Photosynth Res
DOI 10.1007/s11120-011-9622-8

REGULAR PAPER

High-throughput pyrosequencing of the chloroplast genome of a highly neutral-lipid-producing marine pennate diatom, *Fistulifera* sp. strain JPCC DA0580

Tsuyoshi Tanaka · Yorikane Fukuda · Tomoko Yoshino · Yoshiaki Maeda · Masaki Muto · Mitsufumi Matsumoto · Shigeki Mayama · Tadashi Matsunaga

Received: 25 August 2010 / Accepted: 5 January 2011
© Springer Science+Business Media B.V. 2011

Abstract The chloroplast genome of the highly neutral-lipid-producing marine pennate diatom *Fistulifera* sp. strain JPCC DA0580 was fully sequenced using high-throughput pyrosequencing. The general features and gene content were compared with three other complete diatom chloroplast genomes. The chloroplast genome is 134,918 bp with an inverted repeat of 13,330 bp and is slightly larger than the other diatom chloroplast genomes due to several low gene-density regions lacking similarity to the other diatom chloroplast genomes. Protein-coding genes were nearly identical to those from *Phaeodactylum tricornutum*. On the other hand, we found unique sequence variations in genes of photosystem II which differ from the consensus in other diatom chloroplasts. Furthermore, five functional unknown ORFs and a putative serine recombinase gene, *serC2*, are located in the low gene-density regions. *SerC2* was also identified in the plasmids of another pennate diatom, *Cylindrotheca fusiformis*, and in the plastid genome of the diatom endosymbiont of *Kryptoperidinium foliaceum*. Exogenous plasmids might have been incorporated into the chloroplast genome of *Fistulifera* sp. by lateral gene transfer. Chloroplast genome sequencing analysis of this novel diatom provides many important insights into diatom evolution.

Keywords Diatom · Chloroplast genome · Next-generation DNA sequencing · Lateral gene transfer · Endosymbiosis

Electronic supplementary material The online version of this article (doi:10.1007/s11120-011-9622-8) contains supplementary material, which is available to authorized users.

T. Tanaka (✉) · T. Yoshino · Y. Maeda · T. Matsunaga
Department of Biotechnology, Faculty of Engineering, Tokyo University of Agriculture & Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan
e-mail: tsuyo@cc.tuat.ac.jp

Y. Fukuda · M. Muto
The Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture & Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan

M. Matsumoto
Biotechnology Laboratory, Electric Power Development Co., Ltd, 1, Yanagisaki-machi, Wakamatsu-ku, Kitakyusyu 808-0111, Japan

S. Mayama
Department of Biology, Tokyo Gakugei University, 4-1-1, Nukuikita-machi, Koganei, Tokyo 184-8501, Japan

T. Tanaka
JST, CREST, Sanbancho 5, Chiyoda-ku, Tokyo 102-0075, Japan

Introduction

Diatoms are a widespread group of eukaryotic microalgae found in freshwater and marine environments, and are the main contributors to global carbon cycling. Furthermore, diatoms play a key role in controlling silicon cycling in the ocean (De La Rocha et al. 1998). Therefore, many studies have examined the ecological and geological significances of diatoms in natural environments. Recent attention has been focused on biofuel production by diatoms via photosynthetic conversion of carbon dioxide as a source of neutral lipids (Courchesne et al. 2009; Rodolfi et al. 2009; Schenk et al. 2008). Diatoms have relatively high growth rates (approximately 1.0 day^{-1}) and do not compete with food crops.

Diatoms are also important for understanding the evolutionary history of the eukaryotic cell. The endosymbiotic

Published online: 03 February 2011

hypothesis suggests that plastids are primarily derived from endosymbiotic cyanobacteria. Chloroplasts, the light-harvesting plastids in diatoms, were derived from a red alga by secondary endosymbiogenesis, rather than directly from prokaryotes as occurred in plants. To date, three diatom chloroplast genomes have been fully sequenced, including the centric diatoms *Odontella sinensis* and *Thalassiosira pseudonana*, and the pennate diatom *Phaeodactylum tricornutum* (Kowallik et al. 1995; Oudot-Le Secq et al. 2007). Furthermore, analysis of the complete plastid genome of the dinoflagellate *Kryptoperidinium foliaceum*, the plastid likely derived from the pennate diatom by a tertiary endosymbiosis, revealed that gene content and genome organization were similar to that of the pennate diatom *P. tricornutum*. Two exogenous plasmids originated from a pennate diatom, and *Cylindrotheca fusiformis* might have been incorporated by a lateral gene transfer (Imanian et al. 2010).

Genome sequence analysis of more diverse diatom species is a promising approach for further understanding the evolutionary history of eukaryotic cells. In recent years, several technologies for large-scale DNA sequencing, designated as next-generation sequencing technologies, have been developed and provide faster and more cost-effective sequencing throughput (Hert et al. 2008; Ansorge 2009; Schuster 2008). Pyrosequencing, developed by Roche 454 Life Sciences, has been applied to de novo sequencing of microbial and other small genomes (Hongoh et al. 2008; Argueso et al. 2009).

Here, we describe chloroplast genome analysis of a newly isolated marine pennate diatom, *Fistulifera* sp. strain JPCC DA0580, by next-generation sequencing technology using the Genome Sequencer FLX System. *Fistulifera* sp. (formerly *Navicula* sp.) strain JPCC DA0580 was identified as the highest neutral lipid-producer among 1393 strains by screening a marine microalgal culture collection (Matsumoto et al. 2010). Furthermore, comparative analysis of genome among five species (*Fistulifera* sp. strain JPCC DA0580, *T. pseudonana*, *P. tricornutum*, *O. sinensis*, and *K. foliaceum*) was performed. This molecular sequence analysis of a novel diatom provides important insights into diatom evolution.

Materials and methods

DNA extractions

Marine diatom strain JPCC DA0580 was used in this study. Strain JPCC DA0580 was isolated from the junction of the Sumiyo-River and Yakugachi River in Kagoshima, Japan (Matsumoto et al. 2010). On the basis of the phenotypic and genotypic comparison of this isolate with the other strains, JPCC DA0580 was identified as a strain closely related morphologically to *Fistulifera saprophila*. Strain JPCC DA0580 was cultured in f/2 medium (Guillard 1975) (75 mg $NaNO_3$, 6 mg $Na_2HPO_4{\cdot}2H_2O$, 0.5 μg vitamin B12, 0.5 μg biotin, 100 μg Thiamine HCl, 10 mg $Na_2SiO_3{\cdot}9H_2O$, 4.4 mg Na_2-EDTA, 3.16 mg $FeCl_3{\cdot}6H_2O$, 12 μg $CoSO_4{\cdot}5H_2O$, 21 μg $ZnSO_4{\cdot}7H_2O$, 0.18 mg $MnCl_2{\cdot}4H_2O$, 70 μg $CuSO_4{\cdot}5H_2O$, and 7 μg $Na_2MoO_4{\cdot}2H_2O$) dissolved in 1 l of artificial seawater. Cultures were bubbled with sterile air at 20°C under 140 μmol/m^2/s illumination for 14 days.

Cells at late logarithmic growth phase (4.5 l of culture) were collected by centrifugation at 10,000 × *g* for 10 min at 4°C. Cell pellets were frozen in liquid nitrogen, resuspended in 15 ml of lysis buffer (50 mM Tris–HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS, and 10 mM DTT), and incubated at 50°C for 30 min (Bowler et al. 2008). Genomic DNA was stained by Hoechst33258 dye (Dojindo, Japan) and purified by cesium chloride centrifugation (Armbrust et al. 2004).

Genome sequencing

The chloroplast genome of strain JPCC DA0580 was sequenced using a GS FLX Titanium DNA pyrosequencer (Roche 454 Life Sciences, Branford, CT, USA). The library for the GS FLX Titanium was constructed using the GS FLX Titanium General Library Preparation Kit. The library was amplified onto DNA capture beads by emulsion PCR according to the manufacturer's instructions. The collected beads were quantified, and the genome was sequenced using a GS Titanium Sequencing Kit XLR70 and Genome Sequencer FLX System.

The nucleotide sequences were assembled using GS De Novo Assembler version 2.3. The chloroplast genome sequence was compared with the reference sequence from the complete chloroplast genome of *Phaeodactylum tricornutum* (Oudot-Le Secq et al. 2007). Scaffolds containing the chloroplast genome were identified by their similarities to other chloroplast genomes. The remaining gaps were closed by primer walking of gap-spanning PCR products that were identified using linking information from forward and reverse reads with a BigDye Terminator v3.1 Cycle sequencing kit.

Genome annotation and analysis

The genome was examined for open reading frames (ORFs) using Artemis software (Rutherford et al. 2000). ORFs were annotated using BLAST 2.2.23 (Camacho et al. 2009) against the NCBI nr and nt databases. tRNAScan-SE v1.23 (Lowe and Eddy 1997) was used to identify transfer RNAs and BLASTN for identification of ribosomal RNAs.

The regions of the inverted repeat were defined using BLASTN against the sequence of *Fistulifera* sp. Physical maps were generated using GenomeVx (Conant and Wolfe 2008) and further edited manually.

Results and discussion

General features of the chloroplast genome of *Fistulifera* sp. strain JPCC DA0580

The chloroplast genome of *Fistulifera* sp. strain JPCC DA0580 (DDBJ Accession: AP011960) was fully sequenced by high-throughput pyrosequencing using GS FLX Titanium. A total of 273,968 sequences were generated covering 114.5 Mb with an average read length of 418 bases. The sequences were assembled into several contigs using GS De Novo Assembler version 2.3 software to cover the entire chloroplast genome. Two major gaps of 10 and 2.7 kbp are located in inverted repeats (IRs). Gaps within IRs were also observed in the sequencing results of the mungbean (*Vigna radiata*) chloroplast genome by high-throughput pyrosequencing (Tangphatsornruang et al. 2010). The remaining gaps were filled through PCR and Sanger sequencing. The chloroplast genome of *Fistulifera* sp. strain JPCC DA0580 was 134, 918 bp, containing a large single copy (LSC) region of 62,994 bp and a small single copy (SSC) region of 45,264 bp, divided by two IRs of 13,330 bp (Fig. 1).

The general features of the chloroplast genomes from four diatoms, including strain JPCC DA0580 and one diatom endosymbiont of dinoflagellate are summarized in Table 1. A previous report examined common features of chloroplast genomes from five chromists, including three diatoms (*Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, and *Odontella sinensis*), one haptophyte

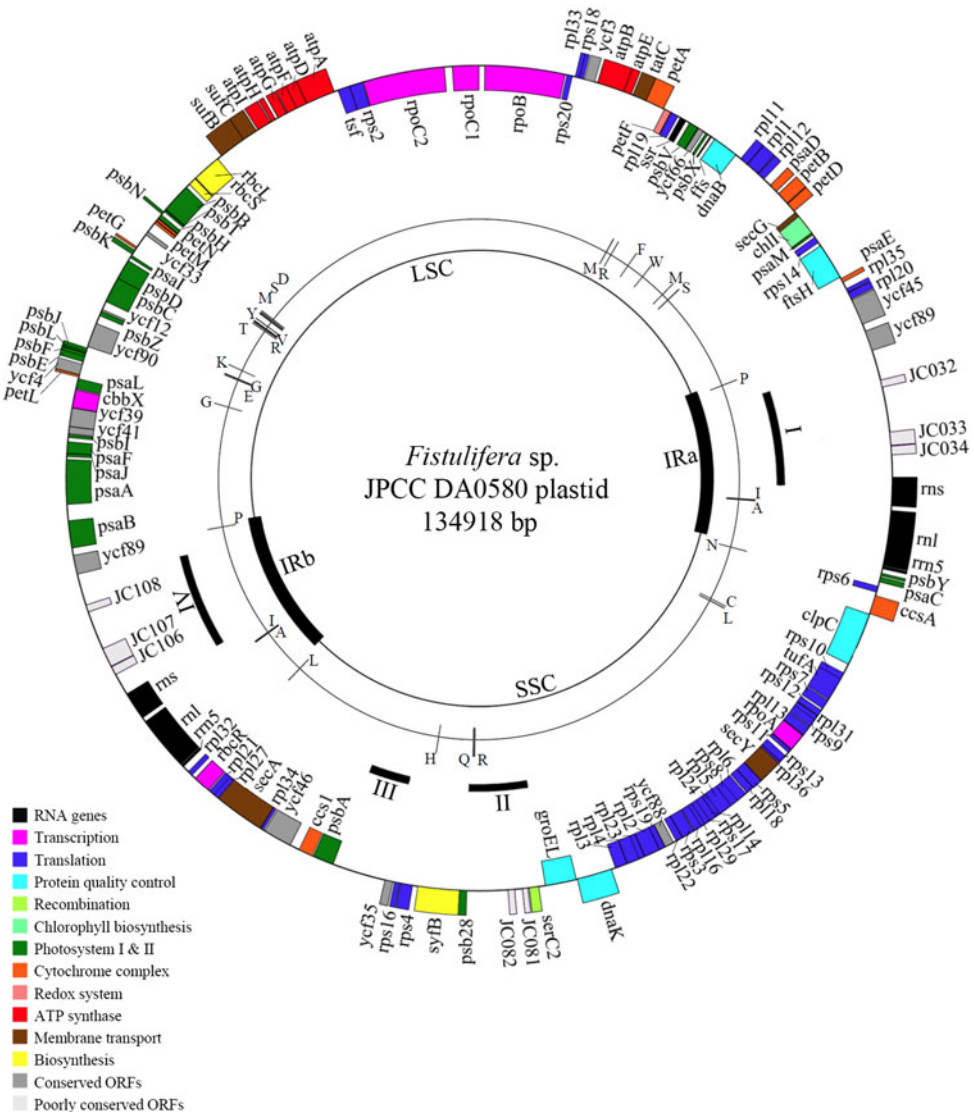

Fig. 1 A chloroplast genome map of *Fistulifera* sp. strain JPCC DA0580. Annotated genes are colored according to the functional categories. *First circle* gene organization, *second circle* tRNA position, *third circle*; inverted repeat regions (IR), small single copy (SSC) and large single copy (LSC). *Roman numerals* (I–IV) mark the locations of four distinct regions in the chloroplast genome of strain JPCC DA0580

Table 1 Chloroplast genome summary

	Diatom				Dinoflagellate
	Fistulifera sp. JPCC DA0580	*Phaeodactylum tricornutum*[a]	*Thalassiosira pseudonana*[a]	*Odontella sinensis*[a]	*Kryptoperidinium foliaceum*[b]
Size (bp)					
Total (% GC)	134918 (32.2)	117369 (32.6)	128814 (30.7)	119704 (31.8)	140426 (32.4)
Inverted repeat	13330	6912	18337	7725	6017
Small single-copy region	45264	39871	26889	38908	56521
Large single-copy region	62994	63674	65250	65346	71871
Average size of intergenic spacer	179.5	88.4	108.2	115.7	246.7
Gene content[c]					
Total	164	162	159	160	160
Protein-coding genes	132	130	127	128	128
rRNA genes	3	3	3	3	3
tRNA genes	27	27	27	27	27
Other RNAs	2	2	2	2	2
Introns	0	0	0	0	0
Overlapping genes	4	4	4	4	4
Coding sequence (%)	77.5	87.5	85.2	84.1	71.9

[a] Data taken from published information (Oudot-Le Secq et al. 2007)

[b] Data taken from published information (Imanian et al. 2010)

[c] Excepting duplicated genes

(*Emiliania huxleyi*), and one cryptophyte (*Guillardia theta*) (Oudot-Le Secq et al. 2007). The features included compact size, complete lack of introns, four identical overlapping genes, and small intergenic spacers (88–116 bp) among three diatoms. As expected, the chloroplast genome from strain JPCC DA0580 was also compact, completely lacked introns, and contained four overlapping genes, including *sufC–sufB* overlapped by 1 nt, *atpD–atpF* by 4 nt, *rpl4–rpl23* by 8 nt, and *psbD–psbC* by 53 nt. On the other hand, the chloroplast genome of strain JPCC DA0580 contained much larger intergenic spacers with an average length of 179.5 bp, which was the largest of this sequence among the four diatoms compared in this study. Low gene-density regions (I, II, and IV in Fig. 1), which included unidentified genes, and a long intergenic region were identified. These regions showed no similarity to other chloroplast genomes in diatoms. This specific feature was also confirmed in the chloroplast genome of the diatom endosymbiont of dinoflagellate *K. foliaceum* (Imanian et al. 2010). The average intergenic spacing in *K. foliaceum* is 246.7 bp.

Gene content in the chloroplast genome

Table 2 listed gene content in the chloroplast genome of the *Fistulifera* sp. strain JPCC DA0580. Protein-coding genes were found to be nearly identical to those of *P. tricornutum,* with several exceptions, such as *psbE* and *psbL* genes. These two genes showed the highest similarity with those of *O. sinensis*. The most common features among the four compared diatoms were three rRNA subunits (*rns*, *rnl*, and *rrn5*) in the IRs and 27 tRNAs (Oudot-Le Secq et al. 2007).

Five ORFs lacked significant similarity to any entry in the public domain sequence databanks. These putative genes were named JC032, -033, -034, -081, and -082. Three genes, JC032, -033, and -034, were detected in the IRs and duplicated. The others are located in the low gene density region (I, II, IV in Fig. 1). In addition, the chloroplast genome encoded a putative serine recombinase gene, *serC2*.

Comparison of sequence identity

The percentages of sequence identity between genes of JPCC DA0580 and other diatoms, raphidophyte, pelagophyte, or diatom endosymbiont of dinoflagellate in chloroplast genomes are listed in Supplementary Tables S-1 and S-2. These comparisons suggest that the gene components of diatom species are almost identical. The closest homologous species of strain JPCC DA0580 is *P. tricornutum*, and a highly homologous genome was also seen with the symbiotic chloroplast genome of dinoflagellate. On the other hand, the *tsf* (translational elongation factor) gene was retained in only two diatom's chloroplasts, strain

Table 2 Functional classification of Chloroplast-encoded genes in *Fistulifera* sp. strain JPCC DA0580

Genetic system	
Transcription	*cbbX, rbcR, rpoA, B, C1, C2*
Translation	*tufA, tsf*[a]
Ribosomal proteins	
Small subunit	*rps2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20*
Large subunit	*rpl1, 2, 3, 4, 5, 6, 11, 12, 13, 14, 16, 18, 19, 20, 21, 22, 23, 24, 27, 29, 31, 32, 33, 34, 35, 36*
Protein quality control	*clpC, dnaK, dnaB, groEL, ftsH*
Recombination	*serC2*
Photosynthetic apparatus	
Chlorophyll biosynthesis	*chlI*
Photosystem I	*psaA, B, C, D, E, F, I, J, L, M*
Photosystem II	*psbA, B, C, D, E*[b]*, F, H, I, J, K, L*[b]*, N, T, V, X, Y, Z, psb28*
Cytochrome complex	*ccs1, ccsA, petA, B, D, G, L, M, N*
Redox system	*petF*
ATP synthase	*atpA, B, D, E, F, G, H, I*
Membrane transport	*secA, G, Y, sufB, C, tatC*
Biosynthesis	*syfB*[a]*, rbcL, S*
Unknown function	
Conserved ORFs	*ycf3, 4, ycf12, 33, 35, 39, 41, 45, 46, 66, 88, 89*[c]*, 90*
Poorly conserved ORFs[d]	JC032[c], 033[c], 034[c], 081, 082
RNA genes	
Ribosomal	*rns, rnl, rrn5*
Transfer	*trnA(ugc)*[c]*, trnC(gca), trnD(guc), trnE(uuc), trnF(gaa), trnG(gcc), trnG(ucc), trnH(gug), trnI(gau)*[c]*, trnK(uuu), trnL(uaa), trnL(uag), trnM(cau), trnM(cau), trnM(cau), trnN(guu), trnP(ugg)*[c]*, trnQ(uug), trnR(acg), trnR(ccg), trnR(ucu), trnS(gcu), trnS(uga), trnT(ugu), trnV(uac), trnW(cca), trnY(gua)*
Others	*ffs, ssr*

Underline found in *Fistulifera* sp. JPCC DA0580 only among diatoms

[a] *Fistulifera* sp. JPCC DA0580 and *P. tricornutum* only among diatoms

[b] The highest similarity to *O. sinensis*

[c] Duplicated gene

[d] Gene ID was indicated

JPCC DA0580 and *P. tricornutum* (Supplementary Table S-1). The *tsf* gene is found on the chloroplast genomes of *Guillardia theta* and *Porphyra purpurea*. This result suggested that the loss in *T. pseudonana* and *O. sinensis* may be relatively recent. The comparison of amino acid sequences of the fundamental components in photosystem I and II showed high-homology (average 89.3% identities) of each gene component among chloroplasts derived from diatoms (Supplementary Table S-2). On the other hand, compared with the raphidophyte, pelagophyte, the average homology was 77.2%, which suggests that there is a difference in the evolution between the diatoms and the other chromalveolata species. Furthermore, strain JPCC DA0580 possesses a unique sequence difference relative to the consensus in other chloroplasts of diatoms (Supplementary Table S-3). The polymorphism appears focused on the *psbL* gene, where amino acid 19 is Phe in three diatoms except JPCC DA0580, in which it is Tyr. A previous report on the *psbL* gene in *Synechocystis* sp. PCC 6803 suggested that the amino acid substitutions in this position influenced photoautotrophic doubling time (Luo and Eaton-Rye 2008). These polymorphisms might be related to unique characteristics of high growth rate and high-accumulating triglyceride in *Fistulifera* sp. JPCC DA0580.

Comparison of gene content and order in a region surrounding *serC2*

Gene content and order in the 4.5-kb region surrounding *serC2*, including a low gene density region (region II in Fig. 1) and a non-coding region (region III in Fig. 1), were compared with three other diatoms and the dinoflagellate, *K. foliaceum* (Fig. 2a). In the 4.5-kb region, the gene order between *trnR* and *ycf35* was conserved among

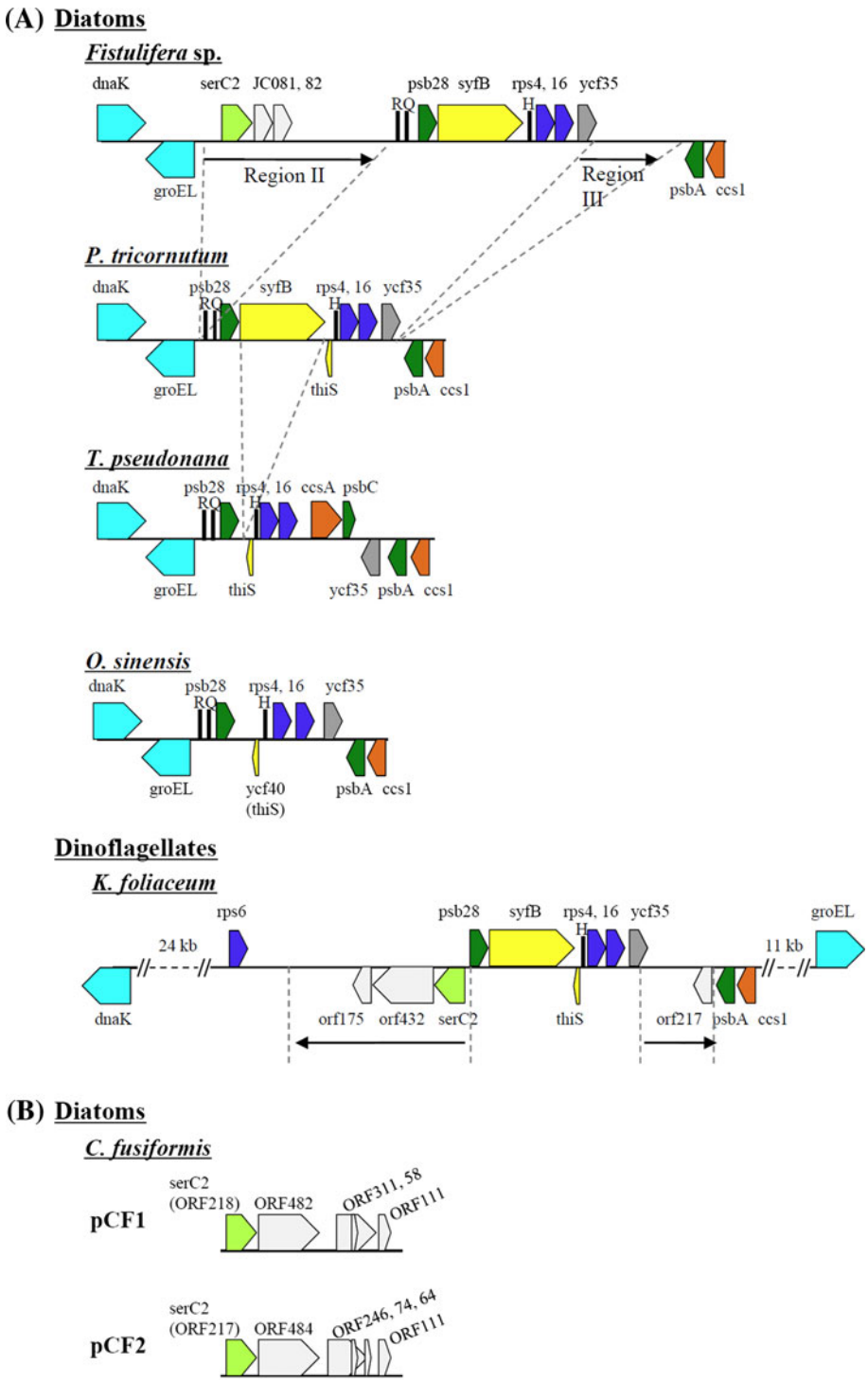

Fig. 2 Gene order surrounding the *serC2* gene. Chloroplast genomes of four diatoms and one diatom endosymbiont of *K. foliaceum* (**a**), and two plasmids of diatom *C. fusiformis* (**b**)

all four diatoms, although some deletions or insertions were observed. Regions II and III were located upstream of *trnR* and downstream of *ycf35*, respectively. Interestingly, similar tendencies were observed in *K. foliaceum*; i.e., low gene density regions were also located upstream of *psb28*, and downstream of *ycf35*, respectively. Furthermore, the intergenic regions showed similarity between strain JPCC DA0580 and *K. foliaceum*. A previous report suggested that the non-coding regions in *K. foliaceum* showed strong similarity to the pCf1/pCf2 plasmid sequences of the marine diatom *C. fusiformis*, and could be incorporated into the chloroplast genome of *K. foliaceum* by a lateral gene transfer event (Imanian et al. 2010). It is therefore possible that a similar event could occur in the chloroplast genome of *Fistulifera* sp. strain JPCC DA0580 chloroplast genome. Region II

shares similarity to the pCf1 and pCf2 plasmids of the diatom *C. fusiformis* (Hildebrand et al. 1992), and includes a putative serine recombinase gene, *serC2*. The *serC2* gene of *Fistulifera* sp. strain JPCC DA0580 shares 45.2, 39.4, and 42.9% aa identity with *K. foliaceum serC2*, ORF 218 from pCf1, and ORF 217 from pCF2, respectively. This is the first report that the trace of plasmid transfer was found in the diatom chloroplast genome, and these investigations will probably bring new insights regarding the diatom ancestral state.

Conclusions

In this study, we sequenced the chloroplast genome of a novel marine pennate diatom, *Fistulifera* sp. strain JPCC DA0580, using high-throughput pyrosequencing, and presented new information on diatom chloroplast genome architecture. The general features and gene content are broadly similar to those of three other fully sequenced diatom chloroplast genomes. However, there are some unique sequence variations in genes of photosystem II that differ from the consensus in other diatom chloroplast genomes. Furthermore, several identical regions that are low gene-density and show no similarity to the other diatom chloroplast genomes were identified in the strain JPCC DA0580 chloroplast genome. One identical region retained a putative serine recombinase also identified in the dinoflagellate *K. foliaceum* chloroplast genome and in two plasmids from another pinnate diatom, *C. fusiformis*. These results suggest that DNA sequences could be incorporated by lateral gene transfer of exogenous DNA. Currently, only four chloroplast genomes, including strain JPCC DA0580, are available. However, the database will continually grow due to next-generation DNA sequencing, and these advances will contribute novel insights to diatom evolution.

Acknowledgments This work was supported by JST, CREST.

References

Ansorge WJ (2009) Next-generation DNA sequencing techniques. New Biotechnol 25(4):195–203

Argueso JL et al (2009) Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. Genome Res 19(12):2258–2270

Armbrust EV et al (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306(5693):79–86

Bowler C et al (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456(7219): 239–244

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) Blast+: architecture and applications. BMC Bioinformatics 10:421

Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24(6): 861–862

Courchesne NMD, Parisien A, Wang B, Lan CQ (2009) Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. J Biotechnol 141(1–2): 31–41

De La Rocha CL, Brzezinski MA, DeNiro MJ, Shemesh A (1998) Silicon-isotope composition of diatoms as an indicator of past oceanic change. Nature 395(6703):680–683

Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates. In: Smith WL, Chanley MH (eds) Culture of marine invertebrate animals. Plenum Press, New York, pp 26–60

Hert DG, Fredlake CP, Barron AE (2008) Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. Electrophoresis 29(23):4618–4626

Hildebrand M, Hasegawa P, Ord RW, Thorpe VS, Glass CA, Volcani BE (1992) Nucleotide sequence of diatom plasmids: Identification of open reading frames with similarity to site-specific recombinases. Plant Mol Biol 19(5):759–770

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M (2008) Complete genome of the uncultured termite group 1 bacteria in a single host protist cell. Proc Natl Acad Sci USA 105(14):5555–5560

Imanian B, Pombert JF, Keeling PJ (2010) The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. PLoS One 5(5):e10711

Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic P, Freier U (1995) The chloroplast genome of a chlorophyll a + c-containing alga, Odontella sinensis. Plant Mol Biol Rep 13(4):336–342

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955–964

Luo H, Eaton-Rye JJ (2008) Directed mutagenesis of the transmembrane domain of the PsbL subunit of photosystem II in *Synechocystis* sp. PCC 6803. Photosynth Res 98(1–3): 337–347

Matsumoto M, Sugiyama H, Maeda Y, Sato R, Tanaka T, Matsunaga T (2010) Marine diatom, *Navicula* sp. strain JPCC DA0580 and marine green alga, *Chlorella* sp. strain NKG400014 as potential sources for biodiesel production. Appl Biochem Biotechnol 161:483–490

Oudot-Le Secq MP, Grimwood J, Shapiro H, Armbrust EV, Bowler C, Green BR (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):427–439

Rodolfi L, Zittelli GC, Bassi N, Padovani G, Biondi N, Bonini G, Tredici MR (2009) Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. Biotechnol Bioeng 102(1):100–112

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16(10):944–945

Schenk PM, Thomas-Hall SR, Stephens E, Marx UC, Mussgnug JH, Posten C, Kruse O, Hankamer B (2008) Second generation biofuels: high-efficiency microalgae for biodiesel production. BioEnergy Res 1(1):20–43

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods 5(1):16–18

Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N, Tragoonrung S (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17(1):11–22