Complete Chloroplast Genome Sequence of Freshwater Araphid Pennate Diatom Alga *Synedra acus* from Lake Baikal

Yuri P. Galachyants (Corresponding author)
Limnological Institute, Siberian Branch of the Russian Academy of Sciences
3 Ulan-Batorskaya st., Irkutsk 664033, Russia
Tel: 7-395-242-6504 E-mail: yuri.galachyants@lin.irk.ru

Alexey A. Morozov
Limnological Institute, Siberian Branch of the Russian Academy of Sciences
3 Ulan-Batorskaya st., Irkutsk 664033, Russia

Andrey V. Mardanov, Alexey V. Beletsky & Nikolay V. Ravin
Centre "Bioengineering", Russian Academy of Sciences
pr. 60-letiya Oktyabrya 7/1, Moscow 117312, Russia

Darya P. Petrova & Yelena V. Likhoshway
Limnological Institute, Siberian Branch of the Russian Academy of Sciences
3 Ulan-Batorskaya st., Irkutsk 664033, Russia

Received: November 28, 2011 Accepted: December 13, 2011 Published: January 1, 2012
doi:10.5539/ijb.v4n1p27 URL: http://dx.doi.org/10.5539/ijb.v4n1p27

This work was financially supported by the Program of the Presidium of the Russian Academy of Sciences "Molecular and Cell Biology" (project # 6.3)

Abstract

Complete chloroplast genome of diatom alga *Synedra acus* possesses canonical quadripartite structure with two inverted repeats containing ribosomal RNA gene loci that separate small and large single-copy regions. Chloroplast genome maps as a circular molecule of 116 251 bp. It encodes 27 tRNAs, three rRNAs, two small RNA genes, and 128 protein-coding genes. Comparison of the genic features across diatom chloroplast genomes reveals the absence of an overlap between *atpD* and *atpF* gene coding sequences that is present in other plastid genomes of diatom origin. This feature is a clear synapomorphy of *S. acus* plastid genome that is likely a result of either relaxed constraints or extensive selection forces acting upon *atpF* gene. We also characterized nuclear-encoded acyl-carrier protein gene with chloroplastic targeting in *S. acus*. The transfer of *acpp* gene into the nuclear host genome is hypothesized to have occurred independently in several lineages of diatoms.

Keywords: Diatoms, Chloroplast genome, Next-generation DNA sequencing

1. Introduction

Diatom algae are a group of unicellular eukaryotes, which is one of the major photosynthetic producers of organic carbon in the world, providing about 20% of net primary production (Falkowski *et al.* 2004). Chloroplast of diatoms had initially a cyanobacterial origin and was inherited from an ancient red alga during its secondary endosymbiosis with a heterotrophic eukaryote. In the course of endosymbiosis, several genes were transferred to the nuclear host genome, whereas some others were lost since the host genes had incorporated into the chloroplast molecular machinery. Diatoms inhabit almost all water environments, from marine to ultra-oligotrophic freshwater bodies and from pack-ice to thermal springs (Round *et al.* 1997). During their

evolution from *ca.* 240 mya (Medlin 2009), many species of diatoms were formed and disappeared. The present number of diatom species is reported to be up to 10^5-10^6 including cryptic ones (Round *et al.* 1997). Based on cell wall symmetry and other morphological features, diatoms could be separated into several large groups which are listed in the order of their divergence: radial and bipolar centrics, araphids, and raphid pennates, with only the last group believed to be monophyletic (Sims *et al.* 2006).

Several chloroplast genomes of diatoms have been sequenced to date, including radial centrics *Thalassiosira pseudonana* (Oudot-Le Secq *et al.* 2009) and *Thalassiosira oceanica* (Lommer *et al.* 2010), bipolar centric *Odontella sinensis* (Kowallik *et al.* 1995), and raphid pennates *Phaeodactylum tricornutum* (Oudot-Le Secq *et al.* 2009) and *Fistulifera* sp. JPCC DAO0580 (Tanaka *et al.* 2011). Additionally, the plastid genomes of two dinoflagellates *Kryptoperidinium foliaceum* and *Durinskia baltica*, also known as dinotoms because they harbouring a diatom endosymbiont, were sequenced recently (Imanian *et al.* 2010). Here we report the complete chloroplast genome sequence of a freshwater araphid pennate diatom *Synedra acus* subsp. *radians* (Kütz.) Skabitsch. from Lake Baikal. *S. acus* species is a common inhabitant of freshwater environments. In Lake Baikal, eventual under-ice blooms of *S. acus* have been reported to feed zooplankton and left records in Pleistocene and Holocene sediments of Lake Baikal (Grachev *et al.* 1998). Like with the mitochondrial genome of *S. acus* (Ravin *et al.* 2010), we used total DNA isolation procedure followed by shot-gun pyrosequencing to decipher its chloroplast DNA (cpDNA).

2. Materials and Methods

2.1 Culture growth and DNA isolation

Culture of *S. acus* was isolated from the phytoplankton of Listvennichny Bay of Lake Baikal. An axenic culture was subsequently obtained in the Limnological Institute by a series of single-cell passaging combined with antibiotics and detergent treatments (Shishlyannikov *et al.* 2011). Diatom cells were cultivated using sterile technique and 15-liter glass bottles with DM medium (Thompson *et al.* 1988) up to a density of $3\text{-}4*10^4$ cells/mL. Diatom cells were collected on a filter with 5 μm pores and washed with sterile medium. Total DNA was isolated as in Ravin *et al.* (2010).

2.2 Sequencing and assembly

The chloroplast genome of *S. acus* was sequenced using GS FLX instrument (Roche/454 Life Sciences, Branford, CT, USA). A shotgun genome library was constructed using GS FLX Titanium General Library Preparation Kit. Shot-gun reads from 2 FLX runs were assembled with MIRA 3.0.4 and two chloroplast-specific contigs were identified by their similarities to other chloroplast genomes. We also prepared a mate-pair genomic library using Mate Pair Library Prep Kit v2 (Illumina, CA, USA). Shortly, genomic DNA was sonicated followed by ligation of circularization adapters. Fragments in the size range of 1.5-4 kbp were eluted from a preparative agarose gel, and amplified with eight PCR cycles using a proof-reading polymerase. After circularization, the linear fragments were digested with DNase I followed by fragmentation of circular DNA with ultrasound. Paired-end adapters were ligated to fragments carrying the circularization adapters for sequencing with Genome Analyzer IIx instrument (Illumina, CA, USA). Finally, mate-pair DNA fragments were PCR-amplified with 17 cycles using a proof-reading enzyme. One lane of GAIIx run resulted in 18M of 2X52bp reads. These reads were mapped to chloroplast-specific contigs with MIRA 3.0.4 to boost the quality of the 454-derived sequences. The remaining two gaps were closed by combining the overlapped parts of the high-quality contigs in Consed v. 20. The complete cpDNA sequence of *S. acus* is available in GenBank (JQ 088178).

2.3 cpDNA genome annotation

Gene content was determined by BLAST similarity searches (Altschul *et al.* 1997) against the non-redundant database of National Center for Biotechnology Information. ORFs were localized using Artemis 12.0; tRNA-coding sequences were identified by tRNAscan-SE 1.23 (Lowe and Eddy 1997). Small and large ribosomal RNA subunit genes were identified by RNAmmer software (Lagesen *et al.* 2007) and by comparing *S. acus* cpDNA with rRNA genes from chloroplast genomes of other diatoms. Among ORFs, protein-coding genes were manually annotated on the basis of similarity derived from BLAST-searches. Transfer-messenger RNAs and signal recognition particle genes were predicted by ARAGORN (Laslett & Canback 2004) and SRPscan (Regalia *et al.* 2002), respectively.

2.4 Search for nuclear-encoded genes with chloroplastic targeting

To look for the specific genes in nuclear genome of *S. acus*, we searched for the similar sequences in a bulk set of the assembled contigs using tblastn protocol. To confirm the plastid targeting, analysis for the presence of the bipartite signal peptide with HECTAR tool was performed (Gschloessl *et al.* 2008).

2.5 Phylogenetic analysis of atpF amino-acid sequences

Sequences, which are closely related to *S. acus* a*tpF* gene, were initially retrieved from GenBank on the basis of BLAST-hits against the non-redundant NCBI sequence database. Several distant sequences of red algae, cryptomonads, and haptophytes were used in the subsequent analysis as an outgroup. The sequences were aligned by MUSCLE (Edgar 2004) under default settings. The alignment was analyzed using the PhyML 3.0 package (Guindon & Gascuel 2003) with the LG amino acid substitution model (Le & Gascuel 2008), and the among-site rate variation approximated by a discrete gamma distribution with four rate categories (LG+Γ model). The initial tree was constructed using BioNJ. Subsequent tree topologies were searched by subtree pruning and regrafting (SPR). Bootstrap analysis was performed for 100 replicates. The consensus network was computed in SplitsTree program (Huson *et al.* 2006) with a threshold of 0.1 and edge weights representing split counts. A similar procedure was performed to reconstruct the phylogenetic network based on *rbcL* amino-acid sequences. Alignments of *atpF* and *rbcL* sequences are available on request.

3. Results and Discussion

The chloroplast genome of *S. acus* was sequenced and assembled using the whole-genome approach as described in section *2.2*. This cpDNA is compared to other chloroplast genomes of diatom origin to integrate the newly available data into the overall concept of the plastid genomics of diatoms. Chloroplast genome of *S. acus* has a size of 116 251 bp and is mapped as a circular molecule. It is characterized as a canonical quadripartite structure with two inverted repeats, IRa and IRb, containing ribosomal RNA gene loci, that separate small (SSC) and large (LSC) single-copy regions (Figure 1). Plastid genome of *S. acus* is somewhat smaller than those of other diatom algae sequenced to date, whereas other general features of the genome such as GC-content and coding capacity are very similar between cpDNAs of diatoms/dinotoms group.

Gene set of *S. acus* cpDNA consists of 160 genes (Table 1). There are three rRNAs composing ribosomal loci in IRs, 27 transfer RNAs, which are sufficient for messenger RNA translation inside the organelle, and 128 protein-coding genes. We have also found a transfer-messenger RNA gene *ssra* which is believed to play a role in trans-translational termination (Roche & Sauer 1999) and signal-recognition particle RNA *fss* which participates in the transmembrane transport of nuclear-encoded genes with plastid localization. The protein-coding gene set includes 44 ribosomal protein genes, 44 photosynthesis-associated genes, and 40 other proteins.

Gene complements of diatom-related cpDNAs share a core set of 150 common genes. This includes 3 rRNA genes, 27 transfer RNAs, and 120 protein-coding genes. However, there are differences in gene content which are outlined in Table 2. Specifically, the *S. acus* plastid genome contains 10 genes absent from at least one of diatom-related cpDNA (Table 1).

S. acus plastid protein-coding genes use the standard plastid/bacterial genetic code (code table 1). In *S. acus* cpDNA, there are five genes with alternative start codons: *rbcS*, *ccsA*, *rps8,* and *rpl23* use GTG, whereas ATT is used in *secY* instead of ATG. The most frequent stop-codon is TAA which is used in 117 genes. Ten genes end with TAG, two with TTA, and one with TGA. All diatom/dinotom plastid genomes are compact and have no intron sequences. Intergenic spacers do not exceed 120 bp.

Like in other diatom genomes, there is no intergenic spacer between *rpl14* and *rpl24* genes in *S. acus* cpDNA. Another similar feature is the presence of three identical cases of overlapping genes in *S. acus* cpDNA: *sufC-sufB* (1 bp), *rpl4-rpl23* (8 bp) and *psbD-psbC* (53 bp) (Oudot-Le Secq *et al.* 2009). However, genes *atpD* and *atpF* do not overlap in contrast with other plastid genomes of diatoms and dinotoms. This feature is a clear synapomorphy of *S. acus* cpDNA which is probably a result of either relaxed constraints or extensive selection forces acting to *atpF* gene in this diatom. To further investigate the reason of low similarity of *atpF* gene, we performed phylogenetic analysis as described in section *2.5. S. acus atpF* gene was shown to belong to the diatom/dinotom group of sequences as revealed by the phylogenetic network reconstruction (Figure 2A). The analysis failed to reveal the order of divergence within the diatom lineage. We performed phylogenetic network reconstruction based on *rbcL* gene amino-acid sequences from a similar set of taxons to see whether this polytomy could be explained by poor taxon sampling (Figure 2B). Like in previous case, diatom divergence is not resolved by the *rbcL*-network. However, according to *rbcL* data, *S. acus* falls into the same clade as other *Synedra*-specific sequences. Additional taxon sampling is still required for a solid understanding of *atpF* evolution in diatoms. Nevertheless, *S. acus atpF* is unlikely to have been acquired by horizontal gene transfer from some distant lineage, since the closest *atpF* sequences still belong to diatom/dinotom group.

As noted above, protein-coding gene sets in the plastid genomes of diatom origin share 120 common genes. It is well known that plastid genes tend to undergo a sequential process of transfer from chloroplast to nucleus, as it is

believed to unify regulation of expression and eventually increase integration of the molecular cell machinery during the orchestrated response to changing environmental conditions (Lommer *et al.* 2010). In this respect, at least some of the genes noticed in Table 2 are hypothesized to have been transferred to the nucleus in the course of evolution. Interestingly, we found a sign of such gene transfer in *S. acus.* Acyl-carrier protein gene *acpp,* which is known to be an important component in fatty-acid biosynthesis inside the plastid, is found only in cpDNAs of *O. sinensis* and *P. tricornutum*. The BLAST-search for *acpp*-related sequences in the bulk assembly of *S. acus* 454 shot-gun reads revealed a contig with similar gene. Further analysis showed *S. acus* nuclear *acpp* gene to contain a bipartite signal of plastid localization which is responsible for directing the transcribed polypeptide towards chloroplast stroma through four plasma membranes. Given the paradigm of divergence of diatoms (Sims *et al.* 2006) coupled with the origin of cpDNAs of dinotoms (Imanian *et al.* 2010), a transfer of the plastid *acpp* gene to the nuclear host genome could hardly be considered as a single event that occurred in an ancient diatom ancestor. It should rather be regarded as a series of independent gene transfers in respective lineages of diatoms.

The small RNA genes detected in *S. acus* cpDNA are present in all plastid genomes of diatom origin except for *O. sinensis* (Table 2). Most likely, the absence of *ssra* and *ffs* genes in cpDNA of *Odontella* is a result of recent evolutionary event which occurred after divergence of this lineage from other groups of diatoms under analysis.

4. Conclusion

During the last decade, several complete organellar and nuclear genomes of diatoms have been sequenced. This encouraged establishing of the first model diatoms, namely *T. pseudonana* and *P. tricornutum,* and supported the origin of functional genomics of diatom algae. In terms of evolutionary biology, the presence of genomic data allows us to apply a rich repertoire of phylogeny methods on a whole-genome scale, to identify rearrangement events, cases of horizontal gene transfer, to perform large-scale phylogenetic analysis using multiple genes, *etc*. However, additional data are required for deep and solid understanding of all aspects of biology of diatoms. Chloroplast DNA sequence of *S. acus* is a small step forward in deciphering a complex quest which is challenged by diatoms, an extremely diversified and distinctive group of unicellular algae.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., *et al*. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25 (17), 3389-3402. http://dx.doi.org/10.1093/nar/25.17.3389

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32 (5), 1792-1797. http://dx.doi.org/10.1093/nar/gkh340

Falkowski, P. G., Katz, M. E., Knoll, A. H., *et al*. (2004). The evolution of modern eukaryotic phytoplankton. *Science,* 305 (5682), 354-360. http://dx.doi.org/10.1126/science.1095964

Grachev, M. A., Vorobyova, S. S., Likhoshway Ye.V., *et al*. (1998). A High-Resolution diatom record of the palaeoclimates of East Siberia for the last 2.5 My from Lake Baikal. *Quat Sci Rev,* 17 (12), 1101-1106.

Gschloessl, B., Guermeur, Y. & Cock, J. M. (2008). HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*, 9, 393. http://dx.doi.org/10.1186/1471-2105-9-393

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol,* 52 (5), 696-704. http://dx.doi.org/10.1080/10635150390235520

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol,* 23 (2), 254-267. http://dx.doi.org/10.1093/molbev/msj030

Imanian, B., Pombert, J. F. & Keeling, P. J. (2010). The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS One*, 5 (5), e10711. http://dx.doi.org/10.1371/journal.pone.0010711

Kowallik, K. V., Stoebe, B., SchaVran, I., Kroth-Pancic & Freier, U. (1995). The chloroplast genome of a chlorophyll a + c- containing alga, *Odontella sinensis*. *Plant Mol Biol Rep,* 13, 336-342. http://dx.doi.org/10.1007/BF02669188

Lagesen, K., Hallin, P., Rødland, E. A., *et al*. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res,* 35 (9), 3100-3108. http://dx.doi.org/10.1093/nar/gkm160

Laslett, D. & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res,* 32 (1), 11-16. http://dx.doi.org/10.1093/nar/gkh152

Le, S.Q. Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol,* 25 (7), 1307-1320. http://dx.doi.org/10.1093/molbev/msn067

Lommer, M., Roy, A. S., Schilhabel, M., Schreiber, S., Rosenstiel, P. & LaRoche, J. (2010). Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics,* 11, 718. http://dx.doi.org/10.1186/1471-2164-11-718

Lowe, T. M. & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res,* 25 (5), 955-964. http://dx.doi.org/10.1093/nar/25.5.955

Medlin, L. K. (2009) Diatom (Bacillariophyta). In S.B. Hedges, & S. Kuman (Eds.). *The timetree of life,* (pp. 127-130). Oxford University Press.

Oudot-Le Secq, M. P., Grimwood, J., Shapiro, H., *et al*. (2007). Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. *Mol Genet Genomics*, 277 (4), 427-39. http://dx.doi.org/10.1007/s00438-006-0199-4

Ravin, N. V., Galachyants, Y. P., Mardanov, A. V., *et al*. (2010). Complete sequence of the mitochondrial genome of a diatom alga *Synedra acus* and comparative analysis of diatom mitochondrial genomes. *Curr Genet*, 56 (3), 215-23. http://dx.doi.org/10.1007/s00294-010-0293-3

Regalia, M., Rosenblad, M. A. & Samuelsson, T. (2002). Prediction of signal recognition particle RNA genes. *Nucleic Acids Res,* 30 (15), 3368-3377. http://dx.doi.org/10.1093/nar/gkf468

Roche, E. D. & Sauer, R. T. (1999). SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *EMBO J*, 18 (16), 4579-4589. http://dx.doi.org/10.1093/emboj/18.16.4579

Round, F. E., Crawford, R. M. & Mann, D. G. (1990) *The Diatoms: biology and morphology of the genera.* Cambridge University Press.

Shishlyannikov, S. M., Zakharova, Y.R., Volokitina, N.A., M *et al*. (2011). A procedure for establishing an axenic culture of the diatom *Synedra acus* subsp. radians (Kütz.) Skabibitsch. from Lake Baikal. *Limnol. Oceanogr.: Methods,* 9, 478-484.

Sims, P., Mann, D. & Medlin, L. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia,* 45 (4), 361-402. http://dx.doi.org/10.2216/05-22.1

Tanaka, T., Fukuda, Y., Yoshino, T., *et al*. (2011). High-throughput pyrosequencing of the chloroplast genome of a highly neutral-lipid-producing marine pennate diatom, *Fistulifera* sp. strain JPCC DA0580. *Photosynth Res,* 109 (1-3), 223-9. http://dx.doi.org/10.1007/s11120-011-9622-8

Table 1. Gene content of *S. acus* cpDNA

Non-protein-coding genes (32)	
Ribosomal RNA (3)	*rnl*[1], *rns*[1],*rrn5*[1]
Transfer RNA (27)	*trnA(ugc)*[1], *trnC(gca), trnD(guc), trnE(uuc), trnF(gaa), trnG(gcc), trnG(ucc), trnH(gug), trnI(cau), trnI(gau)*[1], *trnK(uuu), trnL(caa), trnL(uaa), trnM(cau)-i, trnM(cau)-e, trnN(guu), trnP(ugg)*[1], *trnQ(uug), trnR(acg), trnR(ccg), trnR(ucu), trnS(gcu), trnS(uga), trnT(ugu), trnV(uac), trnW(cca), trnY(gua)*
Others (2)	*ffs*[2], *ssra*[2]
Protein-coding genes (128)	
Ribosomal proteins (44)	
Small subunit (18)	*rps2-rps14, rps16-rps20*
Large subunit (26)	*rpl1-6, rpl11-14, rpl16, rpl18-24, rpl27, rpl29, rpl31-36*
Photosynthesis (44)	
ATP-synthase (8)	*atpA, atpB, atpD, atpE, atpF, atpG, atpH, atpI*
Photosystem I (10)	*psaA, psaB, psaC, psaD, psaE, psaF, psaI, psaJ, psaL, psaM*
Photosystem II (18)	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbN, psbT, psbV, psbX, psbY*[1,2], *psbZ*[2] *(ycf9), psb28*
Cytochrome complex (8)	*petA, petB, petD, petF, petG, petL (ycf7), petM*[2] *(ycf31), petN (ycf6)*
Other (40)	
Metabolism (5)	*chlI, rbcL, rbcS, thiG*[2], *thiS*[2] *(ycf40)*
Protein folding (5)	*clpC, dnaB, dnaK, ftsH (ycf25), groEL*
Assembly, membrane translocation (8)	*ccs1 (ycf44), ccsA, secA, secG*[2] *(ycf47), secY, sufB (ycf24), sufC (ycf16), tatC (ycf43)*
Transcription (6)	*cbbX, rbcR, rpoA, rpoB, rpoC1, rpoC2*
Translation (2)	*syfB*[2], *tufA*
Conserved ORFs (14)	*ycf3, ycf4, ycf12, ycf33, ycf35, ycf39, ycf41, ycf42*[2], *ycf45, ycf46, ycf66, ycf88, ycf89*[1,2], *ycf90*

Genes marked with «1» exist in chloroplast genome in two copies within the inverted repeats. The ten genes marked with «2» are not in the core set of 150 genes shared between cpDNAs of diatom origin.

Table 2. Gene content differences between diatom plastid genomes

Gene name	Gene synonym	Description	TP	TO	OS	SA	PT	Fsp	KF	DB
acpp	-	acyl carrier protein	N	N	Y	N	Y	N	N	N
ffs	-	signal-recognition particle RNA	Y	Y	N	Y	Y	Y	Y	Y
flrn	-	putative ffs-like RNA	N	Y	N	N	N	N	N	N
orf127	-	hypothetical protein	N	Y	N	N	N	N	N	N
petM	*ycf31*	cytochrome b6-f complex subunit VII	Y	Y	N	Y	Y	Y	Y	Y
psbI	-	photosystem II protein I	Y	Y	Y	Y	Y	N	Y	Y
psbZ	*ycf9*	photosystem II protein Z	Y	Y	N	Y	Y	Y	Y	Y
secG	*ycf47*	preprotein translocase SecG subunit	Y	Y	N	Y	Y	Y	Y	Y
serc1	-	putative serine recombinase	N	N	N	N	N	N	Y	N
serc2	-	putative serine recombinase	N	N	N	N	N	Y	Y	N
ssra	-	transfer-messenger RNA	Y	Y	N	Y	Y	Y	Y	Y
syfB	-	phenylalanyl-tRNA synthetase beta chain	N	N	N	Y	Y	Y	Y	Y
thiG	-	thiamin biosynthesis protein G	Y	Y	Y	Y	Y	N	Y	Y
thiS	*ycf40*	thiamin biosynthesis protein	Y	Y	Y	Y	Y	N	Y	Y
tsf	-	elongation factor Ts	N	N	N	N	Y	Y	N	N
tyrC	-	putative integrase/recombinase protein	N	N	N	N	N	N	Y	N
ycf42	-	conserved ORF, putative peroxiredoxin	Y	Y	Y	Y	Y	Y	N	N

There are 150 common genes in diatom plastid genomes. See Table 1 for exact list of shared genes. List of abbreviations: TP – *T. pseudonana*, TO – *T. oceanica*, OS – *O. sinensis*, SA – *S. acus*, PT – *P. tricornutum*, Fsp – *Fistulifera* sp., KF – *K. foliaceum*, DB – *D. baltica.*

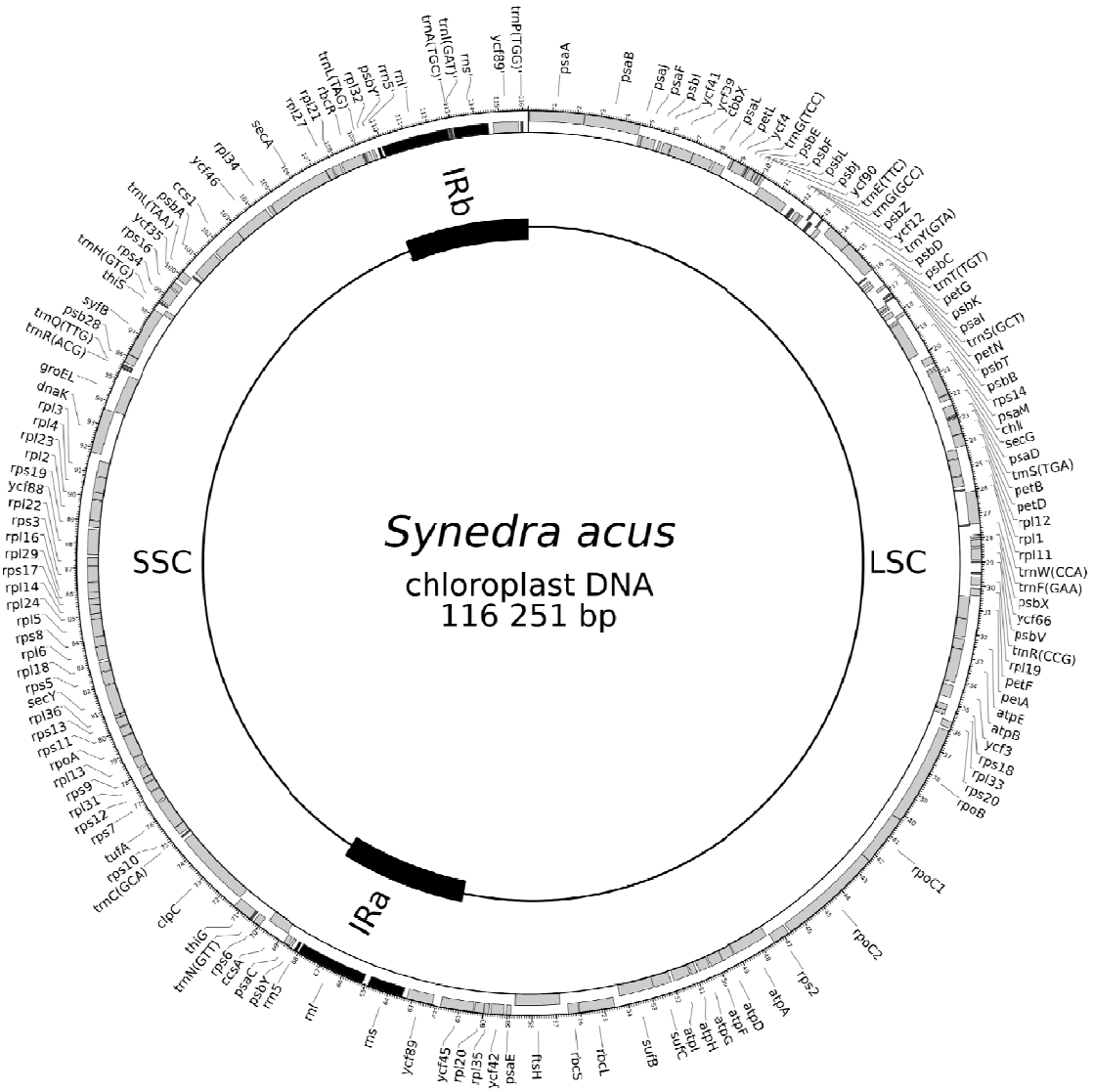

Figure 1. Genome map of *S. acus* cpDNA

Genes inside and outside the outermost circle diagram are transcribed in counter-clockwise and clockwise directions, respectively. Light gray blocks designate the protein-coding genes, black blocks – ribosomal RNAs, dark gray blocks – transfer and other RNA genes. The innermost circle schematically represents the elements of quadripartite structure where inverted repeats IRa and IRb separate large (LCS) and small (SSC) single-copy regions.

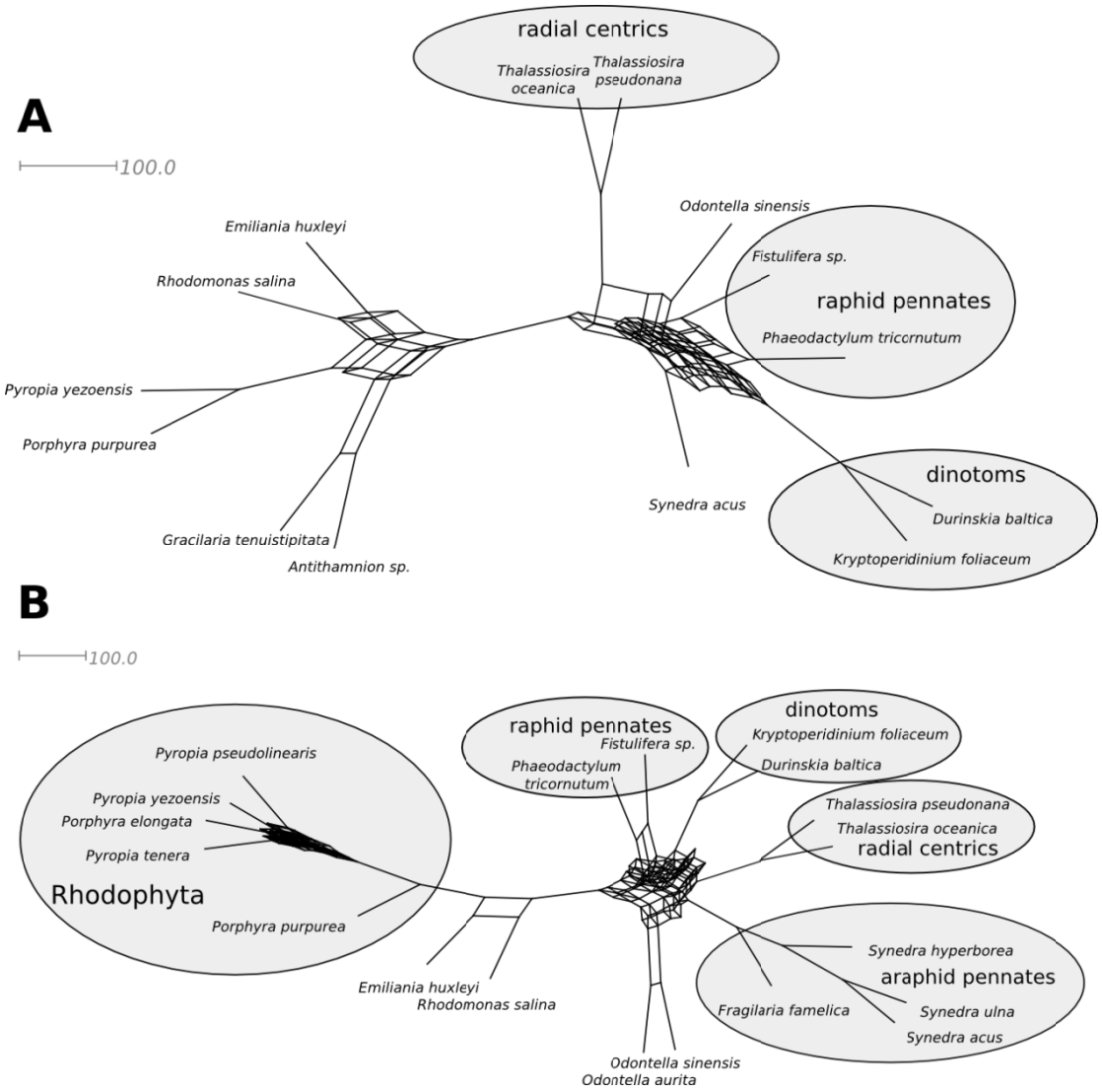

Figure 2. Phylogenetic networks based on chloroplast *atpF* (A) and *rbcL* (B) amino-acid sequences of diatoms

Phylogenetic network was reconstructed as described in section *2.5*. The length of edge is computed as number of initially obtained bootstrap trees which contain the split represented by this edge.