Version of Record: https://www.sciencedirect.com/science/article/pii/S2211926418307938
Manuscript_5c76bd13dbbb0a62915d758bd6a4b744

Chrysochromulina: Genomic assessment and taxonomic diagnosis of the type species for an oleaginous algal clade

Blake T. Hovde[1], Chloe R. Deodato[2], Robert A. Andersen[3], Shawn R. Starkenburg[1], Steven B. Barlow[4], Rose Ann Cattolico[2§]

[1]Los Alamos National Laboratory, Bioscience Division, Los Alamos, NM

[2]Department of Biology, University of Washington, Box 351800, Seattle, WA, USA

[3]Friday Harbor Laboratories, University of Washington, 620 University Road, Friday Harbor, WA, USA

[4]Department of Biology, San Diego State University, 5500 Campanile Drive MC-4614, San Diego, CA, USA

[§]Corresponding author, Co-first author

Email addresses:

BTH: hovdebt@lanl.gov

CRD: chloed@uw.edu

RAA: raa48@uw.edu

SRS: shawns@lanl.gov

SBB: sbarlow@mail.sdsu.edu

RAC: racat@uw.edu

ABSTRACT

Background: Until recent studies documented their extensive contribution to primary productivity and carbon sequestration, haptophytes remained underappreciated players in global ecosystem processes. Contemporary analyses, augmented by the use of molecular probes, show the haptophyte taxon *Chrysochromulina* to be seminal to the ecology of both marine and freshwater ecosystems. Unfortunately, description for the type species for this clade remains enigmatic.

Results: *Chrysochromulina parva* Lackey was re-isolated from Big Walnut Creek (Ohio), the site where the original isolate was obtained. The sequenced haploid genome of this organism is 65.7 Mb in size. Several noteworthy nuclear-encoded genes identified include a novel *fts*Z (mediates organelle division) that phylogenetically clusters with the Chloroarachniophytes. Also revealed, is a complement of genes associated with meiosis and DNA repair, indicating the presence of a sexual cycle in this alga. Mitochondrial genes lost to the nucleus include all extrinsic components of the *nad* complex, completing a punctate pattern of transfer that is observed among algal taxa. Comparison of the newly sequenced *Chrysochromulina parva* Lacky isolate was made with that of *Chrysochromulina tobinii* (59.1 Mb) – a fresh water strain isolated from a lake in Colorado.

Conclusion: Genomic analysis suggests that fresh water *Chrysochromulina* isolates, though geographically well separated, from a related clade. The name of the type species of *Chrysochromulina parva* Lacky is anchored with a lectotype and epitype, and the second isolate is described as *Chrysochromulina tobinii* sp. nov. *Chrysochromulina* represents a new, extremely tractable model organism for experimental studies. This oleaginous alga has a small genome and because it represents only the second haptophyte taxon to be sequenced and assembled, presents new opportunity to examine the evolution of an algal taxon that plays an intrinsic role in ecosystem function.

KEY WORDS: Haptophyta; *Chrysochromulina* genome; *nad* evolution; FtsZ phylogeny; meiosis; taxon diagnosis.

I. Introduction

Haptophytes represent a major taxonomic group of phytoplankters whose ancient evolutionary roots extend from the mid Neoproterozoic-Cryogenian period approximately 820 MA [1, 2]. The ecological relevance of micro-planktonic haptophytes cannot be underestimated. These algae produce copious amounts of fatty acids – an energy currency required for the survival of aquatic eco-cohorts at every trophic level [3-8]. Many haptophyte species are embellished with different types of scales that have a polysaccharide core. Some haptophytes further adorn these organic scales with a complex array of calcium carbonate crystals [9-11]. It has been estimated that select haptophyte taxa generate more than two billion tons of polysaccharide gels annually that serve as a nutrient source for bacteria and zooplankton [12, 13]. Both scales and polysaccharides sediment into the benthos, thus serve as carbon sequestration products. The recent identification of DMSP [14] biosynthesis genes in haptophytes expands the contribution of these algae to the production of organo-sulfur metabolites (which influence sulfur cycling in association with eco-cohorts [Durham, personal communication]), and impact atmospheric geochemical cycles (by regulating weather patterns [15]). High-density haptophyte blooms have been shown to produce toxins that cause extensive finfish losses [16], as well compromise water quality by generating noxious odors [17].

Two classes of Haptophytes are recognized: the monophyletic Pavlovophyceae and the highly diverse Prymnesiophyceae. In spite of their significant ecological contribution, the first haptophyte complete genome sequence only became available to investigators in 2013 [18] wherein the estimated genome of 141.7 Mb for *Emiliania huxleyi* (Prymnesiophyceae;

Isochrysidales) was reported. This well-studied alga is of great interest for it generates bloom events large enough to cover square miles of ocean - so extensive that the highly reflective calcium carbonate scales covering this organism allow satellite monitoring. In contrast, the crucial ecological role played by the minute, fragile members of the *Chrysochromulina* (Prymnesiales) clade has long been observed (e.g., Dahl et al [19]; 2005; Seoane et al, 2009 [20]), but only recently been the focus of more intense study [14, 21]. Data show that both oceanic and fresh water phytoplankton assemblages often contain high percentages of *Chrysochromulina*, in some instances, forming almost uni-algal, high-density blooms [17, 22]. The new awareness of *Chrysochromulina* relevance to global aquatic ecosystem function drove the generation of a second haptophyte genome sequence. *Chrysochromulina tobinii* Cattolico was isolated from a high altitude, freshwater lake in Colorado in 1991 by Dr. Paul Kugrens (Colorado State University). The genome of this alga displays a spectrum of unique nuclear genes and novel chloroplast and mitochondrial genome architectures that are not present in *E. huxleyi* [23, 24].

Chrysochromulina is a large, cosmopolitan complex (~64 species [25]). Confusion concerning the "type species" Diagnosis for *Chrysochromulina* has been ongoing. The initial description of *Chrysochromulina Parva* Lackey was made in 1939 [26], citing a naked cell having two flagella and an unusual flagella-like appendage that was subsequently [27] termed a haptonema. With the onset of electron microscopy, high-resolution studies found ephemeral scales embellished many *Chrysochromulina* isolates, some of which were also identified as *Chrysochromulina parva* Lackey [28], further adding to nomenclature confusion. To untangle this conundrum, we re-isolated the *Chrysochromulina* "type species" (*Chrysochromulina parva* Lacky) from Big Walnut Creek, Ohio, at the same location where the original organism was recovered in 1939. In this study we report on the genetic profile of the scale-less Big Walnut Creek isolate. Our work serves as a conduit for clarifying the Diagnosis of the *Chrysochromulina* type species. The generation of this second *Chrysochromulina* genome has also allowed comparison to be made

between *Chrysochromulina parva* Lackey (65.7 Mb) and *Chrysochromulina tobinii* Cattolico (59.1 Mb). Data suggests that *Chrysochromulina* clade is comprised of a cryptic species complex, for though the morphology of *Chrysochromulina parva* Lackey and *Chrysochromulina parva* Cattolico is almost identical [29], their genetic fingerprints are not [23, 24].

2. Results and discussion

Chrysochromulina parva isolate UW 1161 was established in 2014 from a water sample collected from Big Walnut Creek in Shadeville, Ohio, the type locality for this species. Scanning and transmission electron microscopy of the *C. parva* Ohio isolate shows the small unicell to be 4-6 μm in size. Two flagella (~8 μm) as well as a long haptonema (up to 10x the body length) initiate anteriorly, within a deep groove that runs the length of the cell (Fig. 1). When extended, the haptonema is significantly longer than the cell body of the alga (Fig. 2). Transmission electron studies reveal a simple cellular morphology for *C. parva.* Briefly, a typical eukaryotic nucleus with nucleolus is anteriorly located, and the prolific mitochondria have tubular cristae. The two chloroplasts, which are delineated by four membranes, have internal pyrenoids. Each chloroplast is associated with a large lipid body that lies in the cell anterior. Unlike most *Chrysochromulina* isolates [25, 30], no scales are visible on this organism, either on the cell surface or within the Golgi apparatus. A detailed comparison of *C. parva*, a river dwelling isolate and *C. tobinii*, a lake dwelling isolate, showing these algae to be morphologically similar, will be reported elsewhere [29]. To our knowledge, only three fresh-water *Chrysochromulina* isolates are presently maintained in culture.

2.1.1 Nuclear genome - assembly and annotation:

The *Chrysochromulina parva* genome was shotgun sequenced using Illumina short reads. The final draft genome assembly consisted of 8,362 contigs having an average length of 7,865 bp

(Table 1). A 65.8 Mb genome was assembled representing an average read depth of over 100x (see Materials and Methods), which is larger than that of *C. tobinii* genome assembly (59.1 MB [24]). Whether the *C. parva* versus *C. tobinii* genome size difference is due to a truly larger genome for *C. parva*, issues with assembly methodology, or whether associated bacteria continue to confound the genome assembly process, is not known. Germane to this discussion is the observation that significant genome size differences have been reported among *Nannochloropsis* isolates (up to 27%) and even between putative species (e.g., 15% difference is observed between *N. oceanica* CCMP531 and *N. oceanica* IMET1) [31]. The *C. parva* nuclear genome (Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession PJAB00000000, BioProject PRJNA418464) encodes an estimated 28,138 protein-coding genes (Table 1) with an average length of 1,765 bp. On average, each gene contains a single intron. Similar to *C. tobinii* [24], the *C. parva* nuclear genome encodes several genes that may be of interest to future studies, including a unique complex of non-ribosomal synthetase modules associated with Type 1 polyketide synthetase domains; tylosin and erythromycin antimicrobial peptides; multidrug and toxic compound extrusion proteins, and an alternative RuBisCO activase. A novel xanthorhodopsin is also maintained in *C. parva.* This protein sources non-photosynthetically generated energy that can be used in cellular metabolism. The xanthorhodopsin gene found in *C. parva* as well as *C. tobinii* is of particular interest, since a highly conserved ortholog has been found encoded in the genome of the bacterium *Sphingobium* sp. RAC03 (Supplementary Fig. S1; NCBI: GCF_001713425) which is a member of the nine-membered bacterial biome associated with *C. tobinii* [32]. Both the algal and bacterial proteins maintain identical sites needed for green wavelength spectral tuning, proton acceptor and donor function as well as retinal binding. Such strong sequence similarities suggest the possibility of a lateral transfer origin for the *Chrysochromulina* protein from a bacterial source.

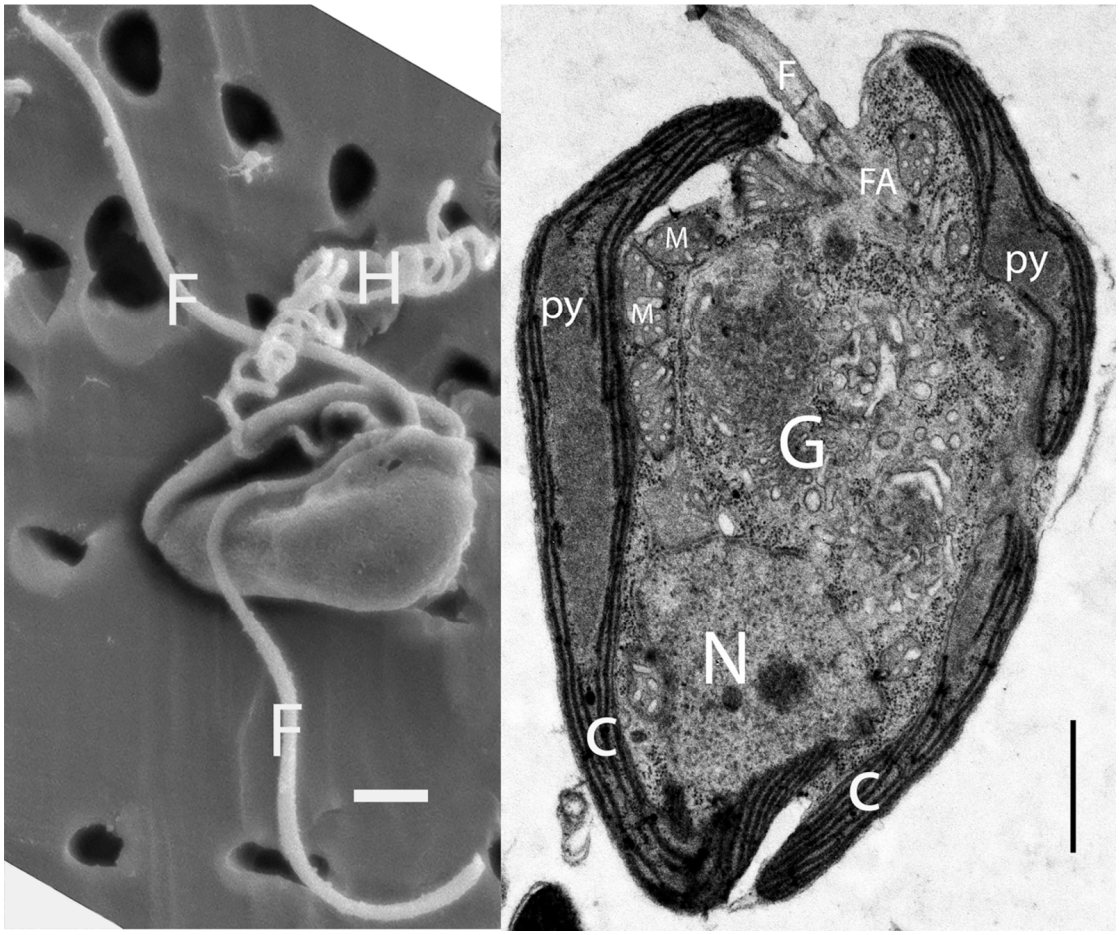

Fig 1: **Images of *Chrysochromulina parva*:** (Left) Scanning electron microscopy image of cell showing two flagella (F), and the partially uncoiled haptonema (H) emerging from a groove or depression along the length of the cell. (Right) Transmission electron microscopy image of a longitudinal slice through the cell showing the two peripheral chloroplasts (C) and displaying an internal pyrenoid (py); mitochondria (M); Golgi apparatus (G); nucleus (N); and one of the two flagella (F), as well as the flagellar and haptonemal complex (FA). Scale bar = 1 μm.

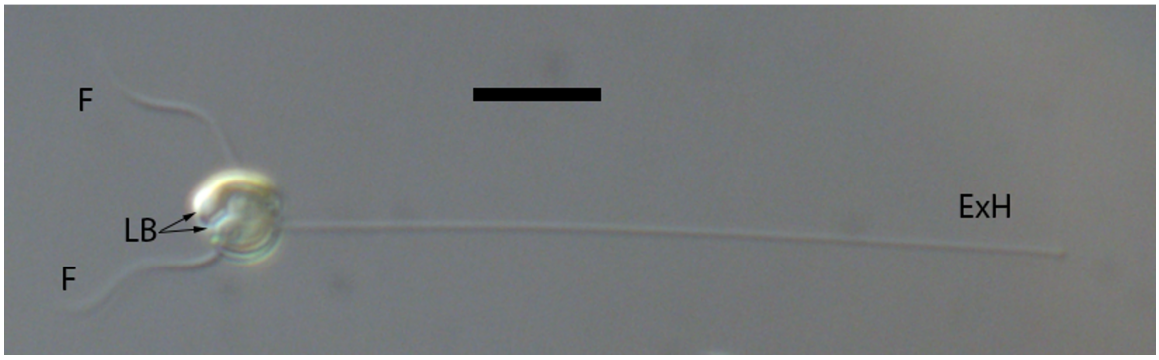

Fig 2: **Light microscopic image of *Chrysochromulina parva*:** There are two equal flagella (F), and a long, retractable haptonema extending from the cell (ExH). One lipid body (arrows) is associated with each chloroplast. Bar = 5 μm.

Table 1: General characteristics of the *Chrysochromulina parva* genome

Assembled genome size	65.8 Mb
Sequencing coverage	110x
Assembled contigs	8,362
Average contig size	7,865 bp
N50 / L50	16,048 bp / 1243 contigs
Contigs > 100kb	1
Contigs > 10kb	2,199
GC content	63.60%
Chloroplast genome size	104,520 bp (complete)
Mitochondrial genome size	24,009 bp (partial)
Nuclear genome	
Protein coding genes	28,138
Average gene length	1,765 bp
Average CDS length	1,506 bp
Average exon length	762 bp
Average exons per gene	2
Average intron length	264 bp
Average introns per gene	1
Chloroplast genome Protein coding/tRNA/rRNA genes	113/27/6
Mitochondrial genome Protein coding/tRNA/rRNA genes	20/25/2

2.1.2 Nuclear genome – ploidy and sexual cycle

Given the extensive distribution of alternative life history phases among the haptophytes, the haplo-diploid reproductive strategy is considered to represent an ancestral attribute in this taxon [33]. Regardless of ancestry, the multi-phasic life history profiles of haptophytes can be genetically complex and morphologically complicated. Many haptophytes display a change in colony formation, scale ornamentation, or motility that reflects a change in ploidy state (Table 2 [20, 34, 35]). No similar shifts in *Chrysochromulina* morphology or swimming behavior have yet to be observed in the two fresh water isolates, even though cells have been subject to an extremely broad range of culture conditions during extensive physiological studies. In contrast to many algae that are constrained to a non-sexual life cycle, having lost those proteins needed to ensure meiotic function (e.g., the commercially exploited genus *Nannochloropsis* [36]), both *C. tobinii* and *C. parva* appear to have retained a full complement of nuclear-encoded genes that produce proteins critical to meiotic processes (Supplementary Table S1; Supplementary Table S2) in their haploid genome. BLAST searches for core meiotic gene orthologs [37, 38] in *C. tobinii* and *Emiliania huxleyi* (which is known to undergo a sexual cycle) reveals that these two haptophytes share the same complement of meiosis specific core genes – many of which are found in model eukaryotic systems with well-studied reproductive systems. Given the large meiosis-related gene complement that has been identified in both *Chrysochromulina* isolates, there is high probability that these organisms have retained sexual reproductive capacity. Although mating has not been observed in our *Chrysochromulina* cultures, it may simply be that the process occurs rarely and quickly or that the physiological parameters needed to induce sexuality have yet to be identified.

In addition to identifying meiosis specific orthologs, genes associated with DNA repair pathways that utilize homologous recombination have also been identified (Supplementary Table S1). The fact that many DNA repair orthologs are also found in *C. tobinii* and *C. parva* suggest that

genetic modification methods that rely on homologous recombination are likely viable in this alga. Genome editing tools such as mega-nucleases, TALE receptor nucleases (TALENs) and CRISPR/Cas9 mediated technologies [39] that rely on endogenous recombination mechanisms for targeted gene modification are present.

Table 2: Ploidy of various haptophyte genomes

Organism	Taxon	Clade#	Haploid	Diploid	Reference
Chrysochromulina tobinii	Prymnesiales	B2	+ (59Mb)$	?	Hovde et. al. 2015 [24]
Chrysochromulina parva	Prymnesiales	B2	+ (67Mb)$	?	This publication
Prymnesium polylepis (alpha)*	Prymnesiales	B1-5	+ (230Mb)	+	John et. al. 2010 [40]
Phaeocystis antarctica	Phaeocystales	A	+ (117Mb)	+ (215Mb)	Vaulot et. al. 1994 [34]
Phaeocystis (Naples isolate)	Phaeocystales	A	+ (176Mb)	+	Vaulot et. al. 1994 [34]
Phaeocystis (North European isolate)	Phaeocystales	A	+ (205Mb)	+	Vaulot et. al. 1994 [34]
Coccolithus pelagicus	Coccolithales	C	+	+	Edvardsen and Vaulot 1996 [41]
Calcidiscus leptoporus	Coccolithales	C	+	+	Edvardsen and Vaulot 1996 [41]
Coronosphaera mediterranea	Syracosphaerales	C	+	+	Edvardsen and Vaulot 1996 [41]
Emiliania huxleyi	Isochrysidales	C	+ (167Mb)$	+	Read et. al. 2013 [18]

Ploidy and genome size estimation of various haptophytes, with current clade designation within Haptophyta. *Chrysochromulina tobinii* has a much more compact genome than the other haptophytes listed here.
* recently revised (*Chrysochromulina polylepis*)
$ genome size from sequencing
+ present
? unknown

2.2 Mitochondrial genome:

The sequenced *Chrysochromulina parva* mitochondrial genome (Fig. 3; Supplementary Table S3; GenBank: MG520332) is 24,009 bp in length, which is smaller than that of *C. tobinii* (34,288 bp). Like *C. tobinii* [23], the *C. parva* mitochondrial genome contains a large internal repeat structure. Given the extensive and complex structure of the repeat region (9.5 kb in *C. tobinii*), this domain was not fully assembled in *C. parva*. *C. parva* and *C. tobinii* mitochondrial genomes encode 20 and 21 proteins respectively, none of which contain introns, or intergenic repeats. The ribosomal operon is split into separate 16S and 23S domains.

Chrysochromulina parva and *C. tobinii* mitochondrial genomes are co-linear in gene profile (Fig. 3). However, the large open reading frame (*orf*456) that lies between the complex repeat region and *nad*4 is present in *C. tobinii* but not in *C. parva*. Though the function of this gene is unknown, transcriptome and comparative genomic analysis indicates it is expressed in *C. tobinii.* The translated protein sequence does not have a homolog found in public reference protein databases. Only a single conserved Domain of Unknown Function (DUF4143, Pfam13635) is noted as a "domain …almost always found C-terminal to an ATPase core family" [42].

Single nucleotide polymorphisms (SNPs) appear between *C. parva* and *C. tobinii* in all 20 protein coding genes (Table S3), and ribosomal RNA genes. Nine of the 25 transfer RNA genes also contain one or more SNPs. The highest number of synonymous changes are observed in *nad* as well as *cox* genes, and the most non-synonymous changes are in *rps*14. dN/dS calculations suggest stabilizing selection for these changes (Table S3). Post-transcriptional modification, which could complicate this computation has not been observed for these genes.

Sequence analysis of the marine species *Chrysochromulina* NIES-1333 shows a very different mitochondrial structure from the *C. tobinii* and *C. parva* isolates described above. The NIES-1333 genome (34,291 bp) is missing a large, complex repeated sequence array; encodes two small intergenic repeats of 1,624 bp and 1,630 bp; contains two genes having group II introns (*cox*1 and *rnl*), and has a highly rearranged gene order [43].

Mitochondrial genomes often serve as windows of evolution, giving insight to the relatedness among phylogenetic groups and likely reflects a program of genetic restructuring via the evolutionary process of endosymbiotic gene transfer (EGT [44,45]). For example, a significant reduction occurs among different algal taxa (Table 3) in the number of genes that contribute components to the algal respiratory chain supramolecular Complex 1 (Fig. 4) which is comprised of ~45 proteins assigned to P, N and Q functional modules [46-48]. The mitogenome of haptophytes, glaucophytes, rhodophytes, cryptophytes and stramenopiles appear to universally retain all genes of the Complex 1 "P-module" (proton pumping), that include NAD 1 through 6. These proteins are highly hydrophobic [46,49] and comprise the peripheral arm of Complex 1 that is membrane localized (Fig. 4). Genes "lost" to the nucleus in some algal lineages, but not others, include the extrinsic proteins [49] associated with modules N (electron capture from NADH oxidation) and Q (electron transfer to ubiquinone). For example, in stark contrast to the cryptophytes that retain *nad* genes 7, 9, 10, and 11, all haptophytes and rhodophytes have lost these 4 mitochondrial-encoded genes. In those algae that retain *nad*7, 9 and 11 genes (e.g., all stramenopile taxa and glaucophytes), the *nad*10 gene is lost.

Insight to the mechanism of "stepwise" mitochondrial to nuclear EGT may be gained by comparing *nad*11 profiles among taxa, given that this gene appears to be progressively targeted for loss from the mitochondrial genome. The *nad*11 gene is comprised of two functional domains: the 5' end serves as binding sites for iron-sulfur clusters while the 3' terminus is similar to an

enzyme having a molydopterin binding activity [50-52]. Raphidophytes, pelagiophytes, dictyochophytes, synurophytes and chrysophytes encode the intact *nad*11 gene. Although, diatoms (bacilliariophytes) may encode the entire *nad*11 gene (e.g., *Thalassiosira pseudonana; Synedra acus*), *nad*11 in *Phaeodactylum tricornutum* is encoded by two adjacent genes that comprise either the 5' or the 3' domains of the full gene sequence [52, 53]. Fracture of NAD 11 into two functional domains has resulted in classes of algae that lack either the 5' (e.g., glaucophytes, and eustigmatophytes) or 3' (phaeophytes) region of *nad*11, or as noted above for haptophytes and rhodophytes, has been completely eliminated from the mitochondrial genome. Though discussion of the acquisition of photosynthesis in the "red" lineage of algae remains controversial [23, 54-61] we suggest that data concerning symbiotic *host cell* genetic footprints (as evidenced by mitochondrial genomic signature) may be of interest, especially in discussing haptophyte and cryptophyte relatedness.

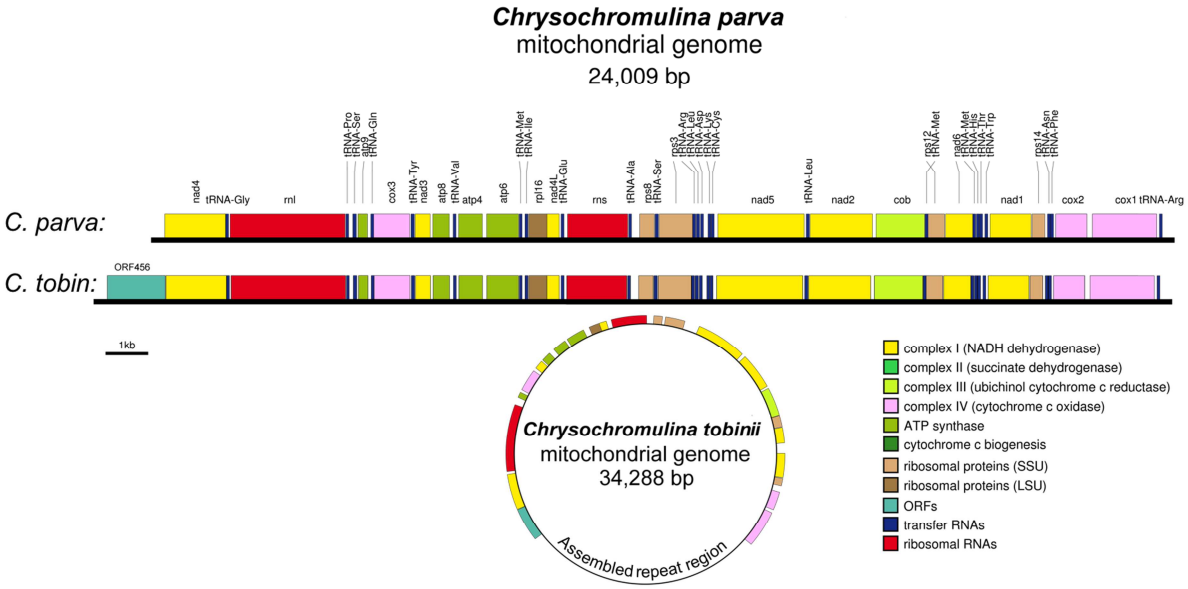

Fig. 3: Comparison of *Chrysochromulina* mitochondrial genome size and gene placement:

The *C. tobinii* genome was previously completed using a combination of high throughput sequencing and Sanger sequencing to assemble the entire repeat region of 9,434 bp [23]. Because short read sequencing was exclusively used in this study, only the coding regions of the *C. parva* mitochondrial genome were assembled. Discounting 987 SNPs within genes and intergenic regions, the overall structure and gene order of the two genomes are the same; the exception being ORF456 that was identified flanking *nad*4 in the *C. tobinii* genome but missing in the *C. parva* genome. The *C. parva* cell contains an estimated 17 copies of the mitochondrial genome per cell based on average read depth of the nuclear and mitochondrial genome assemblies. *C. tobinii* contained an average of 14 mitochondrial genome copies.

Table 3: Reduction in mitochondrial-encoded *nad* genes among different algal taxa

Taxonomic Group	***nad* protein subunit***				
	1 to 6	**7**	**9**	**10**	**11**
Rhodophyte	+	-	-	-	-
Haptophyte	+	-	-	-	-
Cryptophyte	+	+	+	+	+
Glaucophyte	+	+	+	-	Missing 3' end ~120 amino acids, plus several internal deletions
Stramenopile					
Bolidophyceae	N.D.	N.D.	N.D.	N.D.	N.D.
Bacillariophyceae	+	+	+	-	Variable: intact to fragmented
Synurophyceae	+	+	+	-	+
Chrysophyceae	+	+	+	-	+
Eustimatophyceae	+	+	+	-	Missing 5’
Phaeophyceae	+	+	+	-	Missing 3’
Phaeothamniophyceae	N.D.	N.D.	N.D.	N.D.	N.D.
Xanthophyceae	N.D.	N.D.	N.D.	N.D.	N.D.
Raphidophyceae	+	+	+	-	+
Pinguiophyceae	+	+	-	-	+
Dictyochophyceae	+	+	+	-	+
Pelageophyceae	+	+	+	-	+

* + (gene present); - (gene absent); N.D. (gene occurrence not determined)

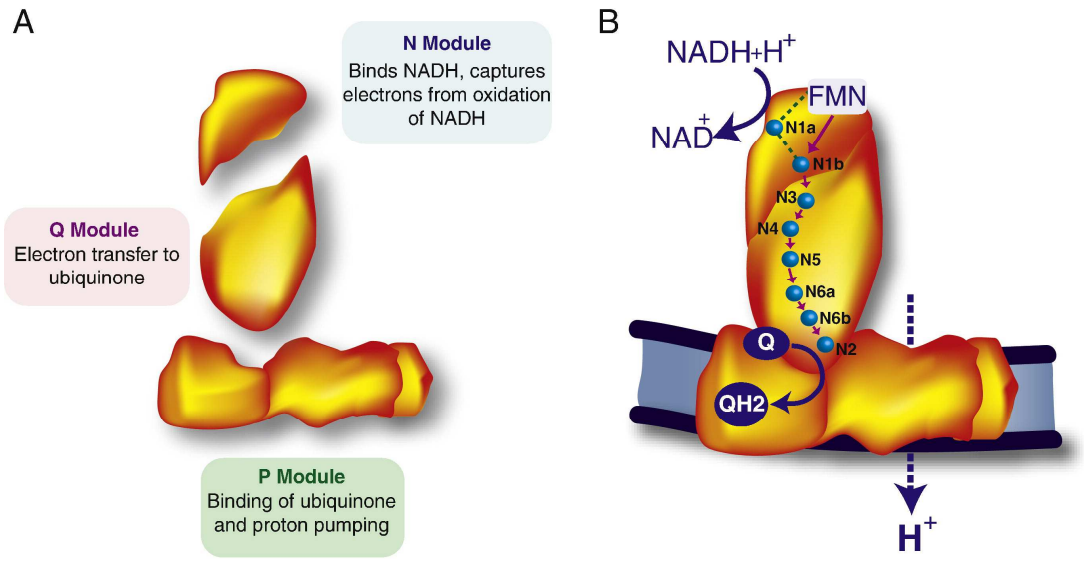

Fig. 4: Schematic of mitochondrial Complex 1 structure. A) Mitochondrial Complex1 is comprised of N, Q and P Modules that B) serve as a conduit for electron transfer and proton pumping processes. These schematics are reproduced from Lazarou et. al. 2009 [49] with permission.

2.3 Chloroplast genome:

The chloroplast genomes of *C. parva* (Fig. 5; Supplementary Table S4; GenBank: MG520331) and *C. tobinii* are 104,520 and 104,518 bp in size respectively. Both genomes encode 112 proteins, 2 ribosomal operons, and 27 tRNAs. A novel restructuring of the conventional chloroplast ribosomal operon within an "inverted repeat" region is seen in both algal isolates. First, the intergenic regions that lie between 23S rRNA and 16S rRNA genes are not identical. The $tRNA_{ala}$ gene is found in one operon, while its sister rRNA operon encodes $tRNA_{ileu.}$ (conventionally, $tRNA_{ala}$ and $tRNA_{ileu.}$ are both present in the intergenic domain). Secondly, exclusive of the tRNA genes, both *C. parva* and *C. tobinii* display significant numbers of SNPs when the two ribosomal "repeats" within a single genome are compared. Analysis of repeated ribosomal genes in "red lineage" plastids reveals that haptophytes, rhodophytes, cryptophytes, and stramenopiles display this non-identity feature in the "inverted repeat" [23]. Similar to previous studies [62-64] the number of SNPs between *C. parva* and *C. tobinii* chloroplast genes (Supplementary Table S4) is significantly less than that seen for the mitochondrial genome. Of the 41 SNPs between *C. tobinii* and *C. parva* chloroplast genomes, 12 occurred within genes and 29 intergenically (0.39 mutations per kb), while among the 987 mitochondrial SNPs observed, 554 occurred within genes and 433 intergenically, (39.6 mutations per kb), when disregarding the large repeat region that remains unassembled in the *C. parva* mitochondrial genome. Similar to the mitochondrial genomes, dN/dS calculations (Supplementary Table S4) suggest stabilizing selection for chloroplast genes. No post-transcriptional editing was noted.

The *Chrysochromulina* cell contains only two chloroplasts. Thus, unlike some algae whose plastid complement is high, where chloroplast and cell division processes can be uncoupled [65] and the distribution of daughter plastids is stochastic [66], the regulation of *Chrysochromulina* chloroplast division at each cell duplication cycle is essential for survival. The process of chloroplast division requires a complex array of protein machinery [67-69]. However, it appears

that whether a cell contains many chloroplasts or few, the genes controlling plastid division have been prime nuclear relocation targets during the evolution of host/symbiont chimeric construction. For example, within the red lineage of algae (haptophytes, rhodophytes, cryptophytes and stramenopiles), only two genes of the chloroplast division complex (*min*D and *min*E) may be found in the plastid. The *min*D gene is found in all haptophytes. The cryptophytes *Guillardia theta* and *Rhodomonas salina* (Pyrenomonadales) also encode the gene in the chloroplast, whereas *Cryptomonas paramecium* (Cryptomonadales) does not. An interesting distribution, perhaps reflecting the wide differences (e.g., phycobiliprotein composition; morphology) noted between these two taxa [70]. The *min*D gene is not encoded in rhodophyte and in stramenopile chloroplast genomes. The second gene, *min*E is solely found in the chloroplast of the cryptophytes *Guillardia* and *Rhodomonas*, but not in any other algal taxa examined to date (including haptophytes).

The division mechanism(s) needed for complex plastids (e.g., 3 or 4 membrane-delineated chloroplasts; glaucophytes, dinoflagellates, haptophytes and stramenopiles) is certainly more complicated than that needed for primary plastids (e.g., 2-membrane enclosed chloroplasts; rhodophytes, chlorophytes). For example, early EM studies [71, 72] report chloroplast division in the stramenopile *Heterosigma akashiwo* (raphidophyte) involves a two-step process wherein the inner 2 membranes (chloroplast envelope) divides before the outer two membranes (chloroplast ER). Thus, one might speculate that the division machinery for plastid might be specialized for each membrane set; that the inner 2 membranes retain the division machinery common to all 2-membrane chloroplast types, whereas the membranes comprising the chloroplast ER have evolved a separate set of partitioning tools for effecting plastokinesis. For example, the nuclear-encoded *fts*Z gene family represents a universal player in the chloroplast division process of both primary and complex plastids within all taxa examined to date [71-77]. The *fts*Z gene produces tubulin-like filaments (cytoskeletal GTPase) that form the dividing ring which is localized on the

stromal side of the most interior set of chloroplast membranes. Chloroplast encoded proteins MinD and MinE, also localized to the chloroplast envelope, modulate the FtsZ filament assembly process [78]. Phylogenetic studies show specialization (sub-functionalization; neo-functionalization?) of FtsZ is quite evident among a broad spectrum of eukaryotic photosynthetic organisms (Fig. 6). Most interestingly, all haptophyte FtsZ proteins, with the exception of members of the *Pavlova* species, lack a C-terminus extension (Fig. 6, Supplementary data S1), which has been implicated in protofilament bundling and anchoring to the inner envelope membrane [74]. The haptophyte FtsZ (excluding the poorly supported Pavlova FtsZ proteins) proteins form two unique clades (putatively named FtsZH and FtsZD-3). The haptophyte FtsZD-3 clade is highly supported as having a common ancestral protein sequence related to the chloroarachniophyte FtsZD gene pairs, while the FtsZH proteins are earlier diverged, bringing into question at what point FtsZ gene duplications took place with both the chloroarachniophyte and haptophyte clades. Unfortunately, detailed mechanistic studies concerning the chloroplast ER membrane division for the complex plastids of haptophyte and stramenopile taxa are lacking. New genomic information allowing the identification of complex chloroplast division components should certainly facilitate these efforts. Recent reports of chloroarachniophyte (four membranes encompassing a green algal- like plastid) FtsZ protein function are noteworthy [79] and may aid in explaining the two gene copy system seen in Haptophyta.

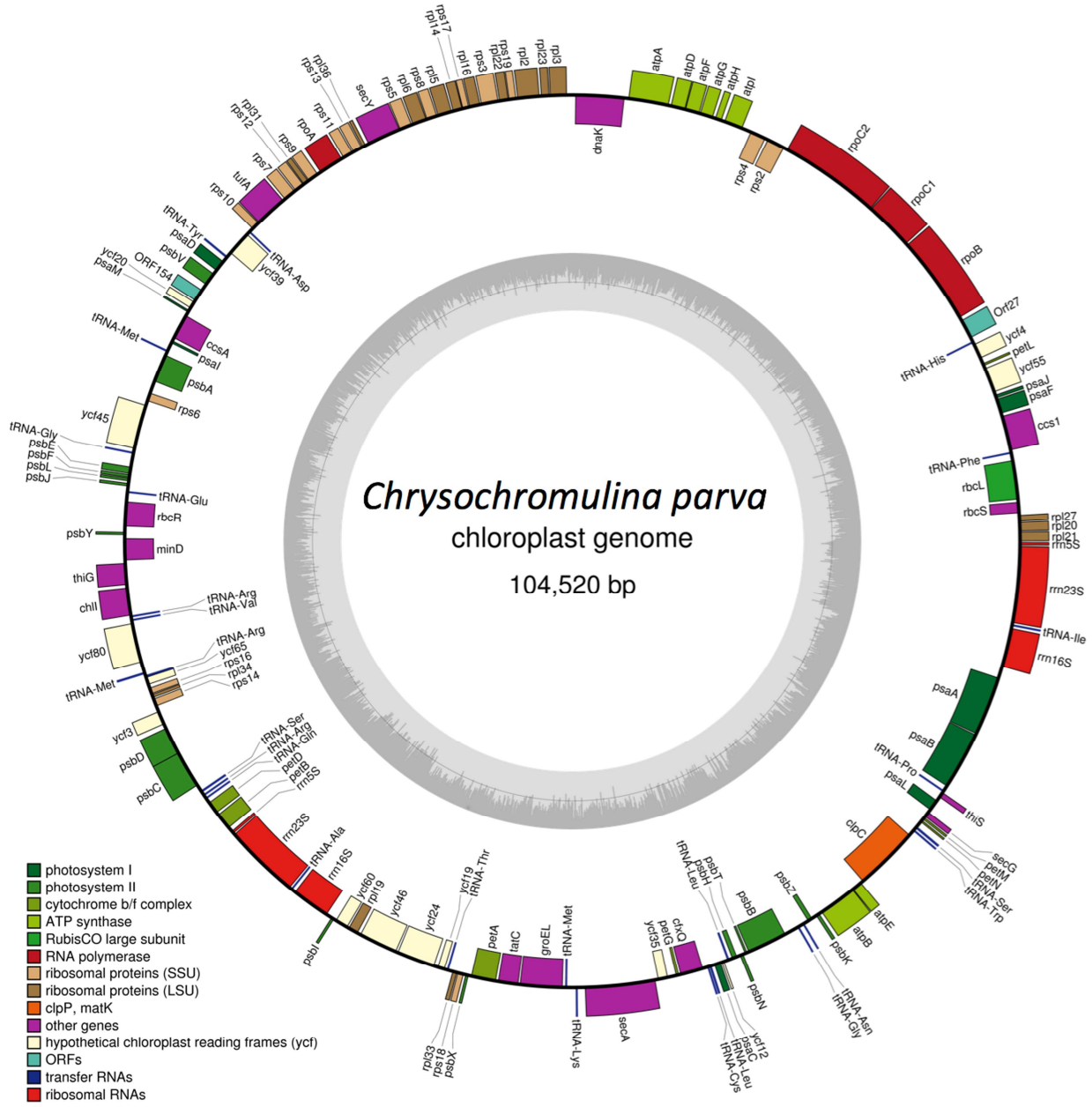

Fig. 5: *Chrysochromulina parva* chloroplast genome: Each cell contains approximately 11 copies of the chloroplast genome.

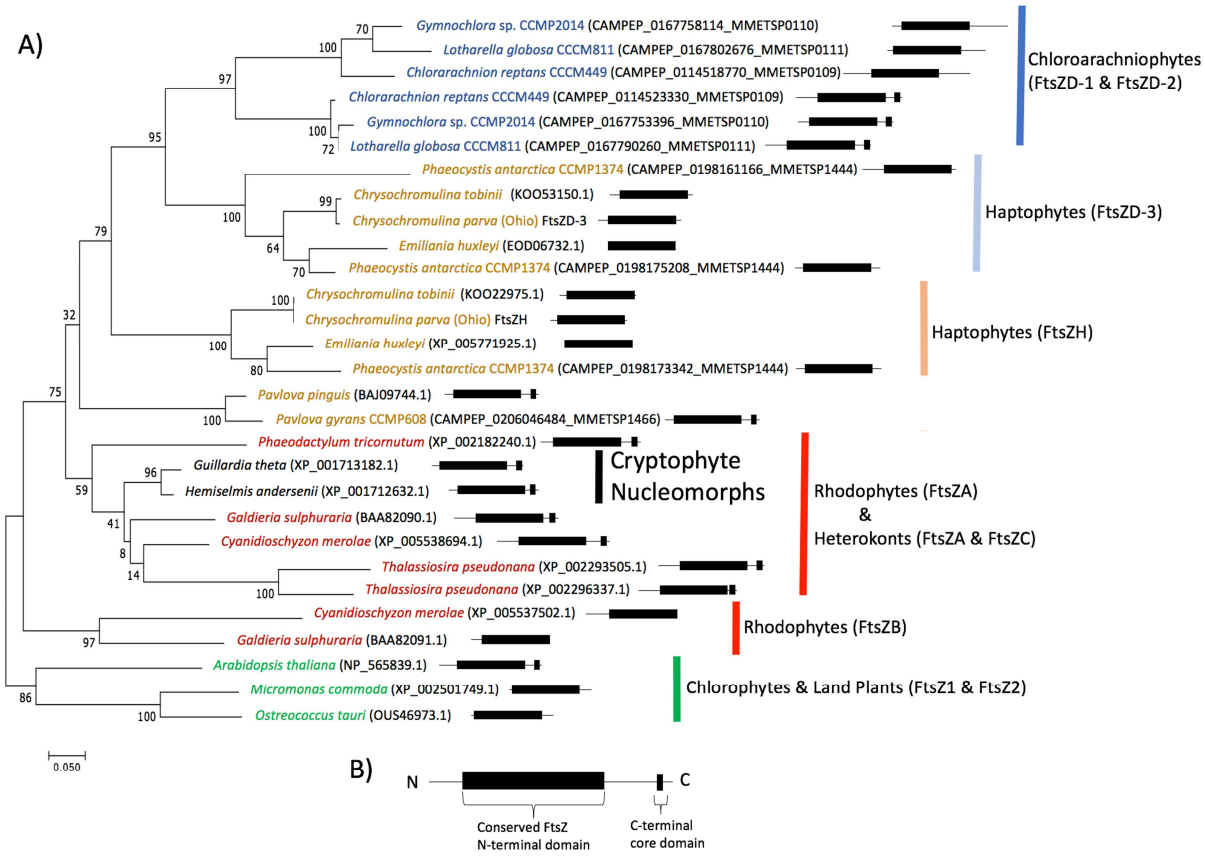

Fig 6. Molecular Phylogenetic analysis of FtsZ: A) Evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [80]. The tree with the highest log likelihood (-7050.62) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated (fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position). There were a total of 289 positions in the final dataset. Analyses were conducted in MEGA7 [81]. B) An approximately 300 amino acid FtsZ domain is highly conserved amongst the FtsZ protein sequences compared in this analysis. Outside of that 300 bp region however, the N and C termini are highly variable and approximate schematics representing the size of the variable regions at

each termini and the presence or absence of a C-terminal core domain as described by Miyagishima et al. 2004 [82] are displayed next to each species within the tree.

2.4 Chrysochromulina/bacterial biome identities

Algal-bacterial biome interdependence has long been recognized. It was of interest to determine if the bacteria associated with *C. parva* and *C. tobinii* isolates (which have different fresh water origins, namely a lake in Colorado and a river in Ohio respectively) associate with similar bacterial taxa. Notably, data show that the most highly represented species within the bacterial biomes of these two algae have similarities (Table 4). The bacterial biome of the bacterized *C. tobinii* culture (P3) consists of nine bacterial species. Eight of these bacterial cohorts have now been isolated, individually cultured, and identified both by 16S rRNA PCR and their full genomes sequenced [32]. As seen in metatranscriptomic data [83] collected for the *C. tobinii* P3 culture, the two bacterial taxa with the highest levels of detectible transcription were *Acidovorax* sp. RAC01 and *Hydrogenophaga* sp. RAC07 - both within Comamonadaceae. All six species of highly represented bacteria identified in the *C. parva* biome are within the Comamonadaceae family. These data certainly suggest that fresh water *Chrysochromulina* might selectively associate with certain metabolic bacterial partners, but also have novel consorts.

Table 4: Bacterial biome associated with two fresh water *Chrysochromulina* isolates.

Algal Isolate	Isolate genome or metagenome sequencing bacterial ID	Proteobacteria: Class; Order; Family
Chrysochromulina tobinii (P5) (cleaned culture)	*Sphingomonas* sp.	*Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae*
Chrysochromulina tobinii (P3)	*Acidovorax* sp. RAC01 *Hydrogenophaga* sp. RAC07	*Betaproteobacteria; Burkholderiales; Comamonadaceae*
	Agrobacterium sp. RAC06 *Sinorhizobium* sp. RAC02	*Alphaproteobacteria; Rhizobiales; Rhizobiaceae*
	Blastomonas sp. RAC04 *Sphingobium sp.* RAC03	*Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae*
	Bosea sp. RAC05	*Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae*
	Methyloversatilis sp. RAC08	*Betaproteobacteria; Nitrosomonadales; Sterolibacteriaceae*
Chrysochromulina parva	*Acidovorax* sp. *Hydrogenophaga* sp. *Delftia* sp. *Alicycliphilus* sp. *Ramlibacter* sp. *Variovorax* sp.	*Betaproteobacteria; Burkholderiales; Comamonadaceae*

2.4.3 Phylogeny:

SNP variation was assessed between *C. parva* and *C. tobinii*, as well as between these two isolates and the fresh water Japanese isolate *Chrysochromulina* strain NIES-562. As shown in Table 5, the mitochondrial genes *cox*1 and *nad*5 differ extensively (38 and 26 SNPs, respectively) between *C. parva* and *C. tobinii*. Interestingly, *C. parva* and *Chrysochromulina* NIES-562 are quite similar, with only 3 (*cox*1) and 2 (*nad*5) differences between them. SNPs present in 18S and 28S rRNA sequences again show greater identity between the *C. parva* and the Japanese isolate. Extending this study, several nuclear-encoded gene sequences were compared between *C. parva* and *C. tobinii* (Table 5; Table S5). SNP differences was variable ranging from marginal nucleotide divergence in sequence fidelity (e.g., RuBisCO activase; 1 SNP in 1245 bp) to highly diverse (e.g, xanthorhodopsin; 25 SNPs and a 3 bp deletion).

Unfortunately, a paucity of non-ribosomal sequences for *Chrysochromulina* species occurs in gene databases. For this reason, the relationship between *C. tobinii*, *C. parva* and *Chrysochromulina* NIES-562 was assessed phylogenetically using the 18S rRNA locus. We note here, a discrepancy between our data (guanine) and the published 18S rRNA $CCMP291_{NCMA}$ sequence (GenBank:AM491019.2) (deletion) at position 181. Our 18S rRNA sequence was used to construct the phylogeny described below. Bayesian and maximum likelihood analyses show *C. parva*, *C. tobinii* and *Chrysochromulina* NIES-562 to be sister species co-occurring within the haptophyte clade B2 (as defined in Edvardsen et. al. 2011 [30]; Fig. 7; Table S6). This sister relationship was also supported by 28S rRNA locus comparison (data not shown). As seen by the branch lengths, *C. parva* and *C. tobinii* demonstrate as much sequence dissimilarity at the 18S rRNA locus as other recognized closely related *Chrysochromulina* species pairs [30] in this clade (e.g., *Chrysochromulina acantha* and *Chrysochromulina throndsenii*). Importantly, this study and that of others have shown that ribosomal genes may be not be the best choice for resolving species identities [61, 62, 84-86]

Table 5: Comparison of SNPs found among three fresh-water *Chrysochromulina* isolates

	C. tobinii vs. *C. parva*	*C. tobinii* vs. NIES-562	*C. parva* vs. NIES-562
Gene			
*nad*5	26 (794)	25 (794)	2 (794)
*cox*1	38 (697)	37 (697)	3 (697)
*rbc*L	0 (1008)	2 (1008)	2 (1008)
18S rRNA	3 (1435)	3 (1435) and 1 deletion/insertion	0 (1435)
28S rRNA	2 (736) and a single 1bp deletion/insertion	2 (736) and a 1bp deletion/insertion	0 (736)
Xanthorhodopsin (KOO22837.1)	25 (789) and a 3bp deletion/insertion	N.A.	N.A
NADPH:adrenodoxin oxidoreductase KOO53708.1	10 (1788)	N.A.	N.A.
Erythromycin esterase (KOO24049.1)	3 (1572)	N.A.	N.A.
Lycopene beta cyclase (KOO22265.1)	14 (1746)	N.A.	N.A.
RuBisCO activase (nuclear genome) (KOO53179.1)	1 (1245)	N.A.	N.A.
Methionine synthase reductase (KOO27933.1)	3 (960)	N.A.	N.A.

SNP identification in various genes between *Chrysochromulina tobinii*, *Chrysochromulina parva* and NIES-562 [# of SNPs (total alignment length)]. Nuclear genome sequence for NIES-562 is not available (N.A.) for comparisons between NIES-562 and the other two *Chrysochromulina* species.

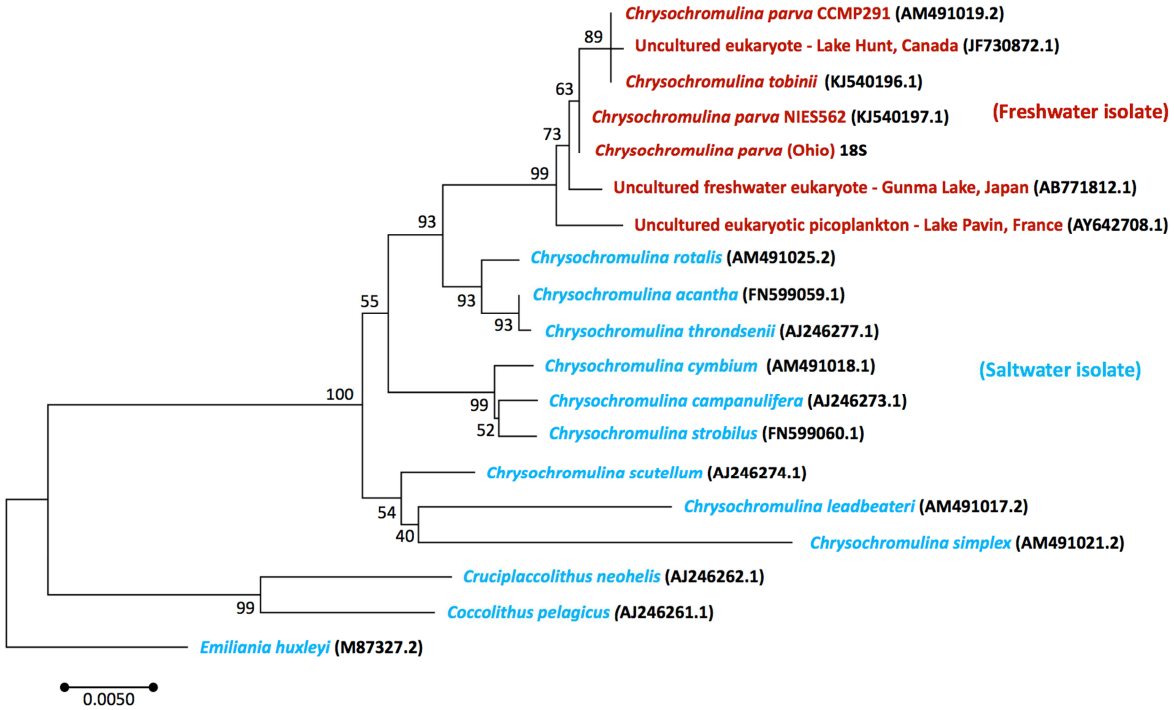

Fig. 7: Neighbor-joining tree of B2-clade Prymnesiophytes*: Result is based on a 1148 bp alignment of 18S rRNA sequences. *Cruciplacolithus neohelis*, *Coccolithus pelagicus* ssp. braarudii and *Emiliania huxleyi* were used as outgroups. The evolutionary history was inferred using the Neighbor-Joining method [87]. The optimal tree with the sum of branch length = 0.14533780 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [88]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [89] and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7 [81]. Additional details of each named species used in this tree are available in Supplementary Table S6.

*****as defined in Edvardsen et. al 2011 [30]

2.5 Taxonomy and Nomenclature:

Chrysochromulina parva Lackey, collected from Big Walnut Creek, Ohio, USA, was initially described as a naked organism that had three flagella [26]. Electron microscopic studies of similar cells revealed that the third flagellum was actually newly recognized structure - the haptonema [27].

Subsequently, three additional species were described, i.e., *C. breviturrita* [90], *C. laurentiana* [91], and *C. inornamenta* [92]. There were also two other organisms that were identified as *C. parva* [28, 93]. All these organisms were embellished with scales. Concurrently, many marine dwelling *Chrysochromulina* species were also described (see Chretiennot-Dinet et al., 2014 for overview [25]), and they all had scale coverings. Further studies ultimately lead to a subdivision of the genus into additional genera [30].

Until our study, no one had returned to Big Walnut Creek, the type locality for the type species. Our *C. parva* isolate (strain UW 1161) was found at the type locality during the same month that Lackey (1939) collected the alga [26]. Electron microscopic analysis of this isolate shows that it has no scales (Figs 1, 2). We also document that two additional freshwater species, *C. tobinii* sp. nov., sourced from a Colorado lake (USA), and an unnamed *Chrysochromulina* sp. obtained from the Fox River, Illinois (USA) also lack scales [29]. Given these observations, we conclude that the type species of *Chrysochromulina* lacks scales and that at least one additional scale-less species exists.

In a review of the North American haptophytes, Nicholls [94] transferred three of the scaled freshwater species to the genus *Prymnesium* [*P. breviturrita* (Nicholls) Nicholls; *P. luarentianum* (Kling) Nicholls; *P. inornamenta* (Wujek & Gardiner) Nicholls]. Unfortunately, no molecular

data exist for these species. We posit that the scaly organisms described by Parke et al. [28] as well as Thompson and Halicki [93] were misidentified and belonged to one or two new, undescribed species that remain unnamed.

Substantial diversity exists among these small flagellates and we conclude that strain UW 1161 from Big Walnut Creek is the best representative for *C. parva* because it was collected at the type locality. We also find that strain NIES-562, based upon molecular data (this paper), is a second representative of *C.* parva. Finally, we conclude that strain CCMP291 is an undescribed species. Morphological comparisons of these three freshwater, naked *Chrysochromulina* isolates will be presented elsewhere [29].

Chrysochromulina parva is the type species for this large genus, and the type material for *C. parva* anchors not only this species but higher ranks (e.g., genus, family names). Unfortunately, Lackey (30) did not designate a holotype, and no subsequent scientist has established a lectotype. Therefore, we establish type material as follows:
Chrysochromulina parva Lackey 1939: **Lectotype here designated: Fig. 23**, page 135, in Lackey, *Lloydia* 2 (1939).

Because the lectotype is an illustration, no molecular data can be retrieved, and therefore we conclude that the lectotype is ambiguous. As a consequence, we establish an epitype as follows:
Chrysochromulina parva Lackey 1939: **Epitype here designated:** cells from strain UW 1161 were preserved as a TEM block and deposited in the New York Botanical Garden herbarium (NY), New York City, NY USA as **No. _____.** Strain UW 1161 was established using cells collected from the type locality (Big Walnut Creek, Shadeville, OH USA (39º 49’ 59.58” N, 82º 59’ 334.64” W).

Strain UW 1161 is available from the Department of Biology, University of Washington, Seattle, WA USA.

Our examination of strain CCMP291 shows that is genetically distinct from strain UW 1161, which was used to establish the *C. parva* epitype. Because strain CCMP291 also lacks scales, it can be distinguished from all species of *Chrysochromulina* that are encased with scales. We therefore describe this organism as a new species as follows:

***Chrysochromulina tobinii* sp. nov. Cattolico**

Diagnosis: single celled flagellate; cell body 4-7 μm wide, 5-8 μm long; two flagella, each approx. 10 μm long; haptonema up to 10x the cell length long; cells naked, without scales; two chloroplasts, each with a pyrenoid; no eyespot; cysts or resistant stage unknown; genome NCBI JWZX00000000 with distinctive DNA sequences.

Holotype designated here: cells from strain CCMP291 were preserved as a TEM block and deposited in the New York Botanical Garden herbarium (NY) as **No. _____**.

Isotype designated here:

cells from strain CCMP291 were preserved as a TEM block and deposited in the New York Botanical Garden herbarium (NY) as **No. _____**.

Etymology: the name honors Tobin Cattolico.

Type locality: unspecified lake, Colorado USA

Authentic culture: Strain CCMP291 is available from the Provasoli-Guillard National Center for Marine Algae and Microbiota, East Boothbay, Maine, 04544 USA.

Highlights:

1.) Genetic assessment supports revisiting the Diagnosis for the *Chrysochromulina* clade.

2.) *Chrysochromulina* isolates often show extensive morphological identity but they differ genetically; thus, cryptic species complexes are expected in this taxon.

3.) Fresh water *Chrysochromulina* isolates appear to have all the genes that insure sexual reproductive capability.

4.) The genomic assessment of two *Chrysochromulina* isolates provides a primary data base for probing the functional specialization and evolutionary process among haptophytes as well as among major algal lineages.

3. Materials and Methods

3.1 Algal Sourcing and Culturing Conditions:

Chrysochromulina tobinii (CCMP291) was acquired from the NCMA (National Center for Marine Algae and Microbiota - East Boothbay, Maine) by our laboratory in 2006. This bacterized isolate was designated as P3. To remove associated bacteria, cells from the P3 cultures were subject to re-iterative flow cytometry in 2009 using BODIPY 505/515 (4,4-difluoro-1,3,5,7-tetramethyl-4-bora-3a,4a-diaza-s-indacene; Invitrogen, Carlsbad, CA; see below for staining procedure) as the fluorophore for cell sorting purposes. Cells obtained from reiterative flow cytometric selection (P5.0) were treated in RAC-1 proprietary medium that contained either streptomycin (P5.5) or hygromycin (P5.6). Treatment for these two antibiotics were identical: cells were exposed to a final concentration of 400 μg/mL antibiotic for 18 hours before 5 mL of treated cultures was transferred to 100 mL of RAC-1 medium lacking antibiotic. Cultures P5.5 and P5.6 were periodically tested for bacterial contamination using Luria-Bertani medium [95] made with RAC-1 medium. Determination of single bacterial contaminant was achieved by plating P5 cells on medium containing 0.1% glycerol. To test whether our P3 and P5 laboratory maintained isolates were genetically identical to the parent isolate that had been maintained in the NCMA culture collection, *Chrysochromulina* (CCMP291) was re-ordered from NCMA in 2011. Sequence analyses of nuclear (18S and 28S rRNA), chloroplast (*rbc*L) and mitochondrial (*cox*1, *nad*5) reveal no differences among UW maintained *Chrysochromulina tobinii* P3, P5 isolates and the 2011 acquired *Chrysochromulina* CCMP291 isolate (data not shown).

Chrysochromulina parva (isolate UW 1161) was obtained from a water sample collected on September 24, 2014 at Big Walnut Creek in Shadeville, OH (Latitude 390 49’60” N; Longitude 820 59’ 35” W), the type locality for this species. Water samples were overnight shipped to the University of Washington and the cultures kept at 20°C on a 12 hr light:12 hr dark photoperiod at 30 $\mu Em^{-2}s^{-1}$ light intensity. Subsamples of the culture were gravity filtered through a 100 μm

nylon mesh to minimize the presence of predatory protists. The filtrate containing *Chrysochromulina* was added to RAC-5 proprietary medium. Once the culture was acclimated to growth chamber conditions, *Chrysochromulina* was further separated from contaminating protists via fluorescence-activated cell sorting (April, 2015) at the Institute for Systems Biology (Seattle, WA).

Chrysochromulina cultures were maintained in 250 mL Erlenmeyer flasks containing 100 mL of RAC-5 medium plugged with silicone sponge stoppers (Bellco Glass, Vineland, NJ) and capped with a sterilizer bag (Propper Manufacturing, Long Island City, NY). Alternatively, large volume experimental cultures were maintained in 1.0 L medium that was contained in a 2.8 L large-mouth Fernbach flask. These flasks were plugged with hand-rolled, #50 cheese cloth-covered cotton stoppers and covered with a #2 size Kraft bag (Paper Mart, Orange, CA). All cultures were maintained at 20°C on a 12 hour light:12 hour dark photoperiod under 100 $\mu Em^{-2}s^{-1}$ light intensity using full spectrum T12 fluorescent light bulbs (Philips Electronics, Stamford, CT). No CO_2 was provided and cultures were not agitated. Unless indicated, cultures were sampled at hour ~6 in the light portion of the 12 hr light:12 hr dark photoperiod for assessing cell counts and for recovering aliquots for GC/MS fatty acid analysis.

3.2 Electron microscopy:

3.2.1 Scanning electron microscopy:

Five hundred microliters of concentrated *C. parva* cell culture was mixed with an equal volume of 2.0% glutaraldehyde in 0.2 M sodium cacodylate buffer (pH 7.2) at room temperature, then 250 µl of 4.0% osmium tetroxide was immediately added. The cells were fixed 15 minutes on ice, then filtered onto a 1.0 µm Nucleopore membrane (Nuclepore Corp., Pleasanton, CA), rinsed with 0.1 M sodium cacodylate buffer, then dehydrated to 100% alcohol using (50, 70, 95, 100%). Dehydrated cells on filters were critical point dried (Samdri 790 critical point dryer, Tousimis

Research Corporation Rockville, MD), coated with 6 nm platinium (ES150T coater, Electron Microscopy Sciences, Hatfield, PA) and viewed in a Quanta 450 FESEM (FEI, Hillsboro, OR).

3.2.2 Transmission electron microscopy:

Chrysochromulina parva cell pellets were fixed for 1 hour in 2.0% glutaraldehyde in 0.15 M sodium cacodylate buffer (pH 7.2) at room temperature, then rinsed three times in buffer alone followed by 1.0% osmium tetroxide treatment in 0.15 M sodium cacodylate buffer (pH 7.2) for 1 hour on ice. Pellets were dehydrated in a graded alcohol series (50, 70, 95, 100%), rinsed three times in 100% dry acetone, infiltrated in a graded acetone-EMBed812 series (33, 66, 100%), embedded in 100% EMBed812 (Electron Microscopy Sciences, Hatfield, PA), and polymerized at 60°C. Silver sections were cut with a Diatome diamond knife (Electron Microscopy Sciences, Hatfield, PA), stained with aqueous 2.0% uranyl acetate followed by lead citrate [96], and viewed stained and unstained on a Tecnai 12 TEM (FEI, Hillsboro, OR). Images were recorded on an XR-41S 2k digital camera (Advanced Microscopy Techniques Corp., Woburn, MA).

3.2.3 Light microscopy:

Chrysochromulina parva was observed on a Leica DMRBE light microscope (Leica Microsystems Inc., Buffalo Grove, IL) and images were recorded with a Lumenera Lt425 color camera using LuCam software (Lumenera Corp., Ottawa, Ontario, Canada).

3.3 Cell counts:

Because of the minute size of *Chrysochromulina*, special care was taken in assessing culture density by using a BD Accuri C6 flow cytometer (BD Biosciences, San Jose, CA). The *Chrysochromulina* cells were counted and positively identified against background noise by exciting the samples with a 488 nm laser and detecting chlorophyll autofluorescence with the FL3 (670nm LP) and FL4 (675/25 nm) channel detectors and then isolating the cellular chlorophyll

signal from non-autofluorescent culture debris. Total overlap (by particle count) between gated cell populations and background was less than 1.0%. Expected count error was less than 1.0%.

3.4 Whole genome sequencing, assembly and gene annotation:

The *Chrysochromulina parva* culture was grown to a concentration of approximately 2 x 10^6 cells/mL before harvest for DNA isolation. 800 mL of culture were pelleted in 500 mL Nalgene bottles by centrifugation at 8663 x g at 4°C for 20 min. The cell pellets were flash frozen and stored at -80°C. High quality genomic DNA was prepared using the Qiagen Genomic-tip protocol (Qiagen, Valencia, CA) with modifications [85] and quantified using the Invitrogen Qubit Fluorometer (Life Technologies, Carlsbad, CA). DNA isolations were stored in the -80°C freezer before being sent to Los Alamos National Laboratories on dry ice for sequencing. The *C. parva* genome was sequenced using a combination of shotgun Illumina libraries for 301 bp paired end reads on the MiSeq (totaling ~28 million reads) and 151 bp reads on the NexSeq platform (totaling ~75 million reads). Genome assembly was performed by using IDBA_UD version 1.1.1 [97] on the EDGE platform [98]. Minimum and maximum kmer length were set to 31 and 121, respectively along with step size of 20 and a minimum assembled contig size of 200 bp. The resulting metagenome (*C. parva* + associated bacteria) assembly was approximately 81 million bp contained in 39,594 contigs. To isolate bacterial sequences from *C. parva* sequences, all assembled contigs were queried against the *C. tobinii* genome using BLASTN (version 2.2.28) with default settings. All contigs with significant hits to the *C. tobinii* reference (E-value < .00001) were classified as *C. parva* contigs. Additionally, the remaining "Non-*C. parva*" contigs were queried against the NCBI "nt" database to identify top hit homology. From this analysis, contigs with top hits identified from other algal species (*Emiliania huxleyi, Aureococcus anophagefferens, Chlorella variabilis* and *Micromonas* sp.) were also classified as *C. parva* contigs. A modified MAKER2 pipeline was used to annotate genes and assign functional annotations to genes [99].

3.5 Bacterial Isolation and bacterial cohort classification:

Bacterial cohorts from the *C. tobinii* genome were previously described in Fixen et al [32] via physical isolation methods. To identify the bacterial cohort within the *C. parva* culture, contig assemblies all "Non-*C. parva*" contigs (as described above) were sorted and enumerated by top blast hit. Highly represented organisms (BLASTN Top hit organism names with > 750,000 bp of total alignment length) were listed in Table 4 as the bacterial cohort of *C. parva*.

3.6 Phylogenetic comparison:

Chrysochromulina isolates were grown in 1.0 L cultures in RAC-1 medium. 800 mL of stationary phase culture were harvested for DNA isolation as described above. *Chrysochromulina parva* NIES-562, DNA was extracted from a lyophilized cell pellet obtained from the National Institute for Environmental Studies (Tsukuba, Japan) using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). Gene sequences from *C. parva* strains UW 1161 and NIES-562 and *C. tobinii* were compared at five loci: 18S rRNA, 28S rRNA, *rbcL, cox1,* and *nad5*. PCR and internal sequencing primers (18S rRNA only) and annealing temperatures used in PCR reactions are shown in Supplementary Table S5. The *nad*5 and 28S rRNA PCR primers were designed using Primer3 [100] based on an alignment of *Emiliania huxleyi*, *Phaeocystis globosa*, and other *Chrysochromulina* species sequences that were obtained from GenBank. Recovered DNA was diluted to 3 – 6 ng/μL and PCR was performed using Thermo Scientific Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) with the following reaction mixture: 1x Phusion HF buffer, 200 μM dNTP, 0.5 μM forward and reverse primers, 2 μL template DNA, and 0.02 U/μL Phusion DNA polymerase in 25 μL reactions. The thermal cycling conditions used for PCR reactions were 98°C for 30 s initial denaturation followed by 40 cycles of 98°C for 10 s denaturation, X°C (See Supplementary Table S5) for 20 s annealing, then 72°C for 1 min extension, followed by a final 72°C for 5 min extension. PCR products were cleaned using the

Qiagen QIAQuick Gel Extraction Kit (Qiagen, Valencia, CA) and sequenced using 1/8th reactions of BDT v3.1 (Life Technologies, Carlsbad, CA). Samples were precipitated with 95% and 70% ethanol, dried at room temperature, resuspended in 10 μL of Hi-Di formamide (Life Technologies, Carlsbad, CA), and run on an ABI3130xl Sequencer (Life Technologies, Carlsbad, CA). Sequence analysis was performed using MacVector 11.1.2 (MacVector, Cary, NC). Sequences were trimmed for quality and aligned by ClustalX. Nucleotide variations among *C. parva* strains UW 1161 and *C. tobinii* were deduced from the sequence alignment of each locus.

4.1 List of Abbreviations:

NCMA – Provasoli-Guillard National Center for Marine Algae and Microbiota

NIES – National Institute for Environmental Studies

SEM – Scanning electron microscopy

TEM – Transmission electron microscopy

BODIPY 505/515 – 4,4-Difluoro-1,3,5,7-Tetramethyl-4-Bora-3a,4a-Diaza-s-Indacene

5.1 Competing Interests:

There are no competing interests associated with the publication of this manuscript.

6.1 Authors Contributions:

RAC conceived, designed, and coordinated the study. BTH accomplished whole genome sequencing, and genome data analysis and mining. SRS supervised the whole genome sequencing and assembly and performed genome comparative analyses. CRD assisted in generating a unialgal *Chrysochromulina parva* culture, performed DNA isolation as well as PCR, sequencing, and sequence analysis. RAA field collected the *Chrysochromulina parva* isolate. SBB was responsible for all SEM and TEM analyses. RAC and BTH wrote the manuscript. All authors

read and approved the final manuscript.

7.1 Acknowledgements:

We wish to thank Dr. Masanobu Kawachi from NIES who graciously facilitated our acquisition of the *Chrysochromulina parva* NIES-562 cell pellet and Dr. Anne Thompson at the Institute for Systems Biology Seattle, for conducting the flow cytometric sorting of *Chrysochromulina parva* that facilitated the generation of a unialgal culture. We would also like to acknowledge Johnathan Patterson and William Yost for their assistance in collection the bacterial cohort lipid analysis data. We wish to thank Heather Hunsperger for assistance in planning the 18S rRNA phylogenetic analysis. Funding for BH was provided by the Interdisciplinary Training in Genomic Sciences NHGRI T32-HG00035. Funding for SS was provided through award NL0029949 from the DOE Bioenergy Technologies Office. The authors acknowledge the use of equipment at the San Diego State University Electron Microscopy Facility acquired by NSF instrumentation grants DBI-0959908 (SEM) and DBI-030829 (TEM) to SB. This research was funded by the US Department of Energy under contract DE-EE0003046 awarded to the National Alliance for Advanced Biofuels and Bioproducts to RAC and SS, and NOAA NA070AR4170007 to RAC. RAC dedicates this manuscript to MAC – Navy Seal Team 5 and Vietnam vet on the occasion of his 73rd birthday.

8.1 Supplementary Data:

Supplementary Table S1: Core genes associated with meiosis and DNA repair in eukaryotic cells

	Chrysochromulina tobinii	*Chrysochromulina parva* (Ohio)	*Emiliania huxleyi* CCMP1516	*Chondrus crispus*	*Ectocarpus siliculosus*	*Chlamydomonas reinhardtii*	*Volvox carteri*	*Arabidopsis thaliana*	
Core meiotic genes									Function
*spo*11	+ ($2e^{-33}$)	+ ($2e^{-33}$)	+ ($5e^{-18}$)	+	+	+	+	+	Transesterase
*hop*1	+ ($3e^{-21}$)	+ ($2e^{-20}$)	+ ($3e^{-7}$)	-	+	+	+	+	DNA DSB binding
*hop*2	-	-	-	+	-	-	+	+	Associated with MND1, homology searching
*mnd*1	+ ($2e^{-12}$)	+ ($3e^{-12}$)	+ ($1e^{-13}$)	+	+	+	+	+	DNA heteroduplex formation
*rec*8	-	-	-	+	-	-	-	+	Sister chromatid binding
*dmc*1	-	-	+*	+	+	+	+	+	Inter-homolog recombination
*rad*51	+ ($1e^{-118}$)	+ ($2e^{-118}$)	+ ($6e^{-58}$)	+	+	+	+	+	Homologous DNA pairing
*msh*4	+ ($8e^{-22}$)	+ ($4e^{-25}$)	+ ($5e^{-40}$)	+	+	+	-	+	Holliday junction resolution w/ MSH5
*msh*5	+ ($6e^{-36}$)	+ ($2e^{-36}$)	+ ($6e^{-32}$)	+	+	+	+	+	Holliday junction resolution w/ MSH4
*mer*3	+ ($2e^{-122}$)	+ ($6e^{-123}$)	+ ($2e^{-109}$)	+	+	+	+	+	Holliday junction resolution
DNA repair and recombination genes									
*mre*11	+ ($2e^{-99}$)	+ ($3e^{-90}$)	+ ($2e^{-32}$)	+	+	+	+	+	dsDNA exonuclease/ssDNA endonuclease
*rad*50	+ ($2e^{-73}$)	+ ($1e^{-74}$)	+ ($2e^{-47}$)	+	+	+	+	+	DNA binding, holds broken DNA ends
*rad*1	+ ($6e^{-118}$)	+ ($7e^{-118}$)	+ ($1e^{-23}$)	+	+	+	+	+	5'-3' endonuclease for nucleotide excision repair
*rad*52	-	+ ($7e^{-19}$)	+ ($4e^{-07}$)	-	+	-	-	+	DSB repair by homologous recombination
*msh*2	+ ($4e^{-157}$)	+ ($1e^{-161}$)	+ ($4e^{-65}$)	+	+	+	+	+	Binds base-base mismatches with MSH6
*msh*6	+ ($9e^{-161}$)	+ ($9e^{-161}$)	+ ($2e^{-59}$)	+	+	+	+	+	Binds base-base mismatches with MSH2
*mlh*1	+ ($4e^{-87}$)	+ ($6e^{-87}$)	+ ($1e^{-41}$)	+	+	+	+	+	Di- and tri-nucleotide mismatch repair
*mlh*2	+ ($7e^{-11}$)	+ ($7e^{-11}$)	+ ($2e^{-14}$)	+	+	-	+	+	Removal of cisplatin adducts
*mlh*3	+ ($2e^{-26}$)	+ ($9e^{-25}$)	+ ($2e^{-19}$)	-	+	+	+	+	Frameshift repair
*pms*1	+ ($3e^{-09}$)	+ ($5e^{-09}$)	+ ($4e^{-7}$)	+	+	-	-	+	DNA mismatch repair
smc1	+ ($2e^{-56}$)	+ ($1e^{-56}$)	+ ($2e^{-32}$)	+	+	+	+	+	Sister chromatid cohesion subunit w/ SMC3
smc2	+ ($3e^{-92}$)	+ ($1e^{-91}$)	+ ($5e^{-74}$)	+	+	+	+	+	Chromosome assembly and segregation
smc3	+ ($1e^{-59}$)	+ ($1e^{-59}$)	+ ($1e^{-14}$)	+	+	+	+	+	Sister chromatid cohesion subunit w/ SMC1
smc4	+ ($5e^{-69}$)	+ ($7e^{-69}$)	+ ($2e^{-66}$)	+	+	+	+	+	Chromosome assembly and segregation

smc5	+ ($1e^{-34}$)	+ ($1e^{-34}$)	+ ($4e^{-40}$)	+	+	+	+	+	DNA repair
smc6	+ ($7e^{-39}$)	+ ($8e^{-39}$)	+ ($6e^{-25}$)	+	+	+	+	+	Post replication DNA repair w/ SMC5
rad21	+ ($1e^{-17}$)	+ ($2e^{-17}$)	+ ($4e^{-20}$)	-	+	-	-	+	Sister chromatid binding
scc3	+ ($2e^{-6}$)	+ ($3e^{-11}$)	+ ($8e^{-30}$)	+	+	+	+	+	Sister chromatid binding
pds5	-	-	-	+	+	-	-	-	Sister chromatid binding in late prophase

Core meiotic gene and DNA repair and recombination gene survey in *Chrysochromulina tobinii, Chrysochromulina parva* and other haptophytes and eukaryotes. Genes were selected based on the meiotic gene survey of *Trichomonas vaginalis* [38]. The "+/-" represents presence or absence of an orthologous gene on the basis of TBLASTN output of less than $1e^{-6}$ E-value. Numbers under the "+" represent the E-value of the orthologous hit.

* The *Emiliania huxleyi* DMC1 gene (NCBI GI# 551629259) was used as the representative sequence for *dmc*1 query.

Supplementary Table S2. Meiotic gene identification

Gene	Full name (GenBank)	GI #	Organism
spo11	Meiotic recombination protein	30696943	*Arabidopsis thaliana*
hop1	Essential protein for meiotic synapsis	37999050	*Oryza sativa Japonica Group*
hop2	Homologous-pairing protein 2	15222250	*Arabidopsis thaliana*
mnd1	Meiotic nuclear division protein 1-like protein	30688234	*Arabidopsis thaliana*
rec8	REC8	289472321	*Daphnia pulex*
dmc1	DMC1 DNA recombinase	551629259	*Emiliania huxleyi*
rad51	DNA repair protein RAD51 homolog 1	585770	*Gallus gallus*
msh4	MutS protein homolog 4 MSH4	308153466	*Saccharomyces cerevisiae*
msh5	Unnamed protein product	9294568	*Arabidopsis thaliana*
mer3	Meiotic recombination protein	75320515	*Arabidopsis thaliania*
mre11	DNA repair and meiosis protein	9759499	*Arabidopsis thaliana*
rad50	DNA repair protein	57013013	*Arabidopsis thaliana*
rad1	Repair endonuclease	22655254	*Arabidopsis thaliana*
rad52	RAD52	33667840	*Giardia intestinalis*
msh2	DNA mismatch repair protein	3914056	*Arabidopsis thaliana*
msh6	DNA mismatch repair protein	15235223	*Arabidopsis thaliana*
mlh1	Putative MLH1	13430732	*Arabidopsis thaliana*
mlh2	Mlh2	33667828	*Giardia intestinalis*
mlh3	Putative protein MLH3	7270503	*Arabidopsis thaliana*
pms1	Pms1	33667830	*Giardia intestinalis*
smc1	Structural maintenance of chromosomes (SMC)-like protein	7258371	*Arabidopsis thaliana*
smc2	Structural maintenance of	15241831	*Arabidopsis*

	chromosomes protein		*thaliana*
smc3	SMC3 protein cohesion	23476966	*Arabidopsis thaliana*
smc4	SMC4 protein	21262152	*Oryza sativa*
smc5	SMC5 protein	27227807	*Oryza sativa*
smc6	SMC6 protein	27227809	*Oryza sativa*
rad21	Sister chromatid cohesion 1 protein 3	15231707	*Arabidopsis thaliana*
scc3	Putative sister-chromatid cohesion protein	20258987	*Arabidopsis thaliana*
pds5	ARM repeat superfamily protein	30699273	*Arabidopsis thaliana*

```
Chysochromulina_parva      MMFPVTAGQFDLVYNALSFTLASMMASTIFFWIRIGSVNEKYKSAMTITGLVTFIAAYHY
Chysochromulina_tobinii    MMFPVTAGQFDLVYNALSFTLASMMASTIFFWIRMGSVSEKYKSAMTITGLVTFIAAYHY
Sphingobium_sp._RAC03      -MDMITAGQYLLVYNAFSFTFASMAAATLFFWFGRSQVGPAYKTALTITGLVTAIAAYHY
                            *  :****: *****:***:*** *:*:***:  ..*.  **:*:******* ******

Chysochromulina_parva      IRIFNSWNESYHYPEAADGVIQDPVITGQPFNDAYRYMDWMLTVPLLMIEIIFVMGLSPE
Chysochromulina_tobinii    IRIFNSWNESYHYPEAADGVVQDPVITGQPFNDAYRYMDWMLTVPLLMIEIIFVMGLSPE
Sphingobium_sp._RAC03      YRIFESWSEAYALTD---GVI---TASGVAFNDAYRYVDWLLTVPLLLIELVLVMRLSQS
                            ***:**.*:*   :   **:   . :*  *******:**:******:**:::** ** .

Chysochromulina_parva      ETASKATSLGVAAGLMIVLGYPGELIIEGDLNVRWMWWTLAMIPFLYVVHTLLIGLQDKI
Chysochromulina_tobinii    ETAAKATSLGVAAGLMIVLGYPGELIIEGDLNVRWMWWTLAMIPFLYVVHTLLIGLQGAI
Sphingobium_sp._RAC03      ATVSTSVRLGSAAALMIILGYPGEIADN--NSTRALWGTLSSIPFLYIVWELFKGLGDAI
                            *.:.:. ** **.***:******:  :   ..* :* **: *****:*  *: ** . *

Chysochromulina_parva      KAEKNEEVKQKLNMVCWATVISWCTYPIVYVFPMFGLDGPSAVVAIQLGYCVSDIISKCG
Chysochromulina_tobinii    KEEKNEEVAKKLNMVCWATVVSWCTYPIVYVFPMLGLDGPSAVVAIQLGYCVSDIISKCG
Sphingobium_sp._RAC03      ERQPES-ARGLIKQARLLTFASWGFYPIVYMIPYTNLAGGAVETGVQIGYTIADLIAKAG
                           : : :. .   :: .   *. **  *****::*  .* * :. ..:*:** ::*:*:*.*

Chysochromulina_parva      VGFLIYNITIAKSN-EGYTQVH--------
Chysochromulina_tobinii    VGFLIYNITIAKSNPEGYAQVH--------
Sphingobium_sp._RAC03      VGILVYMIAVRKSAAEYGEPGKPATVTAVG
                           **:*:* *:: **  *     :
```

Supplementary Fig. S1: Similarity between the bacteria *Sphingobium* sp. RAC03, *Chrysochromulina tobinii* and *Chrysochromulina parva* xanthorhodopsin genes. *C. parva* numbering (38): Asp99 (proton acceptor), Leu107 (green spectral tuning), Glu110 (proton donor), Lys238 (retinal binding).

Supplementary Table S3: Mitochondrial gene compliment and SNPs between *Chrysochromulina parva* and *Chrysochromulina tobinii*

Gene (*C. parva*)	SNPs	Non-synonymous Mutations	Synonymous Mutations	%Non-synonymous Mutations	gene length (bp)	Non-synonymous changes per kB	dN/dS*
rps8	1	1	0	100%	347	2.88	UD
rps3	21	7	14	33%	785	8.92	0.134
nad5	65	11	54	17%	2030	5.42	0.074
nad2	74	10	64	14%	1466	6.82	0.045
cob	23	2	21	9%	1145	1.75	0.027
rsps12	2	0	2	0%	461	0.00	0
nad6	32	8	24	25%	644	12.42	0.084
nad1	35	3	32	9%	968	3.10	0.026
rps14	33	16	17	48%	305	52.46	0.241
cox2	54	7	47	13%	740	9.46	0.054
cox1	98	2	96	2%	1514	1.32	0.001
nad4	74	10	64	14%	1430	6.99	0.044
atp9	2	0	2	0%	227	0.00	0
cox3	12	0	12	0%	890	0.00	0
nad3	4	0	4	0%	359	0.00	0
atp8	3	0	3	0%	392	0.00	0
atp4	8	3	5	38%	563	5.33	0.158
atp6	6	2	4	33%	752	2.66	0.139
rpl16	3	3	0	100%	434	6.91	UD
nad4l	4	0	4	0%	302	0.00	0

Gene (*C. parva*)	SNPs
tRNA-Ser (1)	0
tRNA-Arg (1)	1
tRNA-Leu	4
tRNA-Asp	1
tRNA-Lys	0
tRNA-Cys	2
tRNA-Leu (2)	2
tRNA-Met (1)	0
tRNA-Met (2)	0
tRNA-His	2
tRNA-Thr	1
tRNA-Trp	1
tRNA-Asn	0
tRNA-Phe	0

Gene (*C. parva*)	SNPs
tRNA-Gly	0
tRNA-Pro	0
tRNA-Ser (2)	0
tRNA-Gln	0
tRNA-Tyr	0
tRNA-Val	1
tRNA-Met (3)	0
tRNA-Ile	0
tRNA-Glu	0
tRNA-Ala	0

Gene (*C. parva*)	SNPs	Inser-tions
23S rRNA	43	2
16S rRNA	10	2

tRNA-Arg (2)	0

***** Nei and Gojobori assessment method [101], “Undefined” (UD) resulted from a division by 0

Supplementary Table S4: Chloroplast gene compliment and SNP occurrence between *Chrysochromulina parva* and *Chrysochromulina tobinii*

Gene (*C. parva*)	SNPs	Non-synonymous Mutations	Synonymous Mutations	%Non-synonymous Mutations	gene length (bp)	Non-synonymous changes per kB	dN/dS*
ccs1	2	1	1	50%	1307	0.77	0.27
ycf55	1	0	1	0%	950	0.00	0
trg1	1	1	0	100%	719	1.39	UD
rpoB	3	2	1	67%	3335	0.60	0.57
rpoC1	1	0	1	0%	1847	0.00	0
rps2	1	1	0	100%	689	1.45	UD
atpI	1	0	1	0%	701	0.00	0
atpA	1	0	1	0%	1502	0.00	0
dnaK	1	0	1	0%	1835	0.00	0
rpl16	1	1	0	100%	404	2.48	UD
rps8	1	0	1	0%	398	0.00	0
ycf20	1	0	1	0%	272	0.00	0
ccsA	1	0	1	0%	953	0.00	0
psbA	4	4	0	100%	1082	3.70	UD
rps6	1	1	0	100%	326	3.07	UD
ycf80	1	0	1	0%	1478	0.00	0
ycf3	1	0	1	0%	515	0.00	0
psbC	1	0	1	0%	1415	0.00	0
rpl19	1	0	1	0%	386	0.00	0
ycf46	1	0	1	0%	1481	0.00	0
tatC	1	0	1	0%	773	0.00	0
groEL	5	1	4	20%	1604	0.62	0.08
secA	2	0	2	0%	2618	0.00	0
ycf12	1	0	1	0%	104	0.00	0
psbB	2	0	2	0%	1529	0.00	0
psbZ	1	0	1	0%	188	0.00	0
clpC	2	0	2	0%	2462	0.00	0
psaA	1	0	1	0%	2258	0.00	0
TOTAL	41	12	29				
Genes with no SNPs: *rpl21, rpl20, rpl27, rbcS, rbcL, psaF, psaJ, petL, ycf4, rpoC2, rps4, atpH, atpG, atpF, atpD, rpl3, rpl23, rpl2, rps19, rpl22, rps3, rps17, rpl14, rpl5, rpl6, rps5, secY, rpl36, rps13, rps11, rpoA, rps9, rpl31, rps12, rps7, tufA, rps10, ycf39, psaD, psbV, ORF154, psaM, psaI, ycf45, psbE, psbF, psbL, psbJ, rbcR, psbY, minD, thiG, chlI, ycf65, rps16, rpl34, rps14, psbD, petD, petB, 5S rRNA, 23S rRNA, 16S rRNA, psbI, ycf60, ycf24, ycf19, rpl33, rps18, psbX, petA, ycf35, petG, cfxQ, psaC, psbH, psbN, psbT, atpB, atpE, petN, petM, secG, psaL, psaB*							

* Nei and Gojobori assessment method [101], “Undefined” (UD) resulted from a division by 0

Supplementary data S1: FtsZ alignments in FASTA format. (fasta file download)

Supplementary Table S5: Primers used for sequence comparison and phylogeny of *Chrysochromulina* isolates

Gene	Primer	Sequence	Anneal. Temp. °C	Source
18S rRNA (PCR)	18ScomF1	GCTTGTCTCAAAGATTAAGCCATGC	68	Zhang et. al. 2005 [102]
	18ScomR1	CACCTACGGAAACCTTGTTACGAC		
18S rRNA (Internal Sequencing)	18SISE	CTGACACAGGGAGGTAGTGAC	N/A	Bendiff et. al. 2011 [103]
	18SIAS	TCCTCACTATGTCTGGACCTG		
28S rRNA	Chryso 28SF	AGTCTAGAAAGGCGCCATCG	66	This publication
	Chryso 28SR	GTCGGCGTTCAAGCTATCC		
psbA	psbAF	ATGACTGCTACTTTAGAAAGACG	59	Yoon et. al. 2002 [104]
	psbAR2	TCATGCATWACTTCCATACCTA		
rbcL	PrL1	CCTTATGCAAAAATGGGTTACTGG	63	Fujiwara et. al. 1994 [105]
	PrL4	CCGATTGTACCACCACCGAA		
cox1	LCO1490	GGTCAACAAATCATAAAGATATTGG	59	Folmer et. al. 1994 [106]
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA		
nad5	Chryso nad5F	TGTTGGTGACGTTGGTTTAG	65	This publication
	Chryso nad5R	GGTACAAAATGCAGCACAAC		

Supplementary Table S6: Species information for 18S rRNA sequences used to construct the phylogeny in Figure 7.

Species	Isolate/ Clone	Collection Site	Isolator, Year	Accession	Sequence Source
Chrysochromulina tobinii	CCMP291	Lake Colorado, USA	P. Kugrens 2009	KJ540196	This publication
Chrysochromulina parva	UW isolate 1161	Big Walnut Creek, Shadeville, Ohio	R. Andersen 2014	PJAB00000000	This publication
Chrysochromulina parva Lackey	NIES-562	Lake, Tsukuba, Japan	N. Hatakeyama, 1992	KJ540197	This publication
Chrysochromulina rotalis Eikrem & Throndsen	UIOP16	Skagerrak Strait, Norway	W. Eikrem, 1990	AM491025	Medlin et. al. 2008 [107]
Chrysochromulina acantha Leadbeater & Manton	ALGO HAP 78	Atlantic ocean, France	J. Fresnel, 1994	FN599059	Edvardsen et. al. 2011 [30]
Chrysochromulina throndsenii Eikrem	UIO048 (L12)	Skagerrak, Norway	W. Eikrem, 1989	AJ246277	Edvardsen et. al. 2011 [30]
Chrysochromulina cymbium Leadbeater & Manton	UIOR18	Skagerrak, Norway	W. Eikrem, 1992	AM491018	Medlin et. al. 2008 [107]
Chrysochromulina strobilus Parke & Manton	PCC 43	N. Atlantic	Unavailable	FN599060	Edvardsen et. al. 2011 [30]
Chrysochromulina campanulifera Manton & Leadbeater	UI J10	Skagerrak, Norway	J. Throndsen, 1984	AJ246273	Edvardsen et. al. 2011 [30]
Chrysochromulina scutellum Eikrem & Moestrup	UIO046 (G7)	Skagerrak, Norway	W. Eikrem, 1990	AJ246274	Edvardsen et. al. 2011 [30]
Chrysochromulina simplex Estep, Davis, Hargraves & Sieburth	UIO047 (JomfB)	Oslofjord, Norway	W. Eikrem, 1989	AM491021	Medlin et. al. 2008 [107]

Chrysochromulina leadbeateri Estep, Davis, Hargraves & Sieburth	UIO035 (ERIK)	Lofoten, N. Norway	W. Eikrem, 1991	AM491017	Medlin et. al. 2008 [107]
Cruciplacolithus neohelis (McIntyre & Be) P. Reinhardt	CCMP298	La Jolla, California	K. Lee, 1984	AJ246262	Edvardsen et al. unpub
Coccolithus pelagicus ssp. braarudii (Wallich) Schiller	PLY 182g	English Channel	J.C. Green 1990	AJ246261	Edvardsen et al. unpub
Emiliania huxleyi (Lohmann) Hay & Mohler	PCC 92D	Unavailable	Unavailable	M87327	Bhattacharya et al. unpub

9.1 References:

1) Liu H, Aris-Brosou S, Probert I, de Vargas C. **A time line of the environmental genetics of the haptophytes.** *Mol Biol Evol* 2010, **27**(1):161-176. doi: 10.1093/molbev/msp222

2) Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. **A molecular timeline for the origin of photosynthetic eukaryotes.** *Mol Biol Evol* 2004, **21**:809-818. doi: 10.1093/molbev/msh075

3) Fearman JA, Bolch CJS, Moltschaniwskyj NA. **Energy storage and reproduction in mussels, *Mytilus galloprovincialis*: the influence of diet quality.** *J Shellfish Res* 2009, **28**(2):305-312. doi: 10.2983/035.028.0212

4) Milke LM, Bricelj VM, Parrish CC. **Biochemical characterization and nutritional value of three *Pavlova* spp. in unialgal and mixed diets with *Chaetoceros muelleri* for postlarval sea scallops, *Placopecten magellanicus*.** *Aquaculture* 2008, **276**:130-142. doi: 10.1016/j.aquaculture.2008.01.040

5) Xu J, Zhou H, Yan X, Zhou C, Zhu P, Ma B. **Effect of unialgal diets on the composition of fatty acids and sterols in juvenile ark shell *Tegillarca granosa* Linnaeus.** *J Agric Food Chem* 2012, **60**(15):3973-3980. doi: 10.1021/jf300620e

6) Brown MR, Skabo S, Wilkinson B. **The enrichment and retention of ascorbic acid in rotifers fed microalgal diets**. *Aquaculture Nutrition* 1998, **4**(3):151-156. doi: 10.1046/j.1365-2095.1998.00060.x

7) Kraul S. **Live food for marine fish larvae.** In *Avances en Nutrición Acuicola VIII.* Edited by Suárez LEC, Marie DR, Salazar MT, López MGN, Cavazos DAV, Cruz ACP, Ortega AG. Nuevo Leon: Universidad Autonoma de Nuevo Leon; 2006:55-61.

8) Bigelow N, Barker J, Ryken S, Patterson J, Hardin W, Barlow S, Deodato C, Cattolico RA. ***Chrysochromulina* sp.: A proposed lipid standard for the algal biofuel industry and its application to diverse taxa for screening lipid content.** *Algal Res* 2013, 2:385–393.

9) Andruleit H, Young JR. ***Kataspinifera baumannii*: a new genus and species of deep photic coccolithophores resembling the non-calcifying haptophyte *Chrysochromulina*.** *Journal of Micropalaeontology* 2010, **29**:135-147. doi: 10.1144/0262-821X10-006

10) Yoshida M, Noël MH, Nakayama T, Naganuma T, Inuoye I. **A haptophyte bearing siliceous scales: ultrastructure and phylogenetic position of *Hyalolithus neolepsis* gen. et sp. nov. (Prymnesiophyceae, Haptophyta).** *Protist* 2006, **157**(2):213-234. doi: 10.1016/j.protis.2006.02.004

11) Shutler JD, Grant MG, Miller PI, Rushton E, Anderson K. **Coccolithophore bloom detection in the north east Atlantic using SeaWiFS: algorithm description, application, and sensitivity analysis.** *Remote Sens Environ* 2010, **114**(5):1008-1016. doi: 10.1016/j.rse.2009.12.024

12) Phaeocystis Research Website: http://www.phaeocystis.org/about.html

13) Janse I, Zwart G, van der Maarel MJECVD, Gottschal JC. **Composition of the bacterial community degrading *Phaeocystis* mucopolysaccharides in enrichment cultures.** *Aquatic Microb. Ecol.* 22: 119-133, 2000. doi: 10.4319/lo.1999.44.6.1447

14) A.R.J. Curson, B.T. Williams, B.J. Pinchbeck, L.P. Sims, A.B. Martínez, P.P.L. Rivera, D. Kumaresan, E. Mercadé, L.G. Spurgin, O. Carrión, S. Moxon, R.A. Cattolico, U. Kuzhiumparambil, P. Guagliardo, P.L. Clode, J.-B. Raina, J.D. Todd, **DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton**. *Nat. Microbiol.* 3 (2018) 430. doi:10.1038/s41564-018-0119-5.

15) Jordan RW, Chamberlain AHL**. Biodiversity among haptophyte algae.** *Biodivers Conserv* 1997, **6**(1):131-152. doi: 10.1023/A:1018383817777

16) Eschbach E, John U, Reckermann M, Cembella AD, Edvardsen B, Medlin LK. **Cell cycle dependent expression of toxicity by the ichthyotoxic prymnesiophyte *Chrysochromulina polylepis*.** *Aquat Microb Ecol* 2005, **39**:85-95. doi: 10.3354/ame039085

17) Nicholls KH, Beaver JL, Estabrook RH. **Lakewide odours in Ontario and New Hampshire caused by *Chrysochromulina breviturrita* Nich. (Prymnesiophyceae).** *Hydrobiologia* 1982, **96** (1):91-95. doi: 10.1007/BF00006281

18) Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, Mayer C, Miller J, Monier, A., Salamov A, Young J, Aguilar M, Claverie J-M, Frickenhaus S, Gonzalez K, Herman E, Lin Y-C, Napier J, Ogata H, Sarno A, Shmutz J, Schroeder D, de Vargas C, Verret F, Dassow P, Valentin K, Van de Peer Y, Wheeler G, Dacks J, Delwiche C, Dyhrman S, Glöckner G, John U, Richards T, Worden A, Zhang X, Grigoriev I. **Pan genome of the phytoplankton *Emiliania* underpins its global distribution.** *Nature* 499: 209–213, 2013. doi: 10.1038/nature12221

19) Dahl E, Bagoien E, Edvardsen B, Stenseth N. **The dynamics of *Chrysochromulina* species in the Skagerrak in relation to environmental conditions.** 2005, *J. of Sea Res.* 54: 15-24. doi: 10.1016/j.seares.2005.02.004

20) Seoane S, Eikrem W, Pienaar R, and B. Edvardsen. ***Chrysochromulina palpebralis* sp. nov. (Prymnesiophyceae): a haptophyte, possessing two alternative morphologies.** *Phycologia* 2009; **48**(3):165-176. doi: 10.2216/08-63.1.

21) Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, Nota F, de Vargas C. **Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans.** Proc Natl Acad Sci U S A. 2009;106: 12803–12808. doi:10.1073/pnas.0905841106

22) Hansen, Per; Nielsen, Torkel; Kaas, Hanne (1995). "Distribution and Growth of Protists and Mesozooplankton during a Bloom of Chrysochromulina Spp. (Prymnesiophyceae, Prymnesiales)". Phycologia. 34 (5): 409–416. doi:10.2216/i0031-8884-34-5-409.1.

23) Hovde BT, Starkenburg SR, Hunsperger HM, Mercer LD, Deodato CR, Jha RK, Chertkov O, Monnat RJ, Cattolico RA. **The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles.** *BMC Genomics*. 2014;15: 604. doi:10.1186/1471-2164-15-604

24) Hovde BT, Starkenburg SR, Deodato C, Chertkov O, Monnat R, Cattolico RA. **Genome sequence and transcriptome analyses of *Chrysochromulina*: metabolic tools for enhanced**

algal fitness in the prominent order Prymnesiales (Haptophyceae). (2015) PLoS Genetics. doi: 10.1371/journal.pgen.1005469

25) Chretiennot-Dinet MJ, Desreumaux N, and R Vignes-Lebbe. **An interactive key to the species (Haptophyta) described in the literature.** PhytoKeys 34: 47–60 (2014) doi: 10.3897/phytokeys.34.6242

26) Lackey J. **Notes on plankton flagellates from the Scioto River (with descriptions of new forms).** *Lloydia* 1939, **2**:128-143.

27) Parke M, Green J, Manton I. **Studies in marine flagellates II. Three new species of *Chrysochromulina*.** *Journal of Marine Biology Association of the UK*. 34: 579-609, 1955

28) Parke M, Lund JWG, Manton I. **Observations on the biology and fine structure of the type species of *Chrysochromulina* (*C. parva* Lackey) in the English Lake District.** *Arch Mikrobiol* 1962, **42**(4):333-352.

29) Deodato C, Barlow S, Andersen RA, Hovde BT, Cattolico RA. ***Chrysochromulina* (Haptophyceae) isolates from lake and river ecosystems: An electron microscopic comparison and clarification of type species diagnosis.** (Unpublished – in preparation)

30) Edvardsen B, Eikrem W, Throndsen J, Sáez AG, Probert I, Medlin LK. **Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta).** *Eur J Phycol*. 2011;46: 202–228. doi: 10.1080/09670262.2011.594095

31) Wang D, Ning K, Li J, Hu J, Han D, Wang H, Zeng X, Jing X, Zhou Q, Su X, Chang X, Wang A, Wang W, Jia J, Wei L, Xin Y, Qiao, Y, Huang, R, Jie Chen, J, Bo Han, B, Kangsup Yoon, K, Hill RT, Zohar RY, Feng Chen Y, Qiang Hu Q, Xu J. ***Nannochloropsis* genomes reveal evolution of microalgal oleaginous traits.** *PLOS Genetics*, 2014. doi: 10.1371/journal.pgen.1004094

32) Fixen K, Starkenburg SR, Hovde BT, Deodato C, Harwood C, Cattolico RA. **Genome sequences of 8 bacterial species found in co-culture with the haptophyte *Chrysochromulina tobin*.** *Genome Announcement* 4, 2016 (6):e01162-16. doi:10.1128/genomeA.01162-16.

33) Houdan A, Billard C, Marie D, Not F, Sáez AG, Young JR, Probert I: **Holococcolithophore-heterococcolithophore (Haptophyta) life cycles: flow cytometric analysis of relative ploidy levels.** *Systematics and Biodiversity* 2003, **1**(4):453-465. doi: 10.1017/S1477200003001270

34) Vaulot D, Birrien J-L, Marie D, Casotti R, Veldhuis MJW, Kraay GW, Chretiennot-Dinet M-J. **Morphology, ploidy, pigment composition, and genome size of cultured strains of *Phaeocystis* (prymnesiophyceae).** *J Phycol*. 1994;30: 1022–1035. doi:10.1111/j.0022-3646.1994.01022.x

35) Fujiwara S, Hirokawa Y, Takatsuka Y, Suda K, Asamizu E, Takayanagi T, Shibata D, Tabata S, Tsuzuki M. **Gene expression profiling of coccolith-bearing cells and naked cells in haptophyte *Pleurochrysis haptonemofera* with a cDNA macroarray system.** *Mar Biotechnol (NY)* **9**(5):550-560, 2007.

36) Pan KH, Qin JJ, Li S, Dai WK, Zhu BH, Jin YC, Yu WG, Yang GP. **Nuclear monoploidy and asexual propagation of *Nannochloropsis oceanica* as revealed by its genome sequence.** *Journal of Phycology*, 47: 1425–1432, 2011. doi: 10.1111/j.1529-8817.2011.01057.x

37) Schurko AM, Logsdon JM. **Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex.** *Bioessays* **30**(6):579-589, 2008. doi: 10.1002/bies.20764

38) Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon Jr JM. **An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*.** *PLoS One* 2008, **3**(8):e2879. doi 10.1371/journal.pone.0002879

39) Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Jiang W, Marraffini LA, Zhang F: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339**(6121):819-23. doi: 10.1126/science.1231143

40) John U, Beszteri S, Glöckner G, Singh R, Medlin L, Cembella AD. **Genomic characterisation of the ichthyotoxic prymnesiophyte *Chrysochromulina polylepis*, and the expression of polyketide synthase genes in synchronized cultures.** *Eur J Phycol.* 2010;45: 215–229. doi:10.1080/09670261003746193

41) Edvardsen B, Vaulot D: **Ploidy analysis of the two motile forms of *Chrysochromulina polylepis* (Prymnesiophyceae).** *J Phycol* 1996, **32**(1):94-102. doi: 10.1111/j.0022-3646.1996.00094.x

42) EMBL-EBI Pfam Website: Family: DUF4143 http://pfam.xfam.org/family/PF13635

43) Nishimura Y, Kamikawa R, Hashimoto T, Inagaki Y. **An intronic open reading frame was released from one of group II introns in the mitochondrial genome of the haptophyte *Chrysochromulina* sp. NIES-1333.** *Mobile Genetic Elements* 4, e29384 2014. doi: 10.4161/mge.29384

44) Hunsperger HM, Randhawa T, Cattolico RA. **Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae.** *BMC Evol Biol.* 2015;15: 16. doi: 10.1186/s12862-015-0286-4

45) Adams KL, Palmer JD. **Evolution of mitochondrial gene content: gene loss and transfer to the nucleus.** *Mol Phylogenet Evol* 2003, 29:380–395.

46) Zickermann V, Wirth C, Nasiri H, Siegmund K, Schwalbe H, Hunte C, Brandt U. **Mechanistic insight from the crystal structure of mitochondrial complex I.** *Science* 347: 44-49, 2015. doi: 10.1126/science.1259859

47) Guo, R., Zong, S.,Wu, M., Gu, J., and M. Yang. **Architecture of Human Mitochondrial Respiratory Megacomplex I**. 2017 *Cell* 170, 1247–1257 doi: 10.1016/j.cell.2017.07.050

48) Mimakia M,Wanga X, McKenzieb M, Thorburncd DR, Ryana MT. **Understanding mitochondrial complex I assembly in health and disease**. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*:18:851-862, 2012. doi: 10.1016/j.bbabio.2011.08.010

49) Lazarou M, Thornburn D, Ryan M, McKenzie M. **Assembly of mitochondrial complex I and defects in disease.** *Biochim. Biophys. Acta*. 1793: 78-88, 2009. doi: 10.1016/j.bbamcr.2008.04.015

50) Oudot M, Kloareg B, Loiseaux-de Goer S. **The mitochondrial *Pylaiella littoralis nad*11 gene contains only the N-terminal FeS-binding domain.** *Gene* 235: 131-137, 1999.

51) Starkenburg SR, Kwon J, Chertkov O, Jha R, Twary S, Rocap G, Cattolico RA. **A pangenomic analysis of the *Nannochloropsis* organellar genomes reveals novel genetic variation in key metabolic genes.** *BMC Genomics* **15**:212, 2014. doi: 10.1186/1471-2164-15-212

52) Sevcikova T, Klimes V, Zbrankova V, Strnad H, Hroudova M, Vlcek C, Elias M. **A comparative analysis of mitochondrial genomes in Eustigmatophyte algae.** *Genome Biol. Evol* 2016;8:705-722, 2016. doi:101093/gbe/evw027.

53) Imanian B, Pombert J-F, Dorrell R, Burki F, Keeling P. **Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts.** *PLOS 1*: 7(8), 2012. doi: 10.1371/journal.pone.0043763

54) Patron NJ, Inagaki Y, Keeling PJ. **Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages.** *Curr Biol* 2007, **17**(10):887-891. doi: 10.1016/j.cub.2007.03.069

55) Burki F, Shalchian-Tabrizi K, Pawlowski J. **Phylogenomics reveals a new "megagroup" including most photosynthetic eukaryotes.** *Biol Lett* 2008, 4:366–369. doi: 10.1098/rsbl.2008.0224

56) De Clerck O, Bogaert K, Leliaert F. **Diversity and evolution of algae: primary endosymbiosis.** *Advances in Botanical Research* 64:55-86, 2012. doi: 10.1016/B978-0-12-391499-6.00002-5

57) Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D. **Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates.** *Mol Biol Evol* 2007, **24**(8):1702-1713. doi: 10.1093/molbev/msm089

58) Okamoto K, Chantangsi C, Horak A, Leander BS, Keeling PJ. **Molecular Phylogeny and Description of the Novel Katablepharid Roombia truncata gen. et sp. nov., and Establishment of the Hacrobia Taxon nov.** *PLoS ONE* 4(9):e7080. doi: 10.1371/journal.pone.0007080.

59) Burki F, Okamoto N, Pombert J-F, Keeling PJ. **The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins.** *Proc Biol Sci* 2012, 279:2246–2254. doi: 10.1098/rspb.2011.2301

60) Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, Barbeyron T, Michel G, Noel B, Valentin K, Elias M, Artiguenave F, Arun A, Aury JM, Barbosa-Neto JF, Bothwell JH, Bouget FY, Brillet L, Cabello-Hurtado F, Capella-Gutiérrez S, Charrier B, Cladière L, Cock JM, Coelho SM, Colleoni C, Czjzek M, Da Silva C, Delage L, Denoeud F, Deschamps P, Dittami SM, Gabaldón T, Gachon CM, Groisillier A, Hervé C, Jabbari K, Katinka M, Kloareg B, Kowalczyk

N, Labadie K, Leblanc C, Lopez PJ, McLachlan DH, Meslet-Cladiere L, Moustafa A, Nehr Z, Nyvall Collén P, Panaud O, Partensky F, Poulain J, Rensing SA, Rousvoal S, Samson G, Symeonidi A, Weissenbach J, Zambounis A, Wincker P, Boyen C. **Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida.** *Proc Natl Acad Sci*, 2013 110(13):5247-5252.

61) Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. **The evolution of photosynthesis in chromist algae through serial endosymbioses**. *Nat Commun*. 2014 ;5. doi: 10.1038/ncomms6764

62) Karol K, Jacobs MA, Zhou Y, Sims E, Gillett W, Cattolico RA. **Comparative analysis of complete mitochondrial genome sequences from two geographically distinct *Heterosigma akashiwo* (Raphidophyceae) strains.** *Proc Seventh Int Chrysophyte Symp* :261–282, 2010.

63) Smith DR, Keeling PJ. **Twenty-fold difference in evolutionary rates between the mitochondrial and plastid genomes of species with secondary red plastids.** *J Eukaryot Microbiol*, 59:181–184, 2012. doi: 10.1111/j.1550-7408.2011.00601.x

64) Smith D.R., Arrigo, K.R., Alderkamp, A. C. and A.E. Allen. **Massive difference in synonymous substitution rates among mitochondrial, plastid, and nuclear genes of *Phaeocystis* algae.** *Mol Phylogenet Evol* 71:36–40, 2014. doi: 10.1016/j.ympev.2013.10.018

65) Cattolico RA. **Variation in plastid number: Effect of chloroplast and nuclear deoxyribonucleic acid complement in the unicellular alga *Olithodiscus luteus***. *Plant Physiol*. 62:558-562, 1978.

66) Hennis AS, Birky CW. **Stochastic partitioning if chloroplasts at cell division in the alga *Olisthodiscus*, and compensating control of chloroplast replication**. *J. Cell Sci.* 70: 1-15, 1984.

67) Yoshida Y, Miyagishima SY, Kuroiwa H, Kuroiwa T. **The plastid-dividing machinery: Formation, constriction and fission.** *Curr Opin Plant Biol* 15(6):714–721, 2012. doi: 10.1016/j.pbi.2012.07.002

68) Miyagishima SY, Suzuki K, Okazaki K, Kabeya Y. **Expression of the nucleus- encoded chloroplast division genes and proteins regulated by the algal cell cycle**. *Mol Biol Evol* 29(10):2957–2970, 2012. doi: 10.1093/molbev/mss102

69) Itoh R, Takahashi H, Toda K, Kuroiwa H, Kuroiwa T. **Aphidicolin uncouples the chloroplast division cycle from the mitotic cycle in the unicellular red alga *Cyanidio-schyzon merolae*.** *Eur J Cell Biol* 71(3):303–310, 1996.

70) Deane JA, Strachan IM, Saunders GW, Hill DRA, McFadden G. **Cryptomonad evolution: Nuclear 18S rDNA phylogeny versus cell morphology and pigmentation**. *J. Phycol.* 38: 1236-1244, 2002. doi: 10.1046/j.1529-8817.2002.01250.x

71) Hashimoto, H. **Electron-opaque annular structure girdling the constricting isthmus of the dividing chloroplasts of *Heterosigma akashiwo* (Raphidophyceae, Chromophyta).** *Protoplasma* 197, Issue 3–4, 1997 doi: 10.1007/BF01288030

72) Magnussen C, Gibbs SP. **Behavior of chloroplast ER during chloroplast division in *Olisthodiscus luteus* (chrysophyceae)**. *Journal of Phycology*, Vol 16, Issue 2, 1980. doi:

10.1111/j.1529-8817.1980.tb03035.x

73) Wang W, Li J, Sun Q, Yu X, Zhang W, Jia N, An C, Li Y, Dong Y, Han F, Chang N, Liu X, Zhu Z, Yu Y, Fan S, Yang M, Luo S-Z, Gao H, Feng Y. **Structural insights into the coordination of plastid division by the ARC6–PDV2 complex.** *Nature Plants* 3, 2017. doi:10.1038/nplants.2017.11

74) Terbush A, Yoshida Y, Osteryoung K. **FtsZ in chloroplast division: structure, function and evolution.** *Curr. Opin. in Cell Biol.* 25:461-470, 2013. doi: 10.1016/j.ceb.2013.04.006

75) Sumiya N, Fujiwara T, Era A, Miyagishima S-Y. **Chloroplast division checkpoint in eukaryotic algae.** *PNAS*, 2016. doi: 10.1073/pnas.1612872113

76) Miyagishima S-Y, Kabeya Y. **Chloroplast division: squeezing the photosynthetic captive.** *Curr Opin Microbiol.,* 2010. doi: 10.1016/j.mib.2010.10.004

77) Miyagishima S-Y, Kabeya Y, Sugita C, Sugita M, Fujiwara T. **DipM is required for peptideglyclan hydrolysis during chloroplast division.** *Plant Biol.* 2014. doi: 10.1186/1471-2229-14-57

78) Nakanishi H, Suzuki K, Kabeya Y, Miyagishima SY. **Plant-specific protein MCD1 determines the site of chloroplast division in concert with bacteria-derived MinD.** *Curr Biol.* 2009 Jan 27;19(2):151-6. doi: 10.1016/j.cub.2008.12.018

79) Hirakawa Y, Ishida K. **Prospective function of FtsZ proteins in the secondary plastid of chlorarachniophyte algae.** *BMC Plant Biol.* 2015 Nov 10;15:276. doi: 10.1186/s12870-015-0662-7.

80) Jones DT, Taylor WR, Thornton JM. **The rapid generation of mutation data matrices from protein sequences.** *Computer Applications in the Biosciences* 8: 275-282, 1992.

81) Kumar S, Stecher G, Tamura K. **MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.** *Molecular Biology and Evolution* 33:1870-1874, 2016. doi: 10.1093/molbev/msw054

82) Miyagishima SY, Nozaki H, Nishida K, Nishida K, Matsuzaki M, Kuroiwa T. **Two types of FtsZ proteins in mitochondria and red-lineage chloroplasts: the duplication of FtsZ is implicated in endosymbiosis.** *J Mol Evol.* 2004 Mar;58(3):291-303. doi: 10.1007/s00239-003-2551-1

83) Hovde BT, Starkenburg SR, Deodato C, Fixen K, Cattolico RA. **Physiological cues affect the transcriptional activity of the haptophyte *Chrysochromulina tobin* and its associated bacterial biome: photoperiod responses.** (Unpublished - In preparation)

84) Boczar BA, Delaney TP, Cattolico RA. **Gene for the ribulose-1,5-bisphosphate carboxylase small subunit protein of the marine chromophyte *Olisthodiscus luteus* is similar to that of a chemoautotrophic bacterium.** *Proc Natl Acad Sci USA* 1989, **86**(13):4996-4999. doi: 10.1073/pnas.86.13.4996

85) Cattolico RA, Jacobs MA, Zhou Y, Chang J, Duplessis M, Lybrand T, McKay J, Ong HC, Sims E, Rocap G. **Chloroplast genome sequencing analysis of *Heterosigma akashiwo***

CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. *BMC Genomics* 2008, 9:211. doi: 10.1186/1471-2164-9-211

86) Boo SM, Kim HS, Shin W, Boo GH, Cho SM, Jo BY, Kim J-H, Kim JH, Yang EC, Siver PA, Wolfe AP, Bhattacharya D, Andersen RA, Yoon HS. **Complex phylogeographic patterns in the freshwater alga *Synura* provide new insights into ubiquity vs. endemism in microbial eukaryotes.** *Mol Ecol* 2010, **19**(19):4328-4338.

87) Saitou N, Nei M. **The neighbor-joining method: A new method for reconstructing phylogenetic trees.** *Molecular Biology and Evolution* 4:406-425, 1987. doi: 10.1093/oxfordjournals.molbev.a040454

88) Felsenstein J. **Confidence limits on phylogenies: An approach using the bootstrap.** *Evolution* 39:783-791, 1985. doi: 10.1111/j.1558-5646.1985.tb00420.x

89) Tamura K, Nei M, Kumar S. **Prospects for inferring very large phylogenies by using the neighbor-joining method.** *PNAS* 101:11030-11035, 2004. doi: 10.1073/pnas.0404206101

90) Nicholls KH. ***Chrysochromulina breviturrita* sp. nov., a new freshwater member of the Prymnesiophyceae** *Journal of Phycology* Dec 1978 doi: 10.1111/j.1529-8817.1978.tb02476.x

91) Kling HJ. ***Chrysochromulina laurentiana*: an electron microscopic study of a new species of Prymnesiophyceae from Canadian Shield lakes.** *Nordic Journal of Botany* Vol 1 issue 4 551-555, 1981 doi:10.1111/j.1756-1051.1981.tb00722.x

92) Wujek DE, Gardiner WE. **Chrysophyceae (Mallomonadaceae) from Florida. II. New species of *Paraphysomonas* and the prymnesiophyte *Chrysochromulina***. *Florida Scientist* 48: 56-63, 1985

93) Thompson RH, Halicki PJ. ***Chrysochromulina parva* Lackey in eastern Kansas.** *Trans Am Microsc Soc 1965,* **84**(1):14-17.

94) Nicholls KH. **Haptophyte Algae.** Freshwater Algae of North America (Second Edition) Ecology and Classification A volume in Aquatic Ecology. 2015. ISBN: 978-0-12-385876-4

95) Bertani, G. **Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*.** *J. Bacteriol.* 62: 293–300, 1951.

96) Reynolds, ES. **The use of lead citrate at high pH as an electron-opaque stain in electron microscopy.** *J Cell Biol.* 17(1): 208–212, 1963.

97). Li P-E, Lo CC, Anderson JJ, Davenport KW, Bishop-Lilly KA, Xu Y, et al. **Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform**. Nucleic Acids Res. 2017 Jan 9;45(1):67–80. doi: 10.1093/nar/gkw1027

98). Peng Y, Leung HCM, Yiu SM, Chin FYL. **IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth**. Bioinformatics. 2012 Jun 1;28(11):1420–8. doi: 10.1093/bioinformatics/bts174

99) Holt C, Yandell M. **MAKER2: an annotation pipeline and genome-database**

management tool for second-generation genome projects. *BMC Bioinformatics*, 2011 doi: 10.1186/1471-2105-12-491.

100) Rozen S, Skaletsky HJ. **Primer3 on the WWW for general users and for biologist programmers.** In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Edited by Krawetz S, Misener S. Totowa, NJ: Humana Press; 2000:365-386.

9.2 Supplementary References:

101) Nei M and Gojobori T. **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.** Mol Biol Evol. 1986 Sep;3(5):418-26. doi:10.1093/oxfordjournals.molbev.a040410

102) Zhang H, Bhattacharya D, Lin S. **Phylogeny of dinoflagellates based on mitochondrial cytochrome b and nuclear small subunit rDNA sequence comparisons.** *J. Phycol.* 2004, **41**: 411-420. doi: 10.1111/j.1529-8817.2005.04168.x

103) Bendif EM, Probert I, Hervé A, Billard C, Goux D, Lelong C, Cadoret J-P, Véron B. **Integrative taxonomy of the Pavlovophyceae (Haptophyta): a reassessment.** *Protist* 2011, **162**:738-761. doi: 10.1016/j.protis.2011.05.001

104) Yoon HS, Hackett JD, Battacharya D. **A single origin of the peridinin- and fucoxanthin-containing plstids in dinoflagellates through tertiary endosymbiosis.** *PNAS* 2002, **99**(18):11724-11729. doi: 10.1073/pnas.172234799

105) Fujiwara S, Sawada M, Someya J, Minaka N, Kawachi M, Inouye. **Molecular phylogenetic analysis of rbcL in the Prymnesiophyta.** *J. Phycol.* 1994, **30**:863-871. doi: 10.1111/j.0022-3646.1994.00863.x

106) Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. **DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.** *Mol. Mar. Biol. Biotechnol.* 1994, **3(**5):294-299. doi: doi.org/10.1016/j.marmicro.2007.08.007

107) Medlin LK, Sáez AG, Young JR. **A molecular clock for coccolithophores and implications for selectivity of phytoplankton extinctions across the K/T boundary.** *Marine Micropaleontology* Volume 67, Issues 1–2, April 2008. doi: 10.1016/j.marmicro.2007.08.007