements. Although the three nonphotosynthetic *Cryptomonas*, namely, SAG977-2f, CCAC1634B, and CCAP977/2a, lost both photosynthesis and IRs, it is unclear whether the loss of photosynthesis and loss of the rRNA operon structure are always associated with each other in *Cryptomonas*. It is noteworthy that the loss of IRs has been seen in other lineages, such as photosynthetic green algae (Turmel et al. 2017). In contrast, all of the plastid genomes sequenced from nonphotosynthetic diatoms and a nonphotosynthetic chrysophyte retain IRs (Sabir et al. 2014; Kamikawa, Tanifuji, et al. 2015; Kamikawa et al. 2018; Dorrell et al. 2019). Many "apicoplast" genomes in apicomplexans are also extremely reduced but contain IRs (Arisue et al. 2012). These findings suggest that the loss of photosynthesis might be less commonly associated with the loss of an rRNA operon (Janouškovec et al. 2013; Turmel et al. 2017). More insight into the association between photosynthetic ability and loss of IRs in *Cryptomonas* will hopefully come from investigation of plastid genomes from organisms even more closely related to nonphotosynthetic *Cryptomonas* spp. than the *C. curvata* strains investigated herein.

The correlation between IR structure and frequency of genome rearrangements has been discussed elsewhere (e.g., Palmer and Thompson 1982; Strauss et al. 1988; Wei et al. 2013; Vieira Ldo et al. 2014). Strauss et al. (1988) hypothesized that IRs accelerate the frequency of homologous recombination between individual plastid genome molecules, stabilizing them against structural changes (Strauss et al. 1988). This idea has been discussed in the context of chlorarachniophytes, the other lineage of nucleomorph-bearing organisms (Tanifuji et al. 2016). In this case, a comparison of different chlorarachniophyte genera showed that whereas the IR-lacking mitochondrial genomes have undergone frequent rearrangements, the IR-containing plastid genomes exhibit a very high level of synteny (Tanifuji et al. 2016). Similar trends have been reported in green algae as well. For example, frequent genome rearrangements are observed in the IR-lacking plastid genomes of *Chloropicon* and *Chloroparbula* species (Turmel et al. 2019). The same pattern is seen here in Cryptomonadales. In this study, we have shown that genome rearrangements appear to be more frequent in plastid genomes without IRs (i.e., CCAP977/2a and CCAC1634B) than in IR-containing plastid genomes (fig. 3). Furthermore, genome rearrangements appear frequent in the mitochondrial genomes of Cryptomonadales (Tanifuji and Onodera 2017; Kim et al. 2018). All together, these findings speak to a correlation between the frequency of genome rearrangements and the presence/absence of IRs. Notably, because there is no reason why these phenomena would be limited to particular lineages, we expect the same tendency to be widely observable.

Diversity of Photosynthesis-Related Genes in Nonphotosynthetic *Cryptomonas*

Significant differences were observed in the variety of carbon fixation- and chlorophyll synthesis-related genes in nonphotosynthetic *Cryptomonas*. The plastid genome of *C. paramecium* CCAP977/2a contains genes for three components of RuBisCO, namely, *rbcL*, *rbcS*, and *cbbX*, despite them being absent in the SAG977-2f and CCAP1634B plastid genomes. The existence of residual RuBisCO-related genes in nonphotosynthetic organisms is found not only in *Cryptomonas* but also in several other eukaryotes; for example, some plants, dinoflagellates, stramenopiles, and euglenids retain RuBisCO genes without exhibiting the ability to photosynthesize (Wolfe and dePamphilis 1997; Gockel and Hachtel 2000; Sekiguchi et al. 2002; Sanchez-Puerta et al. 2007; Wickett et al. 2008). Because RuBisCO is the key carbon fixation enzyme, it is unclear why it is retained in those nonphotosynthetic plastids. One proposed explanation is that in nonphotosynthetic strains, RuBisCO is maintained for functions other than the Calvin–Benson cycle, improving carbon efficiency via conversion of hexose to pyruvate in lipid biosynthesis (Schwender et al. 2004; Krause 2008). If RuBisCO plays

an essential role in *C. paramecium* CCAP977/2a, then RuBisCO-lacking *Cryptomonas* spp. SAG977-2a and CCAC1634B may have replaced its function with something else, either a different enzyme or a different process. Another possible explanation is that those RuBisCO proteins encoded in plastid genomes of nonphotosynthetic strains are undergoing evolutionary degradation prior to their complete loss. Consistent with this hypothesis, the RuBisCO protein sequence in the nonphotosynthetic euglenophyte *E. longa* is divergent compared with those of its photosynthetic relatives (Gockel and Hachtel 2000; Záhonová et al. 2016). In addition, the stability and abundance of RuBisCO in the cells are quite low, suggesting that the *E. longa* homolog is undergoing evolutionary degradation (Záhonová et al. 2016). Although a relaxation of selective constraint on RuBisCO genes in *C. paramecium* CCMP977/2a was suggested (Hoef-Emden et al. 2005), the d_N/d_S ratio of *rbcL* in our analyses was 0.0093, suggesting substantial functional constraints still exist (and/or its function has only very recently become unnecessary). At present, it is difficult to determine whether these genes are truly essential or in the process of being lost. Comprehensive comparative analyses between RuBisCO-retaining and RuBisCO-lacking nonphotosynthetic *Cryptomonas* species could help to explain the presence of RuBisCO in nonphotosynthetic plastids.

Another unexpected finding is that all three nonphotosynthetic *Cryptomonas* strains retain a partial chlorophyll biosynthetic pathway, albeit with different gene repertoires. The plastid genomes of *Cryptomonas* spp. SAG977-2a and CCAC1634B retain the light-independent protochlorophyllide reductase (LIPOR) consisting of *chlL*, *N*, and *B*. In contrast, *C. paramecium* CCAP977/2a lost the LIPOR components. All three nonphotosynthetic plastid genomes maintain Mg-protoporphyrin IX chelatase (*chlI*), which plays a role in the first step of the chlorophyll biosynthetic pathway, branching from the plastid heme biosynthetic pathway. In addition, genes for other subunits of Mg chelatase, *chlD* and *chlH*, and Mg-protoporphyrin IX methyltransferase (*bchM*) catalyzing next step of Mg-protoporphyrin IX chelatase were found in transcriptome data of *C. paramecium* CCAP977/2a (Keeling et al. 2014; Johnson et al. 2018). Overall, our bioinformatic investigations suggest that *C. paramecium* CCAP977/2a, *Cryptomonas* spp. SAG977-2f, and CCAP1634B retain at least part of the chlorophyll biosynthetic pathway for synthesizing protochlorophylide (CCAP977/2a) or chlorophylide *a* (CCAC1634B and SAG977-2f), which are precursors of chlorophyll *a* synthesis.

Extant cyanobacteria possess both LIPOR and light-dependent protochlorophyllide reductase (POR). Although these two enzymes are isofunctional, their origins differ. LIPOR likely arose from nitrogenase in anoxygenic photosynthetic bacteria (Fujita and Bauer 2003; Muraki et al. 2010). In addition, POR with high similarity to the short-chain dehydrogenase reductase evolved in oxygenic photosynthetic bacteria

(Suzuki and Bauer 1995; Hunsperger et al. 2015). In eukaryotes, the genes encoding POR have been transferred to the nuclear genome. In contrast, the genes encoding LIPOR remain in some plastid genomes but have been lost outright in many others (i.e., not transferred to the nuclear genome). Additionally, gene duplication and horizontal transfer of POR gave rise to various distribution patterns of coding POR in eukaryotes (Hunsperger et al. 2015; Matsuo and Inagaki 2018). In photosynthetic Cryptomonadales, LIPOR was found to be intact in *C. curvata* and *Storeatula* sp. CCMP 1868 (Kim et al. 2017). In contrast, *G. theta* and *Teleaulax amphioxeia* completely lost LIPOR from their plastid genomes (Douglas and Penny 1999; Kim et al. 2015), and *Rhodomonas salina*, *Chroomonas placoidea*, and *Chroomonas mesostigmatica* appear to have LIPOR only as pseudogenes (Khan et al. 2007; Kim et al. 2017). These distribution patterns of LIPOR suggest that nuclear POR is the main POR in Cryptomonadales, whereas plastid LIPOR is dispensable. On the other hand, it is still unclear whether nonphotosynthetic *Cryptomonas* species are currently undergoing gene loss because, despite the complete loss or degradation of all other photosynthesis-related proteins, they have retained intact chlorophyll biosynthesis enzymes. Furthermore, because chlorophyll precursors are photo-toxins generating reactive oxygen species (ROS), it is likely that those proteins potentially causing ROS generation disappear at an early stage of the loss of photosynthesis. This implies that they might have biological significance, such as in diverting chlorophyll precursors required for still unknown plastid functions in nonphotosynthetic *Cryptomonas* species. Future study of their functions is necessary.

Concluding Remarks

Despite having lost photosynthesis independently, the three nonphotosynthetic *Cryptomonas* species examined herein show similar patterns of gene retention in their plastid genomes. Structural changes in their genomes presumably occurred convergently via the same basic steps: 1) pseudogenization of photosynthesis-related genes tolerated by the loss of photosynthesis (as in SAG977-2f), 2) elimination of pseudogenic DNA, resulting in genome reduction (as seen in the plastid genomes of all three nonphotosynthetic strains), and 3) genome rearrangement accelerated by the loss of the rRNA operon-associated IRs (shown in CCMP977/2a and CCAP1634B). It is not clear whether steps 2 and 3 occurred sequentially or roughly concurrently. Although three nonphotosynthetic *Cryptomonas* have all lost the IRs, it is unclear whether loss of photosynthesis is associated with, and possibly a consequence of, the absence of IRs in Cryptomonadales. At the same time, significant differences in the retention of photosynthesis-related protein-coding genes were found in three nonphotosynthetic plastid genomes in *Cryptomonas*. Intraspecies analyses between two *C. curvata* plastid genomes support the existence of accelerated evolutionary rates of the

plastid genomes in *Cryptomonas*. The extent to which this rate increase is related to the repeated loss of photosynthesis in this genus is unclear. Plastid genome sequence data in Cryptomonadales are currently limited, and the fine-scale processes underlying the loss of photosynthesis in *Cryptomonas* are still uncertain. More genomic data from close photosynthetic relatives of nonphotosynthetic *Cryptomonas* are necessary in order to test the hypotheses present herein (e.g., the relationship between loss of IRs and genome rearrangements). Combined with the generation of experimental data, more genome sequences will help to elucidate the underlying causes of the loss of photosynthesis evolution in *Cryptomonas*.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

Adl SM, et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol. 66(1):4–119.

Archibald JM. 2015. Endosymbiosis and eukaryotic cell evolution. Curr Biol. 25(19):R911–R921.

Arisue N, et al. 2012. The *Plasmodium* apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent Malaria parasites. Mol Biol Evol. 29(9):2095–2099.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Blanc-Mathieu R, Sanchez-Ferandin S, Eyre-Walker A, Piganeau G. 2013. Organellar inheritance in the green lineage: insights from *Ostreococcus tauri*. Genome Biol Evol. 5(8):1503–1511.

Cenci U, et al. 2018. Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. BMC Biol. 16(1):137.

Curtis BA, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492(7427):59–65.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14(7):1394–1403.

Donaher N, 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.

Dorrell RG, et al. 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. Proc Natl Acad Sci U S A. 116(14):6914–6923.

Douglas SE, Penny SL. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol. 48(2):236–244.

Fujita Y, Bauer C. 2003. The light-dependent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark. In: Kadish K, Smith K, Guilard R, editors. The porphyrin handbook. San Diego (CA): Elsevier Science. p 109–156.

Gockel G, Hachtel W. 2000. Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. Protist 151(4):347–351.

Gould SB, Maier UG, Martin WF. 2015. Protein import and the origin of red complex plastids. Curr Biol. 25(12):R515–R521.

Grisdale CJ, Smith DR, Archibald JM. 2019. Relative mutation rates in nucleomorph-bearing algae. Genome Biol Evol. 11(4):1045–1053.

Guillard RL, Lorenzen CJ. 1972. Yellow-green algae with chlorophyllide C. J Phycol. 8:10–14.

Hadariová L, Vesteg M, Hampl V, Krajčovič J. 2018. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. Curr Genet. 64(2):365–387.

Hempel F, Bolte K, Klingl A, Zauner S, Maier UG. 2014. Protein transport into plastids of secondarily evolved organisms. In: Theg SM, Wollman F-A, editors. Plastid Biology. New York: Springer. p. 291–303.

Hoef-Emden K. 2005. Multiple independent losses of photosynthesis and differing evolutionary rates in the genus *Cryptomonas* (Cryptophyceae): combined phylogenetic analyses of DNA sequences of the nuclear and the nucleomorph ribosomal operons. J Mol Evol. 60(2):183–195.

Hoef-Emden K. 2007. Revision of the genus *Cryptomonas* (Cryptophyceae) II: incongruences between the classical morphospecies concept and molecular phylogeny in smaller pyrenoid-less cells. Phycologia 46(4):402–428.

Hoef-Emden K, Archibald JM. 2016. Phylum Cryptophyta (cryptomonads). In: Archibald JM, Simpson AGB, Slamovits C, editors. Handbook of the protists. Berlin: Springer-Verlag. p.851–891.

Hoef-Emden K, Tran HD, Melkonian M. 2005. Lineage-specific variations of congruent evolution among DNA sequences from three genomes, and relaxed selective constraints on *rbcL* in *Cryptomonas* (Cryptophyceae). BMC Evol Biol. 5(1):56.

Hunsperger HM, Randhawa T, Cattolico RA. 2015. Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. BMC Evol Biol. 15(1):16.

Janouškovec J, et al. 2013. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. PLoS One 8(3):e59001.

Johnson LK, Alexander H, Brown CT. 2018. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. Gigascience 8(4):giy158.

Kamikawa R, Azuma T, Ishii K-I, Matsuno Y, Miyashita H. 2018. Diversity of organellar genomes in non-photosynthetic diatoms. Protist 169(3):351–361.

Kamikawa R, Tanifuji G, et al. 2015. Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. Mol Biol Evol. 32(10):2598–2604.

Kamikawa R, Yubuki N, et al. 2015. Multiple losses of photosynthesis in *Nitzschia* (Bacillariophyceae). Phycol Res. 63(1):19–28.

Kamikawa R, et al. 2017. A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. Mol Biol Evol. 34(9):2355–2366.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33(2):511–518.

Keeling PJ, et al. 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12 (6):e1001889.

Khan H, Archibald JM. 2008. Lateral transfer of introns in the cryptophyte plastid genome. Nucleic Acids Res. 36(9):3043–3053.

Khan H, et al. 2007. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. Mol Biol Evol. 24(8):1832–1842.

Kim JI, Yoon HS, Yi G, Shin W, Archibald JM. 2018. Comparative mitochondrial genomics of cryptophyte algae: gene shuffling and dynamic mobile genetic elements. BMC Genomics. 19(1):275.

Kim JI, et al. 2015. The plastid genome of the cryptomonad *Teleaulax amphioxeia*. PLoS One 10(6):e0129284.

Kim JI, et al. 2017. Evolutionary dynamics of cryptophyte plastid genomes. Genome Biol Evol. 9(7):1859–1872.

Kolpakov R, Bana G, Kucherov G. 2003. mreps: efficient and flexible detection of tandem repeats in DNA. Nucleic Acids Res. 31(13):3672–3678.

Krause K. 2008. From chloroplasts to "cryptic" plastids: evolution of plastid genomes in parasitic plants. Curr Genet. 54(3):111–121.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.

Matsuo E, 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.

McFadden GI, Melkonian M. 1986. Use of HEPES buffer for microalgal culture media and fixation for electron microscopy. Phycologia 25(4):551–557.

Moore CE, Archibald JM. 2009. Nucleomorph genomes. Annu Rev Genet. 43(1):251–264.

Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. 2012. Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. Genome Biol Evol. 4(11):1162–1175.

Muraki N, et al. 2010. X-ray crystal structure of the light-independent protochlorophyllide reductase. Nature 465(7294):110–114.

Nowack ECM, Weber APM. 2018. Genomics-Informed insights into endosymbiotic organelle evolution in photosynthetic eukaryotes. Annu Rev Plant Biol. 69(1):51–84.

Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29(2):537–550.

Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10):944–945.

Sabir JSM, et al. 2014. Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. PLoS One 9(9):e107854.

Sanchez-Puerta MV, Lippmeier JC, Apt KE, Delwiche CF. 2007. Plastid genes in a non-photosynthetic dinoflagellate. Protist 158(1): 105–117.

Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432(7018):779–782.

Sekiguchi H, Moriya M, Nakayama T, Inouye I. 2002. Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Dictyochophyceae). Protist 153(2):157–167.

Smith DR, Lee RW. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Physiol. 164(4):1812–1819.

Strauss SH, Palmer JD, Howe GT, Doerksen AH. 1988. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc Natl Acad Sci U S A. 85(11):3898–3902.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acid Res. 34(Web Server):W609–W612.

Suzuki JY, Bauer CE. 1995. A prokaryotic origin for light-dependent chlorophyll biosynthesis of plants. Proc Natl Acad Sci U S A. 92(9):3749–3753.

Suzuki S, Endoh R, Manabe R-I, Ohkuma M, Hirakawa Y. 2018. Multiple losses of photosynthesis and convergent reductive genome evolution in the colourless green algae *Prototheca*. Sci Rep. 8(1):940.

Tanifuji G, Archibald JM. 2014. Nucleomorph comparative genomics. In: Löffelhardt W, editor. Endosymbiosis. Vienna (Austria): Springer. p. 197–213.

Tanifuji G, Archibald JM, Hashimoto T. 2016. Comparative genomics of mitochondria in chlorarachniophyte algae: endosymbiotic gene transfer and organellar genome dynamics. Sci Rep. 6(1):21016.

Tanifuji G, Onodera NT. 2017. Chapter eight—cryptomonads: a model organism sheds light on the evolutionary history of genome reorganization in secondary endosymbioses. In: Hirakawa Y, editor. Advances in botanical research. Cambridge: Academic Press. p. 263–320.

Tanifuji G, et al. 2011. Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. Genome Biol Evol. 3:44–54.

Tanifuji G, et al. 2014. Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. BMC Genomics. 15(1):374.

Turmel M, Lopes Dos Santos A, Otis C, Sergerie R, Lemieux C. 2019. Tracing the evolution of the plastome and mitogenome in the Chloropicophyceae uncovered convergent tRNA gene losses and a variant plastid genetic code. Genome Biol Evol. 11(4):1275–1292.

Turmel M, Otis C, Lemieux C. 2017. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. Sci Rep. 7(1):994.

Vieira Ldo N, et al. 2014. The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. PLoS One 9(3):e90618.

Wei L, et al. 2013. *Nannochloropsis* plastid and mitochondrial phylogenomes reveal organelle diversification mechanism and intragenus phylotyping strategy in microalgae. BMC Genomics. 14(1):534.

Wicke S, et al. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 25(10):3711–3725.

Wickett NJ, et al. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol Biol Evol. 25(2):393–401.

Wilson RJ, et al. 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J Mol Evol. 261(2):155–172.

Wolfe AD, dePamphilis CW. 1997. Alternate paths of evolution for the photosynthetic gene *rbcL* in four nonphotosynthetic species of *Orobanche*. Plant Mol Biol. 33(6):965–977.

Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci U S A. 89(22):10648–10652.

Yabuki A, et al. 2014. *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. Sci Rep. 4(1):4641.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Záhonová K, Füssy Z, Oborník M, Eliáš M, Yurchenko V. 2016. RuBisCO in non-photosynthetic alga *Euglena longa*: divergent features, transcriptomic analysis and regulation of complex formation. PLoS One 11(7):e0158790.

Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. Curr Opin Microbiol. 22:38–48.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31(13):3406–3415.

Associate editor: Gwenael Piganeau