

Comparison of *Auxenochlorella protothecoides* and *Chlorella* spp. Chloroplast Genomes: Evidence for Endosymbiosis and Horizontal Virus-like Gene Transfer

Sang-Hyuck Park ^{1,2}, John A. Kyndt ^{3,*} and Judith K. Brown ¹

¹ School of Plant Sciences, The University of Arizona, Tucson, AZ 85721, USA; sanghyuck.park@csupueblo.edu (S.-H.P.); jbrown@ag.arizona.edu (J.K.B.)

² Institute of Cannabis Research, Colorado State University-Pueblo, Pueblo, CO 81001, USA

³ College of Science and Technology, Bellevue University, Bellevue, NE 68005, USA

* Correspondence: jkyndt@bellevue.edu

Supplementary Figures

Citation: Park, S.-H.; Kyndt, J.A.; Brown, J.K. Comparison of *Auxenochlorella protothecoides* and *Chlorella* spp. Chloroplast Genomes: Evidence for Endosymbiosis and Horizontal Virus-like Gene Transfer. *Life* **2022**, *12*, 458. <https://doi.org/10.3390/life12030458>

Academic Editor: Francois Lefort

Received: 10 February 2022

Accepted: 16 March 2022

Published: 20 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

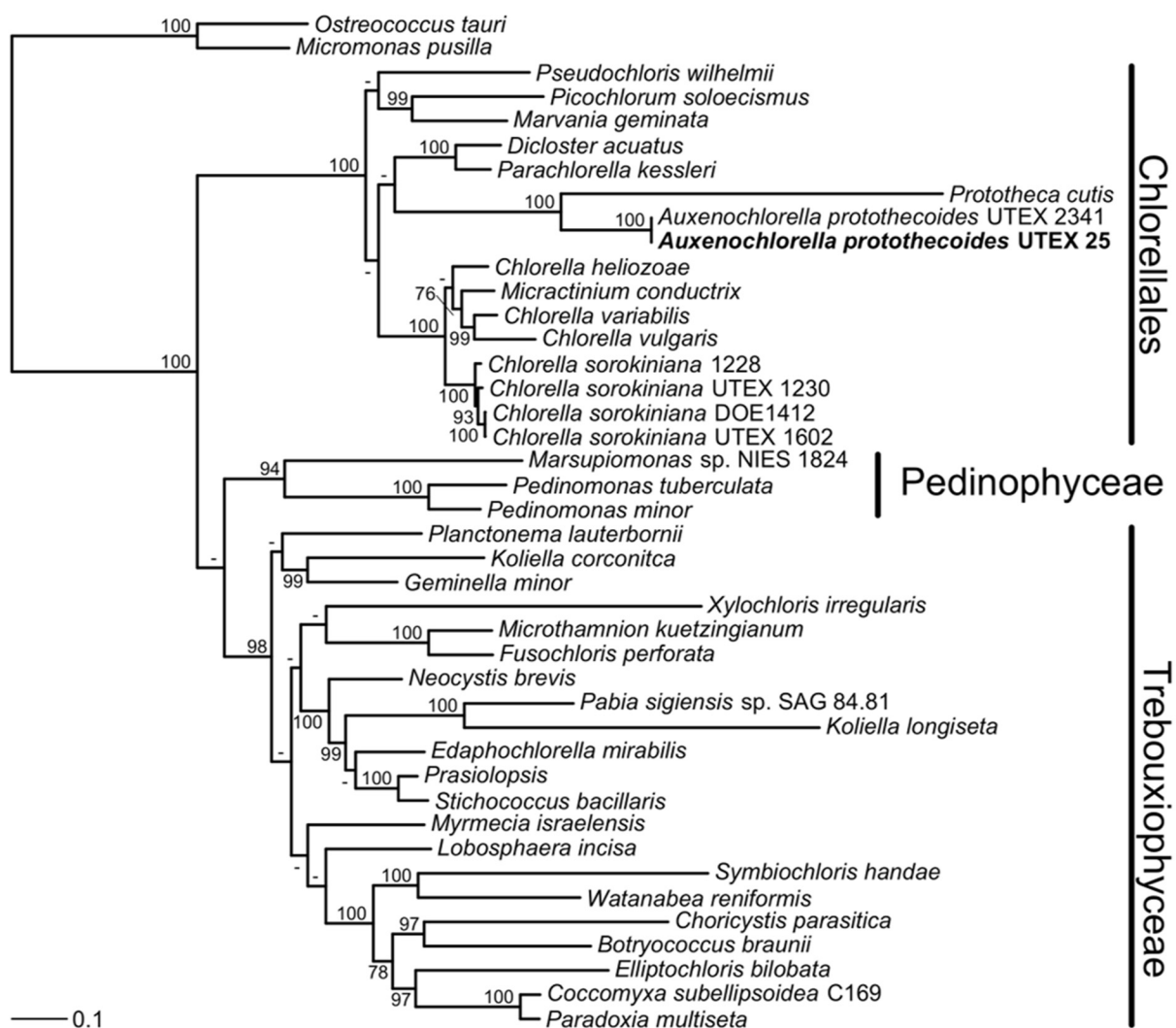
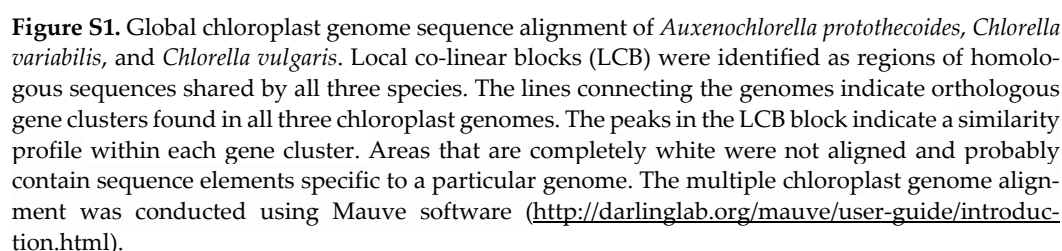


Figure S2. Phylogenetic relationships among the Trebouxiophyceae chloroplast genomes using Maximum Likelihood. Numbers placed at major nodes indicate bootstrap confidence values at $\geq 70\%$ for 1000 iterations.

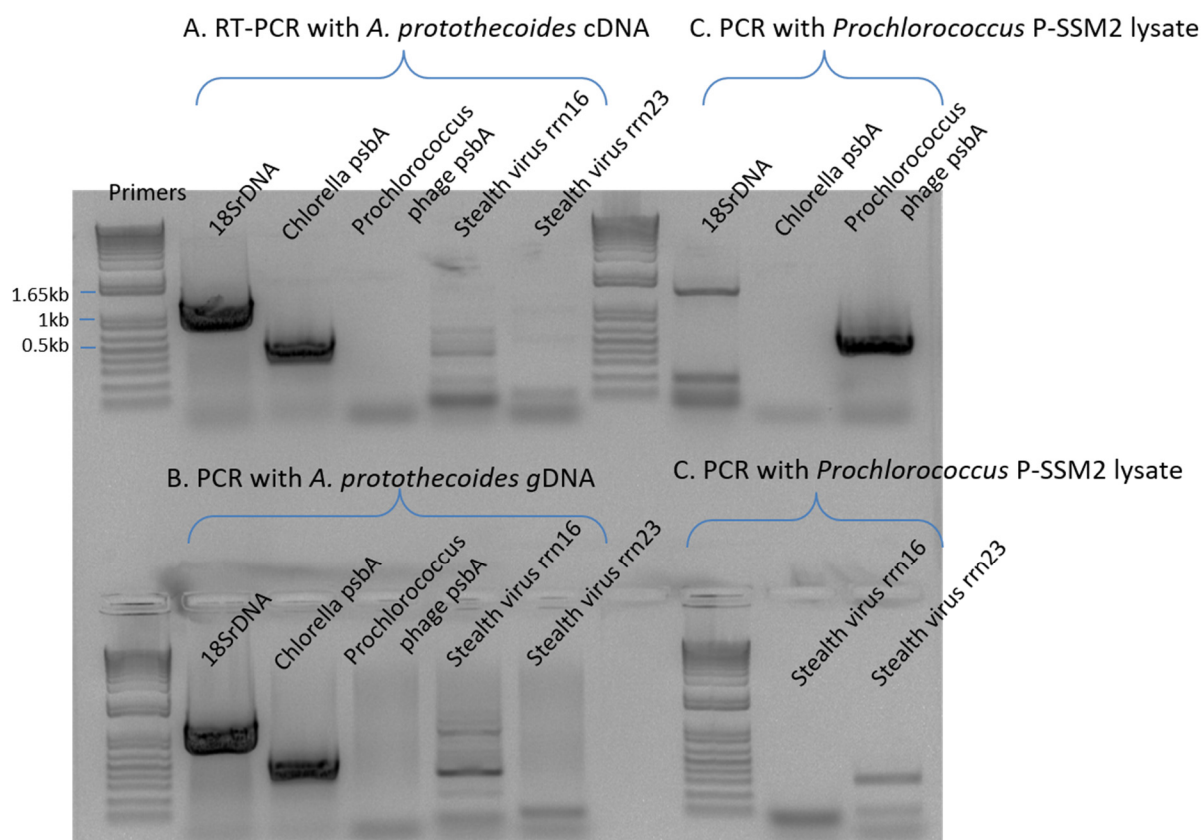


Figure S3. An original gel image used in Figure 6. Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) on *Auxenochlorella protothecoides* cDNA for examination of presence of viral genes (*Prochlorococcus* P-SSM2 *psbA*, *Stealth virus rrn16* and *rrn23*). 18S ribosomal DNA is used as internal control. A. RT-PCR was performed using gene specific primers of *A. protothecoides psbA*, *Prochlorococcus* phage *psbA* (588nt), *Stealth virus rrn16* (230nt) and *Stealth virus rrn23* (331nt). B. Conventional PCR was performed on *C. protothecoides* gDNA to show the presence of viral genes. C. Conventional PCR was performed on *Prochlorococcus* P-SSNM2 lysate to confirm the presence of *psbA*.

Supplementary Tables

Table S1. Inventory of predicted genes within the *Auxenochlorella protothecoides*, *Chlorella variabilis*, and *Chlorella vulgaris* chloroplast genomes. trn: transfer RNA; rrn: ribosomal RNA; rpl: ribosomal protein large subunit; rrs: ribosomal protein small subunit; chl: chlorophyll biosynthesis, psa/psb: photosystem I/II subunits; rpo: RNA polymerase subunit; ycf: hypothetical chloroplast ORF; cysA: sulfate transport ATP-binding protein; ftsH/minE: plastid division; cemA: envelope membrane. †Gene including intron. *Not found in the other Chlorellaceae species in Table 1. § Partial sequences. The copy number for each is indicated parenthetically.

<i>A. protothecoides</i> UTEX 25 84,576 bp (114 genes) (GC: 30.8%)	<i>C. variabilis</i> 124,579 bp (113 genes) (GC: 34%)	<i>C. vulgaris</i> 150,613 bp (120 genes) (GC: 32%)
Photosynthesis		
Photosystem I		
psaA psaB psaC psal psaj psaM	psaA psaB psaC psal psaj psaM	psaA psaB psaC psal psaj psaM
Photosystem II		
psbA psbB psbC psbD psbE psbF psbH psbI psbJ psbK psbL psbM psbN psbT psbZ	psbA (2) psbC(3) psbD psbE psbF psbH psbI psbJ psbK psbL psbM psbN psbT psbZ psi_psbT	psbA psbC psbD psbE psbF psbH psbI psbJ psbK psbL psbM psbN psbT psbZ psi_psbT
Cytochrome		
petA petB petD petG petL	petA petB petD petG petL	petA petB petD petG petL
ATP synthesis		
atpA atpB atpE atpF atpH atpI	atpA atpB atpE atpF atpH atpI	atpA atpB atpE atpF atpH atpI
Chlorophyll synthesis		
chlB chlI (2) chlL chlN	chlB chlI chlL †chlN	chlB chlI †chlL chlN
Rubisco		
rbcL	rbcL	rbcL
Transcription/translation		
RNA polymerase		
rpoA rpoB rpoC1 rpoC2	rpoA rpoB rpoC1 rpoC2	rpoA rpoB rpoC1 rpoC2, §rpoC2
Translation factor		
infA tufa	infA tufa	infA tufa
Chloroplast division		
minD ftsH	minD ftsH	minD *minE ftsH
Ribosomal RNAs		
rrn5 rrn16 rrn23 (2)	rrn5 rrn16 rrn23(2)	rrn5 rrn16 (4) rrn23
Ribosomal proteins		
Large subunits		
rpl2 rpl5 rpl12 rpl14 rpl16 rpl19 rpl20 rpl23 rpl32 rpl36	rpl2 rpl5 rpl12 rpl14 rpl16 rpl19 rpl20 rpl23 rpl32 rpl36	rpl2 rpl5 rpl12 rpl14 rpl16 rpl19 rpl20 rpl23 rpl32 rpl36
Small subunits		
rps2 rps3 rps4 rps7 rps8 rps9 rps11 rps12 rps12_5end rps14 rps18 rps19	rps2 rps3 rps4 rps7 rps8 rps9 rps11 rps12 rps12_3end rps14 rps18 rps19	rps2 rps3 rps4 rps7 rps8 rps9 rps11 rps12 rps12_3end rps14 rps18 rps19
Transfer RNA		
trnA-UGC trnC-GCA trnD-GUC trnE-UUC trnF-GAA trnFM-CAU	trnA-UGC trnC-GCA(2) trnD-GUC trnE- UUC trnF-GAA trnFM-CAU trnG-	trnA-UGC trnC-GCA trnD-GUC trnE- UUC trnF-GAA trnFM-CAU trnG-GCC(2)

<i>trnG</i> -GCC	<i>trnG</i> -UCC	<i>trnH</i> -GUG	GCC(2)	<i>trnG</i> -UCC	<i>trnH</i> -GUG	<i>trnI</i> -CAU	<i>trnG</i> -UCC	<i>trnH</i> -GUG	<i>trnI</i> -CAU	<i>trnI</i> -					
<i>trnH</i> -GUG	<i>trnI</i> -CAU	<i>trnI</i> -GAU	<i>trnK</i> -	<i>trnI</i> -GAU	<i>trnK</i> -UUU	<i>trnL</i> -CAA	<i>trnL</i> -GAU	<i>trnK</i> -UUU	<i>trnL</i> -CAA	<i>trnL</i> -GAG					
UUU	<i>trnL</i> -GAG	<i>trnL</i> -UAA	<i>trnL</i> -UAG	GAG	<i>trnL</i> -UAG	<i>trnM</i> -CAU	<i>trnN</i> -GUU	<i>trnL</i> -UAG	<i>trnM</i> -CAU	<i>trnN</i> -GUU	<i>trnP</i> -				
<i>trnM</i> -CAU	<i>trnN</i> -GUU	<i>trnP</i> -GGG	<i>trnP</i> -GGG	<i>trnP</i> -UGG	<i>trnQ</i> -UUG	<i>trnR</i> -GGG	<i>trnP</i> -UGG	<i>trnQ</i> -UUG	<i>trnR</i> -ACG						
<i>trnP</i> -UGG	<i>trnQ</i> -UUG	<i>trnR</i> -ACG	ACG	<i>trnR</i> -CCG	<i>trnR</i> -UCU	<i>trnS</i> -GCU	<i>trnR</i> -CCG	<i>trnR</i> -UCU	<i>trnS</i> -GCU	<i>trnS</i> -					
<i>trnR</i> -CCG	<i>trnR</i> -UCU	<i>trnS</i> -GCU	<i>trnS</i> -	<i>trnS</i> -GGA	<i>trnS</i> -UGA	<i>trnT</i> -GGU	<i>trnT</i> -GGA	<i>trnS</i> -UGA	<i>trnT</i> -GGU	<i>trnT</i> -UGU					
GGA	<i>trnS</i> -UGA	<i>trnT</i> -GGU	<i>trnT</i> -UGU	UGU	<i>trnV</i> -UAC	<i>trnW</i> -CCA	<i>trnY</i> -GUA	<i>trnV</i> -UAC	<i>trnV</i> -UAC	<i>trnW</i> -CCA	<i>trnY</i> -				
<i>trnV</i> -UAC	<i>trnW</i> -CCA	<i>trnY</i> -GUA								GUA					
Fatty acid biosynthesis															
<i>accD</i>				<i>accD</i>				<i>accD</i>							
Others															
<i>ccsA</i>	<i>cemA</i>	<i>clpP</i>	<i>cysT</i>	<i>TilS</i>	<i>ccsA</i>	<i>cemA</i>	<i>clpP</i>	<i>cysA</i>	<i>cysT</i>	<i>ccsA</i>	<i>cemA</i> (2)	<i>clpP</i>	<i>cysA</i>	<i>cysT</i>	<i>ycf3</i>
<i>ycf3</i>	<i>ycf4</i>	<i>ycf12</i>	<i>ycf78</i>		<i>ycf3</i>	<i>ycf4</i>	<i>ycf12</i>	<i>ycf78</i>		<i>ycf4</i>	<i>ycf12</i>	<i>*ycf62</i>	<i>ycf78</i>		

Table S2. Protein sequences ($n = 26$) identified based on the best-fitting partitioning ($n = 17$) and protein substitution models. The best-fitting partition and protein substitution scheme for 26 proteins, resulting in 17 partitions, determined by PartitionFinder version 2.1.1.

Partition number	Protein	Description	Number of sites	Substitution model
1	<i>atpB</i>	ATP synthase beta-subunit	598	LG4X
2	<i>psaA</i>	P700 chlorophyll a-apoprotein A1	1486	MTZOA+I+G+F
	<i>psaB</i>	P700 chlorophyll a-apoprotein A2		
3	<i>psaC</i>	subunit VII of photosystem I (Fe-Spolypeptide)	556	LG4X
	<i>rbcL</i>	large subunit of Rubisco		
4	<i>psbC</i>	photosystem II CP43 apoprotein	488	LG4X
5	<i>RPL12</i>	ribosomal protein L12	136	LG+G+F
6	<i>RPL14</i>	ribosomal protein L14	217	LG4X
	<i>RPS19</i>	ribosomal protein S19		
7	<i>RPL16</i>	ribosomal protein L16	144	LG4X
8	<i>RPS4</i>	ribosomal protein S4	654	LG+I+G+F
	<i>RPS8</i>	ribosomal protein S8		
	<i>RPL19</i>	ribosomal protein L19		
9	<i>RPL5</i>	ribosomal protein L5	494	LG4X
	<i>RPL2</i>	ribosomal protein L2		
10	<i>RPL20</i>	ribosomal protein L20	122	LG4X
11	<i>RPL23</i>	ribosomal protein L23	737	JTTDCMUT+G+F
	<i>RPS18</i>	ribosomal protein S18		
	<i>RPS2</i>	ribosomal protein S2		
	<i>RPL32</i>	ribosomal protein L32		
12	<i>RPS11</i>	ribosomal protein S11	136	LG4X
13	<i>RPS12</i>	ribosomal protein S12	127	CPREV+I+G
14	<i>RPS14</i>	ribosomal protein S14	102	LG4X
15	<i>RPS3</i>	ribosomal protein S3	425	MTREV+I+G+F
16	<i>RPS7</i>	ribosomal protein S7	157	LG4X
17	<i>RPS9</i>	ribosomal protein S9	205	CPREV+I+G

Table S3. Results of BLASTn searches for the *Auxenochlorella protothecoides* UTEX 25 cpDNA against genome sequences of the Cyanobacteria (taxid:1117) available in the National Center for Biotechnology Information (NCBI) GenBank database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). For top hit selection, the following filtering criteria (percent identity: >86%, E-value: 0, and query coverage: >1%) were used.

	NCBI GenBank Accession #	Cyanobacterium	Gene	Query Coverage	Percent Identity	Gaps
1	EU703214.1	Uncultured cyanobacterium	16S rRNA	1%	88.94%	54/1420(3%)
2	FJ901770.1	Uncultured cyanobacterium	16S rRNA	1%	89.49%	47/1313(3%)
3	CP060212.1	<i>Limnospira</i> sp. BM	23S rRNA	9%	86.55%	31/1323(2%)
4	MT863733.1	<i>Limnospira</i> sp. BM	23S rRNA	2%	86.48%	32/1324(2%)
5	AP012549.1	Cyanobacterium endosymbiont of <i>Epithemia turgida</i> isolate EtSB Lake Yunoko	23S rRNA	4%	87.98%	31/1190(2%)
6	NR076532.1	<i>Crocospaera subtropica</i> ATCC 51142	23S rRNA	2%	87.64%	29/1189(2%)
7	CP000806.1	<i>Crocospaera subtropica</i> ATCC 51142	23S rRNA	4%	87.64%	29/1189(2%)
8	AP018202.2	<i>Synechococcus vulcanus</i> NIES-2134	23S rRNA	4%	87.53%	29/1187(2%)
9	AP018341.1	Cyanobacterium endosymbiont of <i>Rhopalodia gibberula</i>	23S rRNA	4%	87.39%	31/1190(2%)
10	AY584507.1	<i>Eubalthece</i> sp. BAA001	23S rRNA	2%	86.29%	28/1240(2%)
11	NR102529.1	<i>Stanieria cyanosphaera</i> PCC 7437	23S rRNA	2%	86.33%	24/1222(1%)
12	CP003653.1	<i>Stanieria cyanosphaera</i> PCC 7437	23S rRNA	6%	86.33%	24/1222(1%)
13	CP051167.1	<i>Oxynema</i> sp. AP17	23S rRNA	4%	86.81%	16/1183(1%)
14	NR076263.1	<i>Synechococcus elongatus</i> PCC 6301	23S rRNA	2%	86.75%	20/1185(1%)
15	CP000100.1	<i>Synechococcus elongatus</i> PCC 7942 = FACHB-805	23S rRNA	4%	86.75%	20/1185(1%)
16	LC455673.1	<i>Kamptonema laetevirens</i> NIES-31	23S rRNA	2%	86.51%	25/1201(2%)
17	X00512.1	<i>Synechococcus elongatus</i> PCC 6301	23S rRNA	2%	86.67%	22/1185(1%)
18	AP017375.1	<i>Stanieria</i> sp. NIES-3757	23S rRNA	8%	86.59%	19/1186(1%)
19	CP051168.1	<i>Phormidium</i> sp. ETS-05	23S rRNA	4%	86.59%	21/1186(1%)
20	CP000240.1	<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	23S rRNA	4%	86.62%	26/1188(2%)
21	LC455660.1	<i>Planktothrix agardhii</i>	23S rRNA	2%	86.46%	20/1189(1%)
22	LC455656.1	<i>Planktothrix agardhii</i>	23S rRNA	2%	86.46%	20/1189(1%)
23	AP017991.1	<i>Planktothrix agardhii</i> NIES-204	23S rRNA	5%	86.46%	20/1189(1%)
24	LO018304.1	<i>Planktothrix agardhii</i>	23S rRNA	5%	86.46%	20/1189(1%)
25	CP000239.1	<i>Synechococcus</i> sp. JA-3-3Ab	23S rRNA	4%	86.48%	26/1191(2%)
26	NR076758.1	<i>Nostoc azollae</i> 0708	23S rRNA	2%	86.45%	23/1188(1%)
27	CP002059.1	<i>Nostoc azollae</i> 0708	23S rRNA	4%	86.45%	23/1188(1%)
28	AP018254.1	<i>Calothrix</i> sp. NIES-3974	23S rRNA	6%	86.29%	20/1189(1%)
29	LC455612.1	<i>Phormidium</i> sp. NIES-2121	23S rRNA	2%	86.18%	19/1187(1%)
30	CP016483.1	<i>Synechococcus</i> sp. PCC 8807	23S rRNA	4%	86.16%	20/1185(1%)
31	CP040360.1	<i>Synechococcus</i> sp. PCC 11901	23S rRNA	4%	86.08%	20/1185(1%)
32	CP013998.1	<i>Synechococcus</i> sp. PCC 73109	23S rRNA	4%	86.08%	20/1185(1%)
33	CP000951.1	<i>Synechococcus</i> sp. PCC 7002	23S rRNA	4%	86.08%	20/1185(1%)
34	NR102540.1	<i>Gloeocapsa</i> sp. PCC 7428	23S rRNA	2%	86.03%	22/1188(1%)
35	NR102530.1	<i>Cylindrospermum stagnale</i> PCC 7417	23S rRNA	2%	86.04%	23/1189(1%)
36	CP003646.1	<i>Gloeocapsa</i> sp. PCC 7428	23S rRNA	7%	86.03%	22/1188(1%)
37	CP003642.1	<i>Cylindrospermum stagnale</i> PCC 7417	23S rRNA	4%	86.04%	23/1189(1%)