Gene 804 (2021) 145871

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Chloroplast genome analysis of *Chrysotila dentata*

Yagya Prasad Paudel [a,b,1], Zixuan Hu [b,1], Janak Raj Khatiwada [b,1], Lu Fan [c], Shreeti Pradhan [b], Benwen Liu [d,*], Wensheng Qin [b,*]

[a] Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St, Toronto, ON M5S 3E1, Canada
[b] Department of Biology, Lakehead University, 955 Oliver Road, Thunder Bay, ON P7B 5E1, Canada
[c] School of Civil Engineering, Architecture and Environment, Hubei University of Technology, Wuhan 430068, China
[d] Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

ARTICLE INFO

Keywords:
Chrysotila dentata
Haplophyta
Chloroplast
Genome sequencing

ABSTRACT

Chrysotila dentata is an ecologically important marine alga contributing to the coccolith formation. In this study, a complete chloroplast (cp DNA) genome of *Chrysotila dentata* was sequenced by using Illumina Hiseq and was analyzed with the help of a bioinformatics tool CPGAVAS2. The circular chloroplast genome of *Chrysotila dentata* has a size of 109,017 bp with two inverted repeats (IRs) regions (4513 bp each) which is a common feature in most land plants and algal species. The *Chrysotila dentata* cp genome consists of 61 identified protein-coding genes, 30 tRNA genes and 6 rRNAs with 21 microsatellites. The phylogenetic relationship with other select algal species revealed a close phylogeny of *Chrysotila dentata* with Phaeocystis antarctica. This is the first report of the cp genome analysis of genus *Chrysotila* and the results from this study will be helpful for understanding the genetic structure and function of chloroplast in other species of *Chrysotila*.

1. Introduction

Chloroplasts are double membrane bounded organelles specialized in carrying out the photosynthesis in green plants and algae. Chloroplasts also play a significant role in the biosynthesis of amino acids and lipids (Daniell et al., 2016). The chloroplast (cp) genome is circular in most of the plant species and the size of the genome varies from 72 kb to 217 kb (Fang et al., 2020, Méndez-Leyva et al., 2019, Moore et al., 2007). In most cases, the chloroplast genome is a quadripartite structure consisting of pair of inverted repeats (IRs) regions separating a large single copy (LSC) from small single copy (SSC) regions (Chaney et al., 2016, Smith et al., 2014, Odintsova and Yurina 2006).

Haplophytes are photosynthetic microalgae found in marine and freshwater habitats. The genus *Pleurochrysis* consists of 9 species and this genus belongs to the haptophytes algae. There are several taxonomic ambiguities regarding *Pleurochrysis* species. Johansen et al. (1988) described a new variety of *Pleurochrysis carterae* (var. *dentata*) from an inland saline pond in New Mexico, USA based on coccolith morphology and thermal tolerance. Later, Sáez et al. (2003) carried out molecular phylogenetic analysis of coccolithophores species and assigned both *Pleurochrysis carterae* and *Pleurochrysis dentata* as an independent species. Afterwards, Andersen et al. (2014) revised the nomenclature of the genus *Pleurochrysis* and named as *Chrysotila*. Recent publication also follows the *Chrysotila dentata* as the correct nomenclature (Liu et al., 2019). *Chrysotila dentata* is a photoautotrophic unicellular marine alga belongs to the division Haplophyta and class Prymnesiophyceae. This group of the algae can produce, and deposit $CaCO_3$ scales called coccoliths (Chen et al., 2019). This genus *Chrysotila* represents the marine coccolithophore producing algae and has a great economic significance because of their blooms which is released during the formation of cocolith impacting on the carbon cycle (Reid et al., 2011). Coccolithophorids are well known for their roles in the precipitation of biogenic carbonate and their contribution to marine primary production. Despite the important ecological roles of the haplophytes only little is known regarding their genomic organization. The algae database (https://www.algaebase.org/) shows that there are 764 species of phylum haplophyta with the known chloroplast genome structure of only eight species which are *Isochrysis galbana* (Fang et al., 2020), *Tisochrysis lutea* (Méndez-Leyva et al., 2019), *Phaeocystis antarctica* and *Phaeocystis globose* (Smith et al., 2014), *Chrysochromulina tobin* (Hovde et al., 2014), *Chrysochromulina tobin* (Hovde et al., 2014), *Chrysochromulina parva*, *Pavlova lutheri* (Baurain et al., 2010). The study of

* Corresponding authors.
E-mail addresses: liubw@ihb.ac.cn (B. Liu), wqin@lakeheadu.ca (W. Qin).
[1] Co-first authors.

https://doi.org/10.1016/j.gene.2021.145871
Received 7 April 2021; Received in revised form 20 July 2021; Accepted 2 August 2021
Available online 4 August 2021

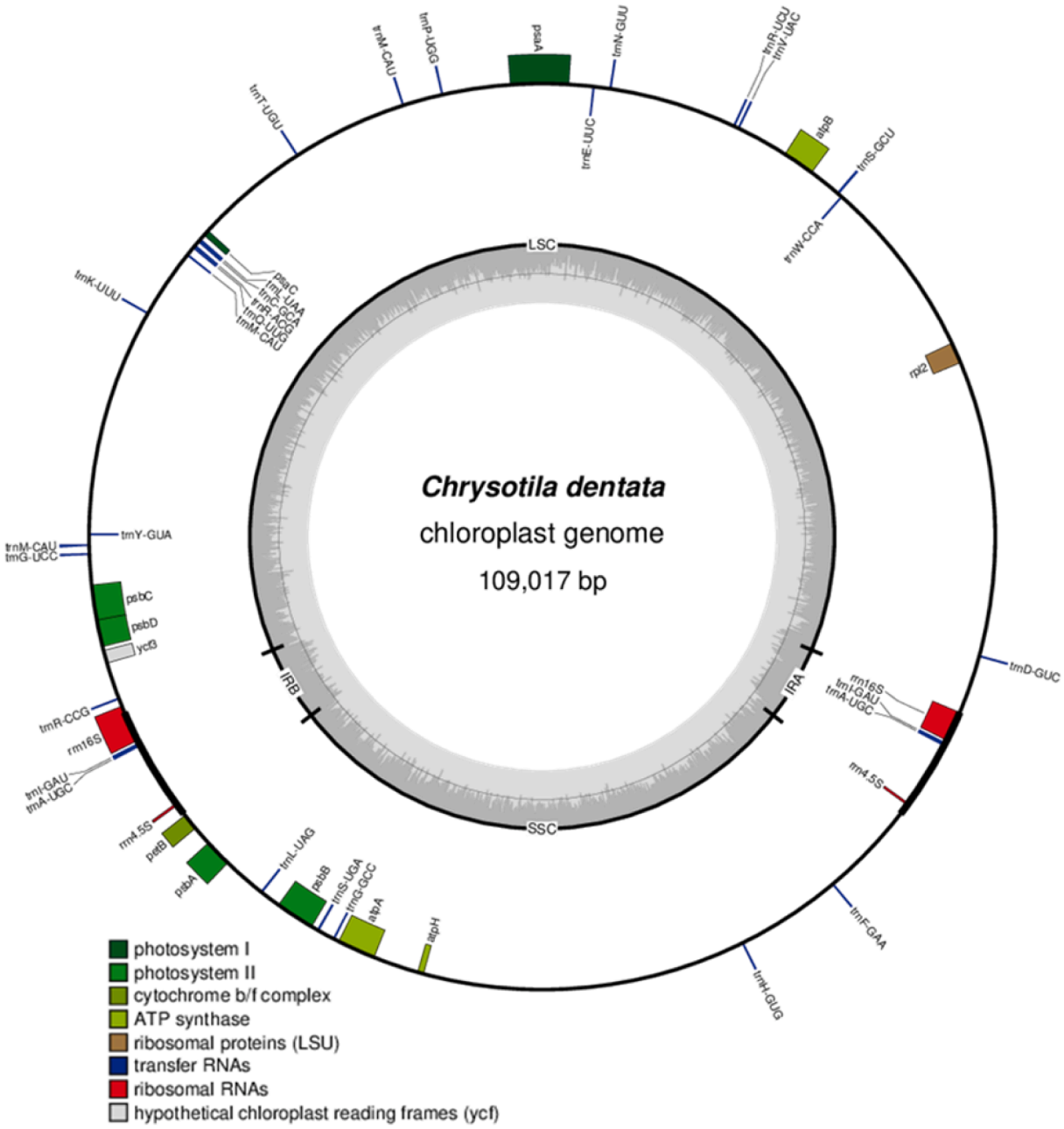

Fig. 1. Gene map of *Chrysotila dentata*. Thick lines indicate the extent of the inverted repeat regions (IRA and IRB), which separate the genome into small single copy (SSC) and large single copy (LSC) single copy regions. Genes shown in inner circle transcribe clockwise whereas the genes in outer circle transcribe counterclockwise. The functional groups of the genes are represented by different colors.

chloroplast genome from other haplophytes will be very helpful to understand the functions of the genes and evolutionary lineage and their relationship with other species. This study reports the chloroplast genome sequencing and analysis of *C. dentata.*

2. Materials and methods

2.1. Algal strain

Chrysotila dentata was purchased from National Center for Marine Algae and Microbiota, East Boothbay ME, USA, strain no. CCMP646 (NCMA- https://ncma.bigelow.org/ccmp646, kindly presented by Dr. Fan Lu, Hubei University of Technology, Wuhan, China) and was cultured using K/2 medium (Keller et al., 1987) at 15 °C on a 12:12 light–dark cycle.

2.2. DNA extraction and chloroplast genome sequencing, assembly, and annotation

Chloroplast DNA of *C. dentata* was isolated using a widely used extraction method (McPherson et al., 2013). Briefly, approximately 40 g of *C. dentata* was homogenized using a blender in 400 mL of cold isolation buffer (1.25 M NaCl 50 mM Tris–HCl pH 8.0, 5 mM EDTA, 1% BSA, 10 mM 2-mercaptoethanol, 5% polyvinylpyrrolidone). The suspension was filtered and centrifuged at 1500 × *g* for 20 min at 4 °C. The resultant pellet was resuspended in wash buffer (10 mM Tris–HCl pH 8.0, 5 mM EDTA, 10 mM 2-mercaptoethanol, 100 μg/mL proteinase K) and the chloroplast DNA was separated using step gradient columns. Finally, the chloroplast DNA was amplified using a whole genome amplification kit (Qiagen) by following the manufacturer's protocol. The fragmented DNA was used to construct short-insert (minimum

Table 1
Gene composition in *C. dentata* chloroplast genome.

Group of genes	Name of genes
Subunits of ATP synthase	atpE, atpI, atpH, atpF, atpB, atpA
ribosomal proteins (LSU)	rpl14, rpl22, rpl2, rpl21, rpl20, rpl5, rpl16, rpl23, rpl19
cytochrome *b*/f complex	petB, petG, petA, petD
RNA polymerase	rpoC1, rpoB, rpoA
Subunit of rubisco	rbcL
other genes	ccs1, C5S92_pgp022, minD, chlI, C5S92_pgp067, tufA, psbN
hypothetical chloroplast reading frames(ycf)	ycf65, ycf4, ycf3
photosystem II	psbZ, psbL, psbK, psbH, psbJ, psbT, psbA, psbC, psbB, psbE psbD, psbF, psbI
photosystem I	psaJ, psaC, psaA
ribosomal proteins (SSU)	rps2, rps11, rps18, rps9, rps14, rps12, rps4, rps19, rps3, rps8 rps16, rps7

length of 50 bp) libraries and sequenced using Illumina Hiseq 4000 (Borgström et al., 2011). The data were trimmed using SOAPnuke 1.3.0 (Chen et al., 2018) and assembled with SPAdes 3.13.0 (Bankevich et al., 2012). The annotation of the cp genome was performed to predict the genes, rRNAs, and tRNAs in the genome using a bioinformatics tool CPGAVAS2 (http://47.96.249.172:16019/analyzer/home) (Shi et al., 2019). The physical cp genome map was drawn using the OGDRAWv1.2 program with default parameters (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Greiner et al., 2019). Similarly, tRNAscan-SE (Schattner et al., 2005) was used to identify the tRNA genes.

2.3. Characterization of repeat structure and simple sequence repeats (SSRs)

The sizes and locations of forward, palindrome, reverse and complement sequences were analyzed using REPuter program (Kurtz et al., 2001) with the following parameters: minimum repeat size of 30 bp, maximum computed repeats of 5000 and a Hamming distance of 3 (a sequence identity greater than 90%). Similarly, microsatellite identification tool MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html) was used for identifying the simple sequence repeats (SSRs). The setting for minimum number of SSRs was ten repeat units for mononucleotide, five repeat units for dinucleotide, four repeat units for trinucleotide. Similarly, three repeat units were used for tetra, penta and hexanucleotide.

2.4. Phylogenetic analysis

Available complete chloroplast sequence data for all related species of algae were downloaded from GenBank. Sequences were assembled and aligned using the Clustalw module in BioEdit v. 7.0.9.0 (Hall, 1999) with default settings. Alignments were also checked and manually edited, if necessary. Maximum Likelihood (ML) analysis with 1000 bootstrap and uncorrected k2p pair-wise distances were conducted using MEGA 7 (Kumar et al., 2016).

2.5. Synteny analysis

Synteny was analyzed using progressiveMauve software (Darling et al., 2010) for accessing the extent of chloroplast genomes more closely. For this, the cp genomes of *Chrysotila dentata, Phaeocystis antarctica* (JN117275) and *Phaeocystis globose* (KC900889) were uploaded to progressiveMauve software and the output results were analyzed.

3. Results and discussion

3.1. Genome analysis of Chrysotila dentata

The complete cp genome of *C. dentata* was assembled with a total of 109, 017 bp in size with the common quadripartite structure as seen in most land plants and is divided in four regions (Chaney et al., 2016, Hovde et al., 2014, Odintsova and Yurina 2006, Jansen et al., 2005). The four regions consist of a large single-copy (LSC) region, a small single-copy (SSC) region, separated by two inverted repeats (IRs) regions each (4513 bp each) (Fig. 1). The presence of identical IRs in the cp genome of *C. dentata* is like most of the land and algal chloroplasts (Hovde et al., 2014; Liu et al., 2017). The GC content in *C. dentata* cp genome is 37.2 %. As shown in Table 1, the chloroplast of *C. dentata* consists of 61 protein-coding genes (6 ATP synthase, 9 ribosomal proteins, 4 cytochrome *b*/f complex, 3 RNA polymerase, 1 RubisCO large subunit, 3 photosystem I, 13 photosystem II, 12 ribosomal proteins (SSU), 3 hypothetical chloroplast reading frames and 7 other genes, 30 tRNA genes and 6 rRNA genes. The protein coding genes in *C. dentata* are less than that of other haplophytes such as *Tisochrysis lutea* and *Isochrysis galbana* in each of which there are more than 100 protein coding genes (Méndez-Leyva AB, 2019, Fang et al., 2020).

Based on tRNA and protein-coding genes, the frequency of codon usage for *C. dentata* cp genome was estimated (Table 2 and Table 3) to have 8362 codons. Glycine is the most frequent amino acid in the genome with 824 codons (9.85%) and the Cysteine with 86 codons (1.02%) is the least frequent amino acid. The chloroplast genomes in algae have conserved features. However, the GC content varies in different algal species with the variation in different regions of the chloroplast genomes. The study of codon usage is helpful for understanding the evolutionary relationship processes, and the selection pressure on genes and genome structure (Yang et al., 2014). The similarities in the codon usage seen in the *C. dentata* and other haplophytes (Fang et al., 2020) shows that these species might have gone through a

Table 2
Information of tRNA of the *C. dentata* chloroplast genome.

tRNA begin	Bounds end	tRNA type	Anticodon	tRNA begin	Bounds end	tRNA type	Anticodon
2243	2316	Asp	GTC	100,723	100,795	Phe	GAA
21,783	21,871	Ser	GCT	111,260	111,333	Asp	GTC
26,355	26,426	Val	TAC	107,735	107,662	Ile	GAT
26,546	26,618	Arg	TCT	107,657	107,585	Ala	TGC
31,505	31,576	Asn	GTT	80,417	80,346	Gly	GCC
37,982	38,055	Pro	TGG	79,738	79,653	Ser	TGA
39,537	39,610	Met	CAT	77,123	77,042	Leu	TAG
44,033	44,105	Thr	TGT	61,371	61,290	Tyr	GTA
52,443	52,514	Lys	TTT	49,803	49,730	Met	CAT
61,705	61,789	Met	CAT	49,440	49,369	Gln	TTG
62,050	62,120	Gly	TCC	49,352	49,279	Arg	ACG
67,724	67,796	Arg	CCG	49,135	49,065	Cys	GCA
69,611	69,684	Ile	GAT	49,054	48,971	Leu	TAA
69,689	69,761	Ala	TGC	32,223	32,151	Glu	TTC
96,592	96,664	His	GTG	21,670	21,598	Trp	CCA

Table 3
Codon usage in *C. dentata* chloroplast genome.

Codon	Amino acid	Frequency	Number	Codon	Amino acid	Frequency	Number
GCA	Ala	38.636	324	AAC	Asn	23.849	200
GCC	Ala	6.201	52	AAT	Asn	12.402	104
GCG	Ala	12.163	102	CCA	Pro	24.326	204
GCT	Ala	35.058	294	CCC	Pro	0.954	8
TGC	Cys	1.431	12	CCG	Pro	4.293	36
TGT	Cys	8.824	74	CCT	Pro	16.456	138
GAC	Asp	10.017	84	CAA	Gln	25.28	212
GAT	Asp	28.858	242	CAG	Gln	8.109	68
GAA	Glu	26.711	224	AGA	Arg	9.54	80
GAG	Glu	19.079	160	AGG	Arg	1.669	14
TTC	Phe	32.435	272	CGA	Arg	5.247	44
TTT	Phe	25.042	210	CGC	Arg	4.77	40
GGA	Gly	17.648	148	CGT	Arg	24.565	206
GGC	Gly	8.824	74	AGC	Ser	4.293	36
GGG	Gly	5.962	50	AGT	Ser	13.117	110
GGT	Gly	65.824	552	TCA	Ser	24.088	202
CAC	His	12.163	102	TCC	Ser	0.954	8
CAT	His	12.163	102	TCG	Ser	6.678	56
ATA	Ile	3.339	28	TCT	Ser	15.502	130
ATC	Ile	18.364	154	ACA	Thr	26.473	222
ATT	Ile	46.029	386	ACC	Thr	1.669	14
AAA	Lys	24.565	206	ACG	Thr	6.439	54
AAG	Lys	11.448	96	ACT	Thr	19.318	162
CTA	Leu	29.573	248	GTA	Val	26.95	226
CTC	Leu	0.954	8	GTC	Val	2.146	18
CTG	Leu	2.385	20	GTG	Val	4.293	36
CTT	Leu	20.749	174	GTT	Val	37.443	314
TTA	Leu	37.92	318	TGG	Trp	23.134	194
TTG	Leu	4.054	34	TAC	Tyr	18.125	152
ATG	Met	23.611	198	TAT	Tyr	15.025	126

Table 4
Repeated structure in the chloroplast genome of *C. dentata.*

Repeat no.	Size(bp)	Type	Repeat 1 start	Repeat 2 Start
1	30	F	34,854	37,094
2	32	F	34,050	36,275
3	4514	P	68,328	104,503
4	146	P	1	68,181
5	108	P	197	68,022
6	106	P	95	68,127
7	82	P	75,389	75,389
8	72	P	129	68,127
9	53	P	148	68,127
10	53	P	58,933	58,933
12	48	P	77,708	77,708
13	45	P	26,709	26,709
14	33	P	74,084	74,124
15	38	P	31,995	31,995
16	34	P	80,527	80,527
17	30	P	18,056	180,096
18	30	P	60,750	60,783

Table 5
Simple sequence repeats in the cp genome of *C. dentata.*

SSR ID	Repeat Motif	Length (bp)	Start	End
1	(ATAA)3	3	5150	5161
2	(GATT)3	3	9046	9057
3	(TCCAAC)3	3	9521	9538
4	(AT)5	5	21,024	21,033
5	(TTTG)4	4	21,297	21,312
6	(A)11	11	21,465	21,475
7	(A)10	10	27,890	27,899
8	(TTAAA)3	3	35,388	35,402
9	(ATTT)3	3	39,238	39,249
10	(TTAT)3	3	39,631	39,642
11	(TAAT)3	3	41,238	41,249
12	(ATT)4	4	43,652	43,663
13	(A)10	10	49,578	49,587
14	(AAAG)3	3	53,332	53,343
15	(TAA)4	4	59,150	59,161
16	(ACCC)3	3	67,881	67,892
17	(AGC)5	5	74,745	74,759
18	(AT)5	5	76,898	76,907
19	(TATT)3	3	80,028	80,039
20	(TA)5	5	82,138	82,147
21	(TTAC)3	3	103,481	103,492

similar environmental stress in their ecological niche.

3.2. Repeat structure and simple sequence repeats (SSRs) analysis

REPuter analysis showed the presence of 18 pairs of repeats in the cp genome of *C. dentata* showing the copy size 30 or longer (Table 4). There are only 2 repeats containing forward repeats whereas 16 repeats are related to the palindromic repeats. There longer repeats are similar to those found in other algal species and such longer repeats might play an important role in sequence divergence of chloroplast genome (Cavalier-Smith, 2002).

Simple sequence repeats (SSRs) are well known as microsatellites. They are short (1–6 bp), tandemly repeated DNA sequences in the genome (Gur-Arie et al., 2000). The study of SSRs has applications in genetic diversity analysis and molecular marker assistance breeding (Barkley et al., 2006, Steele et al., 2006). The analysis of cp of *C. dentata* revealed a total of 21 microsatellites consisting of 3 mononucleotide SSR (14.28%), 3 dinucleotide SSR (14.28%), 3 trinucleotide SSR (14.28%), 10 tetranucleotide SSR (47.61%), 1 pentanucleotide SSR (4.76%) and 1 hexanucleotide SSR (4.76%) showing non-random distribution in the cp genome of *C. dentata* and can be used in future studies related to the genetic diversity and molecular marker development. Most of the SSRs are composed of A/T bases thus contributing to the AT richness of cp genome (Table 5) which is like other chloroplast genomes of algae and land plants (Fang et. al. 2020, Cheng et al 2017).

3.3. Phylogenetic analysis

The phylogenetic data set included 27 species from three major

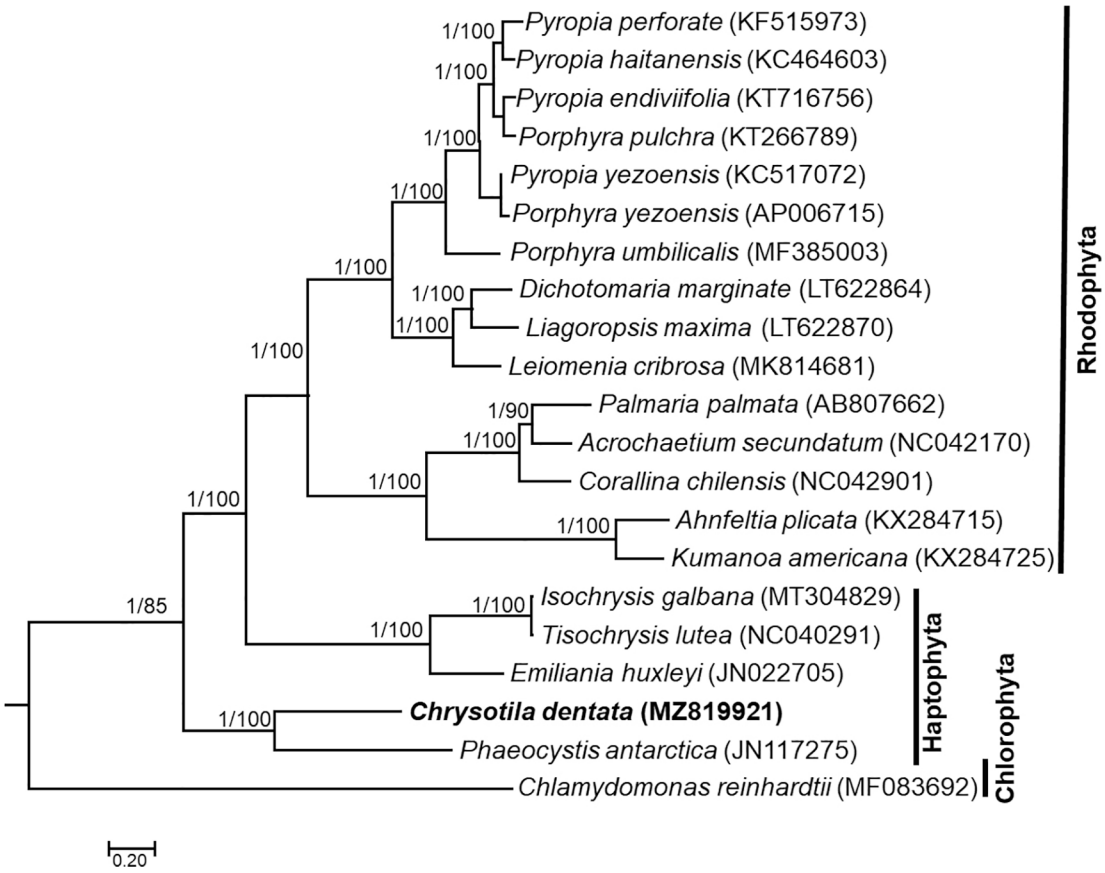

Fig. 2. ML tree based on complete chloroplast genome. Values on branches of the tree are Maximum likelihood (ML). *Chrysotila dentata* in this study shows a close relationship with *Phaeocystis antarctica* and *Phaeocystis globose. Chlamydomonas reinharditii* was selected as an outgroup. Genbank accession number are presented in parenthesis.

taxonomic division: Haptophyte and Rhodophyta including Chlorophyta as an outgroup. Most of the nodes were well-supported with high bootstrap values as shown in Fig. 2. The ML tree constructed with 1000 bootstraps support was consistent with other phylogenetic relationships (Baurain et al., 2010, Petersen et al., 2014, Fang et al., 2020) and indicates that the *Chrysotila dentata* in this study is a sister taxon of *Phaeocystis antarctica* (class: Coccolithophyceae) and *Phaeocystis globose* (Fig. 2). The uncorrelated genetic divergence between *C. dentata, P. antarctica* and *P. globose* were found to be 12.65 % and 12.76 % respectively (Table 6).

Species from Haptophyte and Rhodophyta lineages formed a monophyletic clade with high bootstrap support value (ML = 100). Our study is in line with the previous studies (Fang et al., 2020) that reported a monophyletic clade of Haptophyte and Rhodophyta.

3.4. Synteny analysis

The analysis of synteny helps to understand the ancestry of different species (Douglas and Penny, 1999). Fig. 3 shows the chloroplast genome synteny analysis of *Chrysotila dentata, Phaeocystis antarctica* (JN117275) and *Phaeocystis globose* (KC900889) using progressiveMauve software (Darling et al., 2010). This analysis shows the presence of more than 21 locally collinear blocks (LCBs). The compared cp genomes of *Chrysocapsa* and genus *Phaeocystis* showed confounding LCB connecting lines. The cp genome arrangement of *Phaeocystis antarctica* (JN117275) was similar to *Phaeocystis globose* (KC900889) whereas the cp genome of *Chrysotila* genus was poorly colinear. This study focused in the synteny analysis of closed related species but the inclusion of more cp genomes from different families might be helpful to compare the exact similarities and differences using synteny analysis as shown in the cp genomes of chaetophorales (Liu et al., 2020).

4. Conclusion

In this study, we sequenced and analyzed the complete cp genome of *C. dentata* which is composed of 109,017 bp. This is the first report of the cp genome analysis of genus *Chrysotila* showing the gene contents and orientations like those found in other algae. There are 18 pairs of repeats and 21 microsatellites in the cp genome of *C. dentata* and the distribution and location of these repeated structures might be helpful for developing the microsatellite markers and to understand the cp organization of other species of *Chrysotila.* The study of phylogenetic relationship of *C. dentata* with other algal species at various genetic distance showed that *C. dentata* has a close lineage with another halophyte *Phaeocystis antartica*. The results of this study can be used for discovering the genome organization and evolution of *Chrysotila.*

5. Data availability

The complete chloroplast genome sequence can be found in GenBank with accession no (MZ819921) after acceptance the manuscript.

Credit author statement

YPP, ZH, JRK and SP: Formal analysis, Writing - original draft, Writing - review & editing.

LF: Conceptualization, Data curation, Funding acquisition.

BL and WQ: Supervision, Writing - review & editing, Funding acquisition.

Table 6
Genetic uncorrected *p*-distance matrix of the complete chloroplast genome sequences between different species microalgae used in this study.

S. N.	Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	*C. dentata* (MZ819921)	0.00																								
2	*P. antarctica* (JN117275)	1.27																								
3	*P. globosa* (KC900889)	1.28	0.08																							
4	*A. secundatum* (NC042170)	2.29	2.34	2.35																						
5	*A. plicata* (KX284715)	2.39	2.48	2.52	1.34																					
6	*C. reinhardtii* (MF083692)	2.19	1.84	1.84	2.14	2.32																				
7	*C. parva* (MG520331)	1.96	1.23	1.23	2.47	2.56	2.16																			
8	*C. tobin* (KJ201907)	1.96	1.23	1.23	2.47	2.56	2.16	0.00																		
9	*C. chilensis* (NC042901)	2.31	2.43	2.47	0.40	1.36	2.24	2.54	2.55																	
10	*D. marginata* (LT622864)	2.18	2.32	2.32	1.83	1.91	1.95	2.33	2.33	1.77																
11	*E. huxleyi* (JN022705)	1.73	1.51	1.50	2.05	2.21	2.00	1.24	1.24	2.08	2.00															
12	*I. galbana* (MT304829)	1.81	1.39	1.39	2.03	2.16	1.88	1.14	1.14	2.08	1.98	0.76														
13	*K. americana* (KX284725)	2.40	2.43	2.47	1.30	0.43	2.30	2.47	2.48	1.33	1.87	2.12	2.11													
14	*L. cribrosa* (MK814681)	2.22	2.29	2.28	1.82	1.91	1.91	2.38	2.38	1.81	0.39	2.00	1.97	1.86												
15	*L. maxima* (LT622870)	2.21	2.34	2.34	1.85	1.94	1.98	2.40	2.40	1.84	0.32	2.06	2.03	1.91	0.41											
16	*P. palmata* (AB807662)	2.30	2.41	2.46	0.36	1.36	2.18	2.53	2.53	0.46	1.85	2.11	2.12	1.35	1.87	1.88										
17	*P. lutheri* (KC573041)	2.05	1.74	1.75	1.96	2.07	1.90	1.96	1.96	2.05	1.60	1.89	1.79	2.05	1.63	1.64	1.99									
18	*P. pulchra* (KT266789)	2.00	2.09	2.08	1.59	1.55	1.74	2.16	2.16	1.61	0.67	1.81	1.78	1.54	0.67	0.69	1.63	1.47								
19	*P. umbilicalis* (MF385003)	2.07	2.14	2.14	1.65	1.59	1.82	2.23	2.24	1.64	0.68	1.86	1.85	1.57	0.68	0.71	1.70	1.50	0.21							
20	*P. yezoensis* (AP006715)	2.03	2.10	2.09	1.60	1.56	1.78	2.20	2.20	1.63	0.68	1.86	1.82	1.54	0.69	0.71	1.65	1.48	0.15	0.23						
21	*P. endiviifolia* (KT716756)	2.01	2.09	2.08	1.59	1.56	1.76	2.17	2.17	1.61	0.67	1.82	1.80	1.54	0.67	0.69	1.63	1.49	0.10	0.21	0.15					
22	*P. haitanensis* (KC464603)	2.00	2.09	2.08	1.59	1.56	1.75	2.17	2.17	1.61	0.67	1.85	1.79	1.54	0.67	0.69	1.64	1.49	0.12	0.21	0.15	0.12				
23	*P. perforata* (KF515973)	2.02	2.08	2.07	1.60	1.56	1.75	2.17	2.17	1.62	0.67	1.84	1.80	1.55	0.68	0.69	1.64	1.49	0.15	0.23	0.17	0.15	0.12			
24	*P. yezoensis* (KC517072)	2.02	2.09	2.08	1.60	1.55	1.77	2.18	2.18	1.62	0.67	1.85	1.80	1.53	0.67	0.69	1.63	1.47	0.13	0.22	0.01	0.13	0.13	0.15		
25	*T. lutea* (NC040291)	1.81	1.40	1.39	2.03	2.16	1.88	1.14	1.14	2.08	1.98	0.76	0.00	2.11	1.98	2.03	2.12	1.79	1.78	1.85	1.82	1.80	1.79	1.80	1.80	0.00

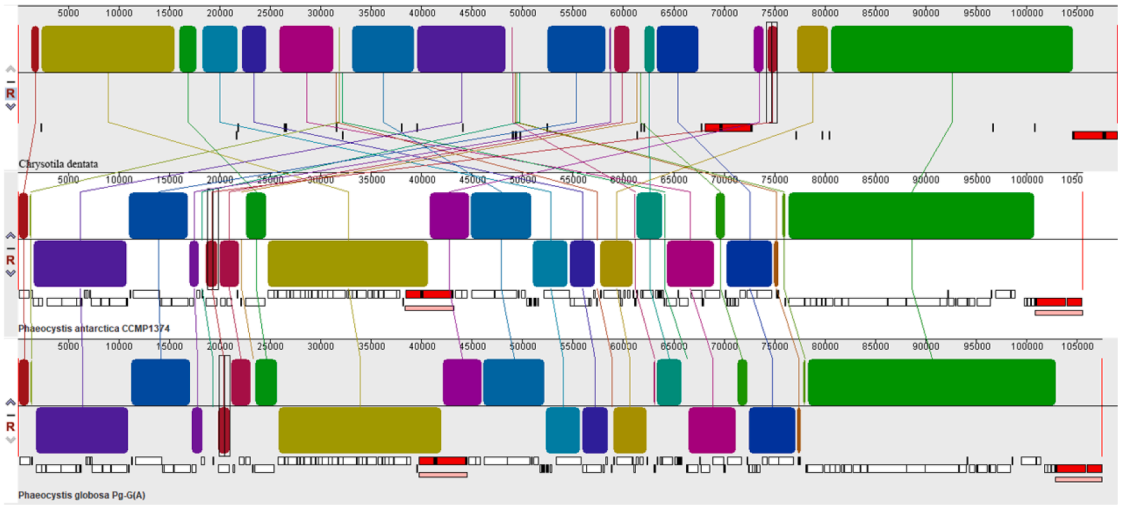

Fig. 3. Synteny comparison of the *Chrysotila dentata, Phaeocystis Antarctica* (JN117275) and *Phaeocystis globose* (KC900889) The coloured syntenic blocks are local collinear blocks; blocks above the centre line indicate they are on the same strand, and blocks below the centre line indicate they are on the opposite strand.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the China Postdoctoral Science Foundation (Grant No. 2019 M662748) to B. L. and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-05366) to W. Q.

References

Andersen, R.A., Kim, J.I., Tittley, I., Yoon, H.S., 2014. A re-investigation of *Chrysotila* (Prymnesiophyceae) using material collected from the type locality. Phycologia 53 (5), 463–473.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19 (5), 455–477.

Barkley, N.A., Roose, M.L., Krueger, R.R., Federici, C.T., 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theoretical and applied genetics 112 (8), 1519–1531.

Baurain, D., Brinkmann, H., Petersen, J., Rodriguez-Ezpeleta, N., Stechmann, A., Demoulin, V., Roger, A.J., Burger, G., Lang, B.F., Philippe, H., 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol Biol Evol 27 (7), 1698–1709.

Borgström, E., Lundin, S., Lundeberg, J., Nitabach, M.N., 2011. Large scale library generation for high throughput sequencing. PLoS One 6 (4), e19119.

Cavalier-Smith, T., 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses. Curr Biol 12 (2), R62–R64.

Chaney, L., Mangelson, R., Ramaraj, T., Jellen, E.N., Maughan, P.J., 2016. The complete chloroplast genome sequences for four Amaranthus species (Amaranthaceae). Applications in plant sciences 4 (9), 1600063. https://doi.org/10.3732/apps.160006310.3732/apps.1600063.s6.

Chen, X., Kameshwar, A.K.S., Chio, C., Lu, F., Qin, W., 2019. Effect of KNO3 on Lipid Synthesis and CaCO3 Accumulation in *Pleurochrysis cartarae* Coccoliths with a Special Focus on Morphological Characters of Coccolithophores. Int J Biol Sci 15, 2844.

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., Zhang, X., Wang, J., Yang, H., Fang, L., Chen, Q., 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 7, 1–6.

Cheng, H., Li, J., Zhang, H., Cai, B., Gao, Z., Qiao, Y., Mi, L., 2017. The complete chloroplast genome sequence of strawberry (Fragaria× ananassa Duch.) and comparison with related species of Rosaceae. PeerJ 5, e3919.

Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17, 134.

Darling, A.E., Mau, B., Perna, N.T., Stajich, J.E., 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5 (6), e11147. https://doi.org/10.1371/journal.pone.0011147.

Douglas, S.E., Penny, S.L., 1999. The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae. Journal of Molecular Evolution 48 (2), 236–244.

Fang, J., Lin, A., Yuan, X., Chen, Y., He, W., Huang, J., Zhang, X., Lin, G., Zhang, J., Xue, T., 2020. The complete chloroplast genome of *Isochrysis galbana* and comparison with related haptophyte species. Algal Res. 50, 101989. https://doi.org/10.1016/j.algal.2020.101989.

Greiner, S., Lehwark, P., Bock, R., 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47, W59-W64.

Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M., Kashi, Y., 2000. Simple sequence repeats in Escherichia coli: abundance, distribution, composition, and polymorphism. Genome research 10 (1), 62–71.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 95–98.

Hovde, B.T., Starkenburg, S.R., Hunsperger, H.M., Mercer, L.D., Deodato, C.R., Jha, R.K., Chertkov, O., Monnat Jr., R.J., Cattolico, R.A., 2014. The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles. BMC Genomics 15, 604.

Jansen, R.K., Raubeson, L.A., Boore, J.L., Depamphilis, C.W., Chumley, T.W., Haberle, R. C., Wyman, S.K., Alverson, A.J., Peery, R., Herman, S.J., Fourcade, H.M., 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods in enzymology 395, 348–384.

Johansen, J.R., Doucette, G.J., Barclay, W.R., Bull, J.D., 1988. The morphology and ecology of *Pleurochrysis carterae* var. *dentata* var. nov. (Prymnesiophyceae), a new coccolithophorid from an inland saline pond in New Mexico, USA. Phycologia 27 (1), 78–88.

Keller, M.D., Selvin, R.C., Claus, W., Guillard, R.R.L., 1987. Media for the culture of oceanic ultraphytoplankton. Journal of Phycology. 23, 633–638.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33, 1870–1874.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29, 4633–4642.

Liu, B., Hu, Y., Hu, Z., Liu, G., Zhu, H., 2020. Taxonomic scheme of the order Chaetophorales (Chlorophyceae, Chlorophyta) based on chloroplast genomes. BMC genomics 21 (1), 1–12.

Liu, H., Zhang, X., Sun, J., Satheeswaran, T., Zhang, G., Li, H., An, X., 2019. Morphology, phylogenetic position, and ecophysiological features of the coccolithophore *Chrysotila dentata* (Prymnesiophyceae) isolated from the Bohai Sea. China. Phycologia 58 (6), 628–639.

Liu, L.X., Li, R., Worth, J.R.P., Li, X., Li, P., Cameron, K.M., Fu, C.X., 2017. The Complete Chloroplast Genome of Chinese Bayberry (*Morella rubra*, Myricaceae): Implications for Understanding the Evolution of Fagales. Front. Plant Sci. 8, 968.

McPherson, H., van der Merwe, M., Delaney, S.K., Edwards, M.A., Henry, R.J., McIntosh, E., Rymer, P.D., Milner, M.L., Siow, J., Rossetto, M., 2013. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. BMC Ecol 13 (1), 8. https://doi.org/10.1186/1472-6785-13-8.

Méndez-Leyva, A.B., Guo, J., Mudd, E.A., Wong, J., Schwartz, J.-M., Day, A., 2019. The chloroplast genome of the marine microalga *Tisochrysis lutea*. Mitochondrial DNA Part B 4 (1), 253–255.

Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. USA 104 (49), 19363–19368.

Odintsova, M.S., Yurina, N.P., 2006. In: Biotechnology Intelligence UnitBiotechnological Applications of Photosynthetic Proteins: Biochips, Biosensors and Biodevices. Springer US, Boston, MA, pp. 57–72. https://doi.org/10.1007/978-0-387-36672-2_6.

Petersen, J., Ludewig, A.K., Michael, V., Bunk, B., Jarek, M., Baurain, D., Brinkmann, H., 2014. Chromera velia, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). Genome biology and evolution 6 (3), 666–684.

Reid, E.L., Worthy, C.A., Probert, I., Ali, S.T., Love, J., Napier, J., Littlechild, J.A., Somerfield, P.J., Allen, M.J., 2011. Coccolithophores: functional biodiversity, enzymes and bioprospecting. Mar Drugs 9 (4), 586–602.

Saez, A.G., Probert, I., Geisen, M., Quinn, P., Young, J.R., Medlin, L.K., 2003. Pseudo-cryptic speciation in coccolithophores. PNAS 100 (12), 7163–7168.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33 (Web Server), W686–W689.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., Liu, C., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47, W65-W73.

Smith, D.R., Arrigo, K.R., Alderkamp, A.-C., Allen, A.E., 2014. Massive difference in synonymous substitution rates among mitochondrial, plastid, and nuclear genes of *Phaeocystis* algae. Mol Phylogenet Evol 71, 36–40.

Steele, K.A., Price, A.H., Shashidhar, H.E., Witcombe, J.R., 2006. Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theoretical and Applied Genetics 112 (2), 208–221.

Yang, X., Luo, X., Cai, X., 2014. Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. Parasit. Vectors 7, 527.