

Article

Sequencing and Phylogenetic Analysis of Chloroplast Genes in Freshwater Raphidophytes

Ingrid Sassenhagen ^{1,*}  and Karin Rengefors ²

¹ Laboratoire d'Océanologie et des Geosciences, UMR CNRS 8187, Université du Littoral Côte d'Opale, 62930 Wimereux, France

² Aquatic Ecology, Department of Biology, Lund University, 22362 Lund, Sweden; karin.rengefors@biol.lu.se

* Correspondence: ingrid.sassenhagen@gmail.com

Received: 8 February 2019; Accepted: 20 March 2019; Published: 22 March 2019



Abstract: The complex evolution of chloroplasts in microalgae has resulted in highly diverse pigment profiles. Freshwater raphidophytes, for example, display a very different pigment composition to marine raphidophytes. To investigate potential differences in the evolutionary origin of chloroplasts in these two groups of raphidophytes, the plastid genomes of the freshwater species *Gonyostomum semen* and *Vacuolaria virescens* were sequenced. To exclusively sequence the organelle genomes, chloroplasts were manually isolated and amplified using single-cell whole-genome-amplification. Assembled and annotated chloroplast genes of the two species were phylogenetically compared to the marine raphidophyte *Heterosigma akashiwo* and other evolutionarily more diverse microalgae. These phylogenetic comparisons confirmed the high relatedness of all investigated raphidophyte species despite their large differences in pigment composition. Notable differences regarding the presence of light-independent protochlorophyllide oxidoreductase (LIPOR) genes among raphidophyte algae were also revealed in this study. The whole-genome amplification approach proved to be useful for isolation of chloroplast DNA from nuclear DNA. Although only approximately 50% of the genomes were covered, this was sufficient for a multiple gene phylogeny representing large parts of the chloroplast genes.

Keywords: raphidophytes; chloroplast genome; whole-genome-amplification; phylogeny; LIPOR

1. Introduction

Plastids are likely derived from a single endosymbiotic event incorporating a free-living cyanobacterium into the ancestor of the *Archaeplastida* (glaucoephytes, red algae and green algae) [1]. The chloroplasts of both red algae and green algae were subsequently transferred to other lineages by secondary endosymbiosis [2]. Red algae appear to have been taken up as endosymbionts only once, giving rise to a diverse group called chromalveolates. Additional layers of complexity come from recurring plastid loss and replacement. Some photosynthetic lineages have thus evolved from ancestors with a plastid of different origin, meaning that an ancestral plastid has been replaced with a new one [2]. Such replacement has taken place in several dinoflagellates by tertiary endosymbiosis with other chromalveolates, or serial secondary endosymbiosis with a green alga. This study focuses on plastids that carry out photosynthesis, and they will be referred to as chloroplasts throughout the text independent of their evolutionary acquisition.

Raphidophytes are eukaryotic, unicellular algae belonging to the chromalveolates. In particular, the marine species *Heterosigma akashiwo* and *Chattonella subsalsa* have received much scientific attention as they form harmful algal blooms and can cause extensive fish kills [3]. However, the freshwater species, such as *Gonyostomum semen* and *Vacuolaria virescens*, have been often overlooked and are

less studied due to their lack of toxins. Most phylogenetic and evolutionary studies of this class are, therefore, limited to the marine species.

A conspicuous difference between freshwater and marine raphidophytes is their large difference in coloration, marine taxa being brownish when observed under the microscope while freshwater taxa appear bright green. Analyses of pigment composition of *G. semen* and *V. virescens* revealed the presence of unusual pigments and identified major differences compared to marine raphidophytes [4]. The marine genera *Chattonella*, *Heterosigma* and *Fibrocapsa* share the pigments chlorophyll *a*, chlorophyll *c1c2*, violaxanthin, zeaxanthin, *a* and *b* carotene with *G. semen* and *V. virescens*, but also contain auroxanthin and fucoxanthin [5], which probably generates the typical brown color of these taxa. The freshwater raphidophytes *G. semen* and *V. virescens* additionally possess trans-neoxanthin, cis-neoxanthin, diadinoxanthin and alloxanthin [4]. The origin of these different pigments in raphidophytes, especially the presence of alloxanthin that typically occurs in cryptophytes, is currently unknown and might result from tertiary endosymbiosis.

To illuminate the evolutionary origin of freshwater raphidophyte chloroplasts, we selectively sequenced the respective genomes of *G. semen* and *V. virescens* after isolation of the organelles and whole-genome-amplification, thereby avoiding sequencing of the large nuclear genome. Assembled and annotated chloroplast genes were phylogenetically compared to genes of the marine raphidophyte *H. akashiwo* and other evolutionarily diverse microalgae.

2. Materials and Methods

The chloroplast genomes were sequenced from the *G. semen* culture GSMJ21 (Lund University, Sweden) and the *V. virescens* culture SAG 1195-1 (Culture Collection of Algae, Göttingen, Germany). For isolation of the chloroplasts, 1.5 mL of algal culture was pelleted by centrifugation and, after removal of the supernatant, resuspended in 1 mL breaking buffer (300 mM sorbitol, 50 mM Hepes-KOH (pH 7.5), 2 mM Na-EDTA, 1 μ L MgCl₂, 1% bovine serum albumin (BSA)) [6]. The algal cells were disrupted by vortexing with glass beads for 5 min. 10–20 chloroplasts were isolated from each sample by micropipetting under an inverted microscope (Nikon Eclipse TS100; Melville, NY, USA) at 400 \times magnification into polymerase chain reaction (PCR) tubes with 3 μ L phosphate buffered saline (PBS). The whole chloroplast genomes from the pooled 10–20 organelles were amplified by multiple displacement amplification (MDA) using the Qiagen RELPI-g kit (Qiagen, Hilden, Germany). Sequencing libraries of two samples from each species were constructed and paired-end sequenced on a partial Illumina MiSeq lane at the Department of Biology DNA sequencing facility in Lund (Sweden).

The quality of the Illumina reads was checked using the software FastQC [7]. Low quality sequences were trimmed or removed with the program Trimmomatic-0.36 [8] using leading and trailing cut-off values of 15 bp and a sliding-window approach with a window size of 10 bp and a quality threshold of 20. Reads from samples of the same species were pooled for assembly with IDBA-ud [9], SPAdes [10] and metaSPAdes [11] and the resulting assemblies were assessed with quast 4.5 [12]. IDBA-ud settings included pre-correction of reads before assembly, a minimum k value of 30 and a maximum k value of 70, k-mer increments of 10 bp, and a seed kmer size of 30 bp. SPAdes and meta-SPAdes were run with default settings and k-mer sizes of 21, 33, 55, 77 and 99 bp. Assemblies were also repeated for reduced datasets (10%, 20%, 25% and 30% of reads) due to very high (>1000 \times) and uneven coverage of chloroplast contigs.

After each assembly, the unconnected contigs were organized into biologically relevant units (bins) with Anvio (v2.2.2) [13] using information about GC content (approx. 30% GC expected based on *H. akashiwo* chloroplast genomes), coverage and sequence similarity. Thereby, contigs belonging to the chloroplasts were separated from potential nuclear DNA and bacterial contaminations. All contigs in bins that potentially represented the chloroplast genomes were compared with the NCBI nucleotide database using blastn [14]. To improve the assemblies, the coverage of the chloroplast contigs was homogenized by extracting the corresponding reads with the program Bowtie2 [15], and python and perl scripts from Albertsen et al. [16]. The size of these contig-specific read sets was reduced to

approximately 100× coverage by splitting them according to their original read depth. The remaining reads were re-assembled with SPAdes or metaSPAdes.

To compare the contribution from replicated libraries of the same species to the chloroplast genome assemblies, trimmed raw reads were mapped with the software Bowtie2-2.3.4.3 [17] against the assembled contigs using default settings. These alignments were analysed with samtools-1.9 [18].

Phylosift [19] was used to identify marker genes in the chloroplast contigs from *G. semen* and *V. virescens*. The chloroplast contigs were afterwards translated into amino acid sequences using the program prodigal [20]. These sequences were functionally annotated in KAAS (KEGG Automatic Annotation Server) [21] using the single-directional best hit method. To identify potentially unassembled fragments of light-independent protochlorophyllide oxidoreductase (LIPOR) genes (*chlB*, *chlN*, *chlL*) in the raw reads from *G. semen*, these reads were mapped with Bowtie2 against the chloroplast contigs of *V. virescens* that contained the genes of interest. Annotated amino acid sequences were aligned with MUSCLE [22] using default settings in the program Geneious 11.1.5 [23] and trimmed using the software trimAl applying a gap threshold of 0.75 [24].

Maximum likelihood trees were build for marker genes from Phylosift, individual and 53 concatenated amino acid sequences annotated with KAAS (*16S rRNA*, *eflAlike*, *hsp70*, *rplA*, *rplV*, *rplB*, *rpsI*, *rplC*, *rpsS*, *rpsG*, *rplP*, *rp L25/L23*, *rplF*, *rplK*, *rplE*, *rp S12/S23*, *rpsK*, *rpsH*, *rp L18P/L5E*, *rplM*, *rp L24*, *psaA*, *psaB*, *psaC*, *psaD*, *psaE*, *psaF*, *psaL*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbH*, *psbL*, *psbV*, *rbcS*, *rbcL*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *ilvB*, *sufB*, *petJ*, *chlB*, *chlN*, *chlL*, *atpA*, *atpD*, *atpE*, *clpC*, *groEL*) using RAxML [25]. The RAxML settings included rapid bootstrap analysis and automatic determination of protein substitution model while the number of distinct starting trees was based on bootstrapping criteria. For these trees, corresponding sequences from 47 reference organisms, selected to cover the genetic diversity within most microalgae lineages, were downloaded from NCBI. All trees were visualized with the online application iTOL [26]. tRNAs in the chloroplast bins were identified using the online program tRNAscan-SE [27]. Similarity and approximate coverage of the *H. akashiwo* chloroplast genome by *G. semen* and *V. virescens* contigs was assessed using blastn and visualized with the online application CGview [28].

3. Results

3.1. Processing of Reads

For *G. semen*, the two libraries yielded 702,765 and 706,755 forward and reverse reads (in total 1,409,520 reads), while 392,682 and 397,490 forward and reverse reads were generated for *V. virescens* (in total 790,172 reads) (Table 1). Different assembly approaches yielded the best results for *G. semen* and *V. virescens* probably due to differences in read quality and coverage. The reads from the *V. virescens* sample were assembled with metaSPAdes to 7172 contigs with on average 75× coverage. The best assembly was achieved for *G. semen* by assembling 25% of the reads with SPAdes, identification of chloroplast contigs through binning, homogenization of coverage of these selected contigs and reassembly of the remaining reads with SPAdes to 1649 contigs. After binning with Anvio and identity control with blastn, 35 contigs with a total length of 88,279 bp were identified as belonging to the chloroplast genome of *V. virescens* and 39 contigs with a total length of 74,603 bp to *G. semen* (Table 1). In comparison, the size of the chloroplast genome of the marine raphidophyte *H. akashiwo* ranges between 159,321 and 160,152 bp [29,30] suggesting approximately 50% completeness of the novel freshwater raphidophyte chloroplast genomes. The chloroplast contigs from *G. semen* and *V. virescens* had a mean coverage of 613× and 472× respectively with individual contigs having more than 1000× coverage. The majority of the contigs had in both cases a length between 1000–5000 bp and only 3 and 1 contigs respectively in each assembly were longer than 5000 bp. The assembled contigs can be found at NCBI under the accession numbers MK658836 and MK674494.

No contigs from the nuclear genome of the two algal species were identified in the other bins, which instead consisted of contigs from various bacterial species. Based on mapping with Bowtie2,

the majority of reads, 81% for *G. semen* and 85% for *V. virescens*, belonged to bacteria despite the high coverage of chloroplast contigs. Reads from replicated libraries were similarly distributed over the individual chloroplast contigs, meaning that contigs with high coverage attracted many reads from both libraries, while relatively few reads from both replicates mapped to contigs with lower coverage.

Table 1. Sequencing statistics of chloroplast (chl) samples from *G. semen* and *V. virescens*.

	<i>G. semen</i>	<i>V. virescens</i>
#reads (fwd and rev)	1,409,520	790,172
% bacterial reads	81.35	85.05
mean coverage of all contigs	95.3	75.4
#chl contigs	39	35
chl contigs length (bp)	74,603	88,279
mean coverage of chl contigs	612.9	472.3

3.2. Annotation

Phylosift identified 23 marker genes in the contigs from *G. semen* including the 16S rRNA, the RNA polymerase β' subunit (V136), the translation elongation factor 1, heat shock protein 70, cytochrome b6 (petB) and 18 ribosomal subunits. In the contigs from *V. virescens*, the 16S rRNA, the translation elongation factor 1, heat shock protein 70 and 26 ribosomal subunits were found. tRNAscan-SE found 10 tRNA sequences in the *G. semen* contigs (Ser, Gly, Phe, Ile, 2 \times Ala, 2 \times Leu, Asn, Asp) and seven in the *V. virescens* contigs (His, Arg, Ile, Ala, Val, Arg, Gly).

Prodigal identified 95 potential protein-coding sequences for *G. semen* and 115 for *V. virescens*. Out of these, 72 amino-acid sequences could be annotated with KAAS for *G. semen* and 84 for *V. virescens*. All genes were previously described in other microalgal chloroplast genomes and most were also found in the closest sequenced relative *H. akashiwo*. The complete chloroplast genome of *H. akashiwo* contains 156 protein coding genes, 130 with assigned function. Although the chloroplast genomes from this study remained incomplete preventing synteny analyses with reference genomes, the order of genes is likely very similar to *H. akashiwo*, as the freshwater raphidophyte plastid contigs mapped very well to large parts of the chloroplast genome of the marine raphidophyte (Figure 1). High sequence similarity was observed, especially in coding genomic regions with higher than average GC content (30.5%).

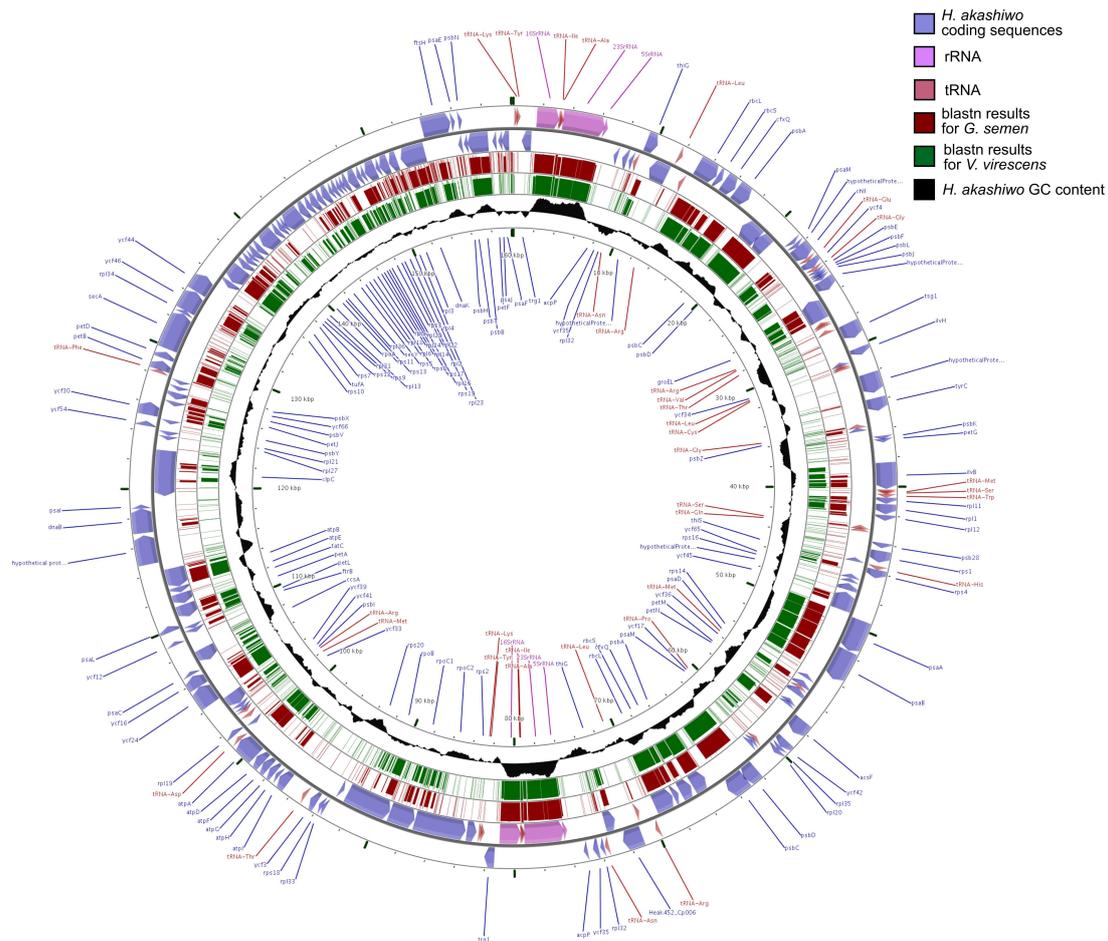


Figure 1. Sequence similarities based on blastn between to the circular chloroplast genome of *H. akashiwo* (~160 kbp) and chloroplast contigs of *G. semen* and *V. virescens* visualized with CGView. Coloured regions on the *G. semen* and *V. virescens* rings correspond to fragments matching the reference genome. The GC content of the *H. akashiwo* chloroplast genome averages at 30.5%.

3.3. Phylogenetic Analyses

The phylogenetic analyses of 53 different chloroplast protein sequences using maximum likelihood trees showed a high relatedness of *G. semen*, *V. virescens* and *H. akashiwo* (Figure 2). Individual and concatenated sequences from the raphidophyte species clustered together in all phylogenetic analyses. A maximum likelihood tree based on 53 chloroplast genes concatenated and trimmed to 14,896 amino acids confirmed the monophyly of *stramenopiles*. The raphidophytes represented the sister group to *xantophyceae* and *phaeophyceae*. The arrangement of genes in the different chloroplast genomes could not be compared due to the incompleteness of the *G. semen* and *V. virescens* genomes and the limited length of the assembled contigs.

One significant difference in gene composition compared to the completely sequenced chloroplast genomes of *H. akashiwo* [29,30] was, however, identified. The genes for the three subunits of the light-independent protochlorophyllide oxidoreductase (LIPOR: *chlB*, *chlL*, *chlN*) are encoded in the chloroplast genome of *V. virescens*, but were not identified in *G. semen*. The genes have previously been found in the marine raphidophyte *Chattonella subsala*, but appear to be absent in *H. akashiwo* and all diatoms. However, the genes are present in other *stramenopiles* such as *phaeophyceae* and *xantophyceae* (Figure 2), suggesting independent loss of these genes in different taxonomic lineages [31].

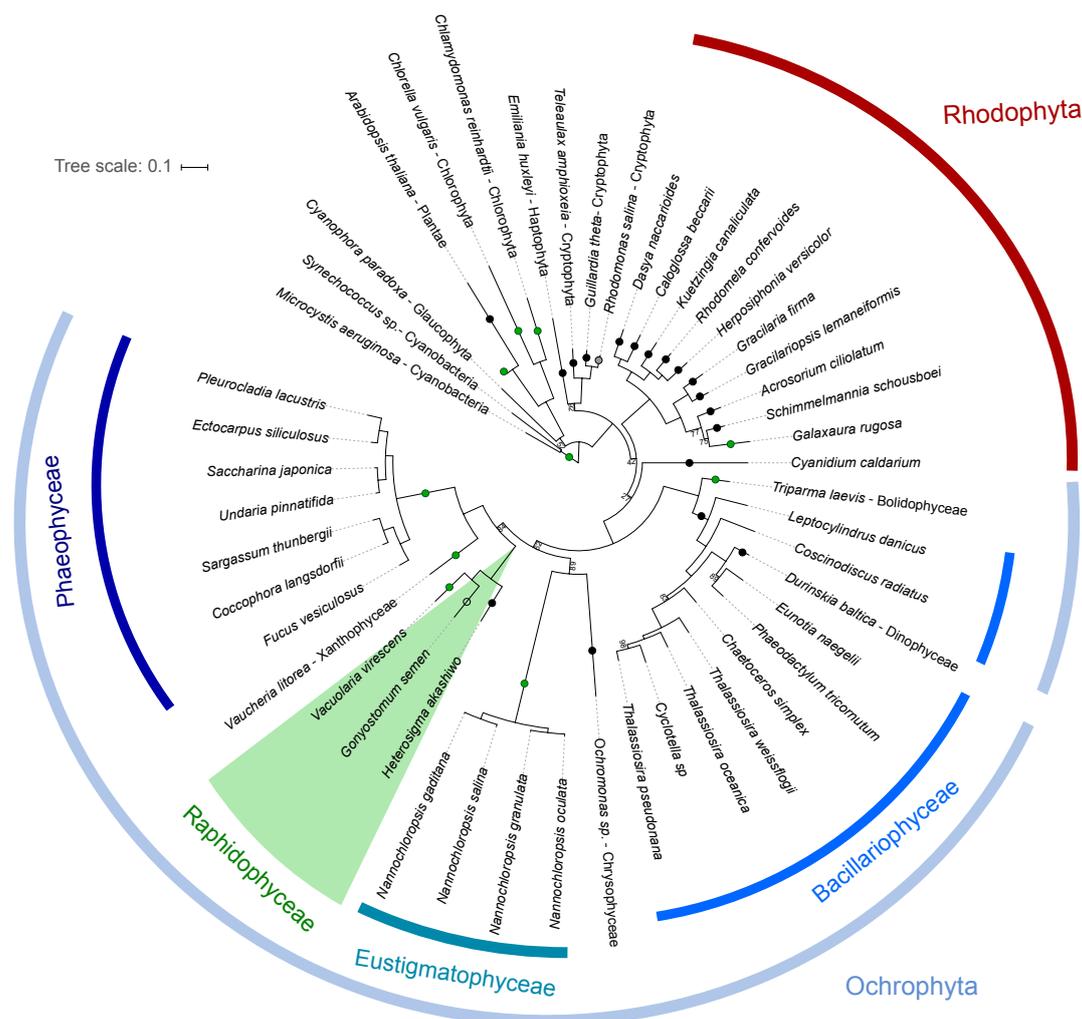


Figure 2. Maximum likelihood phylogeny of 53 concatenated chloroplast protein sequences, 14,896 amino acids in length, built with RAxML from an alignment of 49 taxa with bootstrap values <100 displayed on the internal branches. *Raphidophyceae* are highlighted in green, while curved lines indicate the position of other taxonomic groups. Circles on the branches indicate presence/absence of light-independent protochlorophyllide oxidoreductase (LIPOR) genes based on [31] and NCBI: green = presence, black = absence, grey = pseudogenes, empty = unknown.

4. Discussion

The whole-genome-amplification approach chosen for this study enabled de-novo sequencing of large parts of the chloroplast genome of two freshwater raphidophyte species. The advantage of this method is that it allows sequencing of the chloroplasts without wasting reads on the large nuclear genome (likely >100 pg per cell, [32]) thereby attaining very high sequencing depth of the targeted genomes. Since the chloroplasts from each library were acquired from single cells, we demonstrate that the method can be useful for taxa that cannot be cultured. Overall, the sequencing data yielded approximately 50% of the chloroplast genomes and 100 protein-coding sequences per species, showing that the approach allows for downstream phylogenomic analyses.

However, the assembled contigs in this study were on average very short, which is likely caused by limitations of multiple displacement amplification (MDA) of the chloroplast genomes. Previous studies have shown that MDA is very sensitive to reaction gain, meaning that overrepresented fragments get amplified more throughout the reaction [33,34]. Thus, high gain potentially led to a strong bias towards certain genomic regions in our study and limited amplification of other parts of the genome.

Similarly biased amplification was observed in both replicates of the sequenced species suggesting a recurring bias towards certain genomic regions such as areas with high GC content. This suggests that additional replicates would not necessarily have improved the chloroplast genome recovery. However, decreased reaction volumes achieved through nanoliter reactors [35] or emulsion-based amplification [36] might result in more uniform and accurate MDA of single cells. Furthermore, new single-cell whole genome amplification approaches, such as the MALBAC [37] or WGA-X [38], promise less bias and higher uniformity.

A major concern when using whole-genome-amplification methods is amplification of non-targeted DNA, such as bacterial and human DNA. Binning of contigs based on similarity, GC-content and coverage allowed distinguishing between chloroplast and bacterial contigs. Treatment of the algal cultures with antibiotics to decrease the abundance of bacteria in the medium prior to the isolation of chloroplasts could likely further improve this approach.

Despite not recovering complete genomes, assembly of many chloroplast genes allowed phylogenetic comparisons with marine raphidophytes and other microalgae. The phylogenetic trees of individual and concatenated chloroplast genes confirmed the high relatedness of freshwater and marine raphidophyte chloroplasts, despite the large differences in pigment composition. This finding largely rejects the hypothesized different evolutionary origin of plastids in freshwater raphidophytes. The chloroplasts of raphidophyceae were closely related to phaeophyceae agreeing with the phylogenetic position of these two groups based on nuclear genes [39].

The origin of unusual accessory pigments in *G. semen* and *V. virescens*, such as alloxanthin, diadinoxanthin, trans-neoxanthin and cis-neoxanthin, however, remains uncertain. Different pathways for the biosynthesis of alloxanthin, which is characterized by two acetylenic groups, have been suggested in previous studies. In cryptophytes, transformations of zeaxanthin or diatoxanthin into alloxanthin by hitherto unknown enzymes have been hypothesized [40–42]. In the dinoflagellate *Amphidinium carterae*, the use of a ^{14}C -tracer has demonstrated that acetylene in alloxanthin likely forms from an allen group [43]. Zeaxanthin is first changed to neoxanthin via violaxanthin. Neoxanthin, which has one allen group, is then converted into diadinoxanthin, which has one acetylenic group. The presence of violaxanthin, neoxanthin and diadinoxanthin in the *G. semen* and *V. virescens* suggest a pathway similar to that of *A. carterae* for the synthesis of alloxanthin.

To ultimately identify the biosynthesis pathway of xanthophylls in freshwater raphidophytes, their nuclear genome needs to be sequenced, where the responsible enzymes are likely encoded. Previous studies have shown that only few genes encoding the chloroplast proteome are located in the chloroplast genome [31,44,45], while the majority has been transferred to the nucleus. The proteins encoded in the nucleus are synthesized as precursor proteins in the cytosol and are imported post-translationally into the chloroplast [46]. Nuclear genes appear to be more affected by horizontal gene transfer than chloroplast genomes [2] suggesting that the enzymes involved in the synthesis of the unusual xanthophylls in freshwater raphidophytes might originate from horizontal gene transfer.

Despite the high similarity in chloroplast genome composition of *G. semen*, *V. virescens* and *H. akashiwo*, one notable difference was observed in this study. The genes *ChlL*, *ChlN* and *ChlB*, encoding the light-independent protochlorophyllide oxidoreductase (LIPOR), were found in *V. virescens* but are absent in *H. akashiwo* chloroplast genomes. These genes have also been reported from the marine raphidophyte *C. subsalsa* [31]. However, the presence of LIPOR genes in *G. semen* remains unfortunately uncertain due to the incompleteness of the chloroplast genomes in this study. Protochlorophyllide oxidoreductases are essential for the transformation of the pigment protochlorophyllide to chlorophyllide during chlorophyll *a* synthesis. Two different enzymes, which are both present in cyanobacteria, can catalyze this step: the light-dependent (POR) and light-independent protochlorophyllide oxidoreductase (LIPOR). As the name suggests, LIPOR facilitates the synthesis of chlorophyll *a* in the dark [47]. After primary endosymbiosis, the POR genes were transferred to the nucleus, while the LIPOR genes remained in the chloroplast genome or were lost in several microalgal lineages [47]. In the phyla *chlorarachniophytes*, *euglenoids* and

haptophytes, the loss of LIPOR genes may have occurred early during establishment of these lineages. In contrast, in chlorophytes, rhodophytes, *cryptophytes*, heterokonts, and *chromerids*, LIPOR genes were only lost by individual taxa [47] resulting in large heterogeneity regarding chlorophyll *a* synthesis in these groups.

The presence of LIPOR genes might be a beneficial adaptation to growth under low light conditions as commonly found in many northern European lakes rich in organic matter and humic substances. Furthermore, high iron concentrations in these lakes [48] might facilitate the synthesis of an iron-requiring protein such as LIPOR. In contrast, the synthesis of LIPOR might be disadvantageous for many marine taxa in iron-depleted regions in the ocean [31]. The loss of the LIPOR enzyme might have also occurred in several microalgal lineages due to its inefficiency under aerobic conditions. Similar to the nitrogenase enzyme, from which the LIPOR enzyme evolved, it is very sensitive to oxygen, allowing chlorophyll *a* synthesis only under anaerobic conditions [49]. Many lakes in northern Europe with high water color experience stratification and hypoxia in bottom waters during summer. Freshwater raphidophytes such as *G. semen* migrate at night several meters into the meta- and hypolimnion to access soluble reactive phosphorus gradually released from the sediment in anoxic conditions [50]. During this migration and presence in the anoxic hypolimnion, these microalgae could utilize LIPOR for additional chlorophyll *a* synthesis.

Author Contributions: I.S. conceived and designed this study with contributions from K.R. and I.S. carried out the lab work, data analyses and wrote the manuscript. K.R. revised the manuscript and approved the final submitted version.

Funding: This research was supported by a grant from the Helge Ax:son Johnson foundation.

Acknowledgments: We thank Nina Dombrowski, Brett Baker and Dag Ahrén for advice about bioinformatic analyses during this project. We also appreciate Allan Rasmusson's help with the isolation of chloroplasts from our algal cultures. Constructive comments from three reviewers significantly improved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Archibald, J.M. The Puzzle of Plastid Evolution. *Curr. Biol.* **2009**, *19*, R81–R88. [PubMed]
2. Keeling, P.J. The endosymbiotic origin, diversification and fate of plastids. *R. Soc. Philos. Trans. Biol. Sci.* **2010**, *365*, 729–748. [CrossRef] [PubMed]
3. Horiguchi, T. Raphidophyceae (Raphidophyta). In *Handbook of the Protists*; Archibald, J.M., Simpson, A.G.B., Slamovits, C.H., Margulis, L., Melkonian, M., Chapman, D.J., Corliss, J.O., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 1–26. ISBN 978-3-319-32669-6.
4. Sassenhagen, I.; Rengefors, K.; Richardson, T.L.; Pinckney, J.L. Pigment composition and photoacclimation as keys to the ecological success of *Gonyostomum semen* (Raphidophyceae, Stramenopiles). *J. Phycol.* **2014**, *50*, 1146–1154. [CrossRef] [PubMed]
5. Mostaert, A.S.; Karsten, U.; Hara, Y.; Watanabe, M.M. Pigments and fatty acids of marine raphidophytes: A chemotaxonomic re-evaluation. *Phycol. Res.* **1998**, *46*, 213–220. [CrossRef]
6. Mason, C.B.; Matthews, S.; Bricker, T.M.; Moroney, J.V. Simplified Procedure for the Isolation of Intact Chloroplasts from *Chlamydomonas reinhardtii*. *Plant Physiol.* **1991**, *97*, 1576–1580. [PubMed]
7. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 1 May 2016).
8. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
9. Peng, Y.; Leung, H.C.M.; Yiu, S.M.; Chin, F.Y.L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28*, 1420–1428. [CrossRef]
10. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
11. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834. [CrossRef]

12. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUASt: Quality assessment tool for genome assemblies. *Bioinforma. Oxf. Engl.* **2013**, *29*, 1072–1075. [[CrossRef](#)]
13. Eren, A.M.; Esen, Ö.C.; Quince, C.; Vineis, J.H.; Morrison, H.G.; Sogin, M.L.; Delmont, T.O. Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* **2015**, *3*, e1319.
14. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421. [[CrossRef](#)]
15. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359.
16. Albertsen, M.; Hugenholtz, P.; Skarshewski, A.; Nielsen, K.L.; Tyson, G.W.; Nielsen, P.H. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **2013**, *31*, 533–538. [[CrossRef](#)]
17. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [[CrossRef](#)]
18. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
19. Darling, A.E.; Jospin, G.; Lowe, E.; Matsen, F.A.; Bik, H.M.; Eisen, J.A. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2014**, *2*, e243.
20. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **2010**, *11*, 119. [[CrossRef](#)]
21. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [[CrossRef](#)]
22. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
23. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)]
24. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973.
25. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
26. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)]
27. Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57.
28. Grant, J.R.; Stothard, P. The CGView Server: A comparative genomics tool for circular genomes. *Nucleic Acids Res.* **2008**, *36*, W181–W184. [[CrossRef](#)]
29. Cattolico, R.A.; Jacobs, M.A.; Zhou, Y.; Chang, J.; Duplessis, M.; Lybrand, T.; McKay, J.; Ong, H.C.; Sims, E.; Rocop, G. Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. *Bmc Genomics* **2008**, *9*, 211. [[CrossRef](#)]
30. Seoane, S.; Hyodo, K.; Ueki, S. Chloroplast Genome Sequences of Seven Strains of the Bloom-Forming Raphidophyte *Heterosigma akashiwo*. *Genome Announc.* **2017**, *5*, e01030-17. [[CrossRef](#)]
31. Hunsperger, H.M.; Randhawa, T.; Cattolico, R.A. Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. *BMC Evol. Biol.* **2015**, *15*, 16.
32. Sassenhagen, I.; Sefbom, J.; Sall, T.; Godhe, A.; Rengefors, K. Freshwater protists do not go with the flow: Population structure in *Gonyostomum semen* independent of connectivity among lakes. *Environ. Microbiol.* **2015**, *17*, 5063–5072.
33. de Bourcy, C.F.A.; Vlaminc, I.D.; Kanbar, J.N.; Wang, J.; Gawad, C.; Quake, S.R. A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. *PLOS ONE* **2014**, *9*, e105585.
34. Borgström, E.; Paterlini, M.; Mold, J.E.; Frisen, J.; Lundeberg, J. Comparison of whole genome amplification techniques for human single cell exome sequencing. *PLOS ONE* **2017**, *12*, e0171566.
35. Marcy, Y.; Ishoey, T.; Lasken, R.S.; Stockwell, T.B.; Walenz, B.P.; Halpern, A.L.; Beeson, K.Y.; Goldberg, S.M.D.; Quake, S.R. Nanoliter Reactors Improve Multiple Displacement Amplification of Genomes from Single Cells. *PLOS Genet.* **2007**, *3*, e155.

36. Fu, Y.; Li, C.; Lu, S.; Zhou, W.; Tang, F.; Xie, X.S.; Huang, Y. Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11923–11928.
37. Zong, C.; Lu, S.; Chapman, A.R.; Xie, X.S. Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. *Science* **2012**, *338*, 1622–1626. [[CrossRef](#)]
38. Stepanauskas, R.; Fergusson, E.A.; Brown, J.; Poulton, N.J.; Tupper, B.; Labonté, J.M.; Becraft, E.D.; Brown, J.M.; Pachiadaki, M.G.; Povilaitis, T.; et al. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun.* **2017**, *8*, 84. [[CrossRef](#)]
39. Derelle, R.; López-García, P.; Timpano, H.; Moreira, D. A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (=Heterokonts). *Mol. Biol. Evol.* **2016**, *33*, 2890–2898.
40. Li, H.; Wang, W.; Wang, Z.; Lin, X.; Zhang, F.; Yang, L. De novo transcriptome analysis of carotenoid and polyunsaturated fatty acid metabolism in *Rhodomonas* sp. *J. Appl. Phycol.* **2016**, *28*, 1649–1656.
41. Takaichi, S. Carotenoids in algae: distributions, biosyntheses and functions. *Mar. Drugs* **2011**, *9*, 1101–1118.
42. Takaichi, S.; Yokoyama, A.; Mochimaru, M.; Uchida, H.; Murakami, A. Carotenogenesis diversification in phylogenetic lineages of Rhodophyta. *J. Phycol.* **2016**, *52*, 329–338.
43. Swift, I.E.; Milborrow, B.V.; Jeffrey, S.W. Formation of Neoxanthin, Diadinoxanthin and Peridinin from [¹⁴C]Zeaxanthin by a cell-free System from *Ampidinium carterae*. *Phytochemistry* **1982**, *21*, 2859–2864. [[CrossRef](#)]
44. Race, H.L.; Herrmann, R.G.; Martin, W. Why have organelles retained genomes? *Trends Genet.* **1999**, *15*, 364–370. [[CrossRef](#)]
45. Leister, D. Chloroplast research in the genomic age. *Trends Genet.* **2003**, *19*, 47–56. [[CrossRef](#)]
46. Soll, J.; Schleiff, E. Protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 198–208. [[CrossRef](#)] [[PubMed](#)]
47. Fong, A.; Archibald, J.M. Evolutionary Dynamics of Light-Independent Protochlorophyllide Oxidoreductase Genes in the Secondary Plastids of Cryptophyte Algae. *Eukaryot. Cell* **2008**, *7*, 550–553. [[CrossRef](#)] [[PubMed](#)]
48. Björnerås, C.; Weyhenmeyer, G.A.; Evans, C.D.; Gessner, M.O.; Grossart, H.-P.; Kangur, K.; Kokorite, I.; Kortelainen, P.; Laudon, H.; Lehtoranta, J.; et al. Widespread Increases in Iron Concentration in European and North American Freshwaters. *Glob. Biogeochem. Cycles* **2017**, *31*, 1488–1500. [[CrossRef](#)]
49. Yamazaki, S.; Nomata, J.; Fujita, Y. Differential Operation of Dual Protochlorophyllide Reductases for Chlorophyll Biosynthesis in Response to Environmental Oxygen Levels in the Cyanobacterium *Leptolyngbya boryana*. *Plant Physiol.* **2006**, *142*, 911–922. [[PubMed](#)]
50. Salonen, K.; Rosenberg, M. Advantages from diel vertical migration can explain the dominance of *Gonyostomum semen* (Raphidophyceae) in a small, steeply-stratified humic lake. *J. Plankton Res.* **2000**, *22*, 1841–1853. [[CrossRef](#)]

