Marine Genomics 10 (2013) 1–9

Contents lists available at SciVerse ScienceDirect

Marine Genomics

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

Chloroplast genome of one brown seaweed, *Saccharina japonica* (Laminariales, Phaeophyta): Its structural features and phylogenetic analyses with other photosynthetic plastids

Xiuliang Wang, Zhanru Shao, Wandong Fu, Jianting Yao, Qiuping Hu [1], Delin Duan *

Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form 19 December 2012
Accepted 20 December 2012

Keywords:
Photosynthetic heterokont
Saccharina japonica
Chloroplast
Phylogenomic
Molecular phylogenetics

ABSTRACT

The chloroplast genome sequence of one brown seaweed, *Saccharina japonica*, was fully determined. It is characterized by 130,584 base pairs (bp) with a large and a small single-copy region (LSC and SSC), separated by two copies of inverted repeats (IR1 and IR2). The inverted repeat is 5015 bp long, and the sizes of SSC and LSC are 43,174 bp and 77,378 bp, respectively. The chloroplast genome of *S. japonica* consists of 139 protein-coding genes, 29 tRNA genes, and 3 ribosomal RNA genes. One intron was found in one tRNA-Leu gene in the chloroplast genome of *S. japonica*. Four types of overlapping genes were identified, *ycf24* overlapped with *ycf16* by 4 nucleotides (nt), *ftrB* overlapped with *ycf12* by 6 nt, *rpl4* and *rpl23* overlapped by 8 nt, finally, *psbC* overlapped with *psbD* by 53 nt. With two sets of concatenated plastid protein data, 40-protein dataset and 26-protein dataset, the chloroplast phylogenetic relationship among *S. japonica* and the other photosynthetic species was evaluated. We found that the chloroplast genomes of haptophyte, cryptophyte and heterokont were not resolved into one cluster by the 40-protein dataset with amino acid composition bias, although it was recovered with strong support by the 26-protein dataset.

1. Introduction

The photosynthetic heterokont, haptophyte and cryptophyte were previously postulated to have arisen from a single secondary endosymbiosis in which a non-photosynthetic eukaryote acquired a plastid by engulfing a red algal (Cavalier-Smith, 1999).

Many molecular phylogenetic studies have been conducted to resolve the phylogenetic relationships of these three groups, and the results were in a flux. Recently the photosynthetic heterokont was classified into one new supergroup Sar, however, haptophyte and cryptophyte were listed as incertae sedis in the eukaryotes, based on nuclear genes phylogenomic analyses to taxon broadly sampled (Adl et al., 2012). In addition to the nuclear DNA sequences, ten mitochondrial genes were used to examine the phylogenetic relationships of the 29 taxa. The results showed that heterokont algae formed strong sister-clusters that separated from the cryptophyte and haptophyte (Oudot-Le Secq et al., 2006). Moreover, the chloroplast genome could give important clues to the phylogenetic relationships among the three groups (Yoon et al., 2002). For instance, an analysis of 62 plastid associated genes of 15 taxa showed that the plastids from heterokont, dinoflagellate, haptophyte and cryptophyte are monophyletic (Sanchez-Puerta et al., 2007). However, based on the plastid protein dataset without composition bias, such phylogenetic relationship was not identified for heterokont, haptophyte and cryptophyte (Khan et al., 2007).

Previous studies showed that the phylogenetic relationships among the three groups are hard to resolve, because the methodologies and analysis results varied significantly using the datasets from nuclear, mitochondria or chloroplast (Green, 2011). Therefore, in order to resolve the deepest branch order in the topology tree for photosynthetic heterokont, haptophyte and cryptophyte, broad sampling and new dataset are needed for phylogenetic analysis (Le Corguillé et al., 2009).

Here, we reported the sequence and structural analyses of the chloroplast genome of *Saccharina japonica* from Laminariales, which is one of the economically important species in large scale brown seaweed aquaculture in East Asian countries (Tseng, 2001). In comparison with other two available brown seaweed plastid genomes of *Fucus vesiculosus* and *Ectocarpus siliculosus* (Le Corguillé et al., 2009), our data showed that there were higher cpDNA structural similarities between *S. japonica* and *F. vesiculosus*. With 40- and 26-plastid protein sequence datasets, we investigated the phylogenetic relationships among the plastids of photosynthetic heterokont, haptophyte and cryptophyte, and demonstrated that the chloroplast genomes of haptophyte, cryptophyte and heterokont were not resolved into one cluster by the 40-protein dataset, although it was recovered with strong support by the 26-protein dataset.

* Corresponding author. Tel./fax: +86 532 82898556.
E-mail address: dlduan@qdio.ac.cn (D. Duan).
[1] Present address: Shanghai Majorbio Bio-pharm Technology C., Ltd., Shanghai 201203, China.

1874-7787/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.margen.2012.12.002

Table 1
Datasets used in the phylogenetic studies.

Dataset	Number of amino acids	Gene names	Taxa that failed the amino acid composition homogeneity test by TreePuzzle
40 proteins for 23 species	7140	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *petA*, *petB*, *petD*, *petG*, *psaA*, *psaB*, *psaC*, *psaJ*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbN*, *psbT*, *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rps2*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps19*, *ycf3*, *ycf4*	*Cyanidium caldarium*, *Cyanidioschyzon merolae*, *Fucus vesiculosus*, *Nephroselmis olivacea*, *Vaucheria litorea*, *Prochlorococcus marinus*, *Synechocystis* sp.
26 proteins for 21 species	5746	*atpA*, *atpB*, *atpH*, *petB*, *petD*, *petG*, *psaA*, *psaB*, *psaC*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbL*, *psbN*, *psbI*, *rpl14*, *rpl2*, *rpl20*, *rps11*, *rps12*, *rps14*, *rps19*, *ycf3*	None

2. Materials and methods

2.1. Algal material and chloroplast DNA purification

S. japonica sporophytes raised from gametophyte cultures were maintained at the Key Laboratory of Experimental Marine Biology, Institute of Chinese Academy of Sciences (Wang et al., 2004). Gametophyte culture and sporophyte cultivation were performed according to the previously described method (Li et al., 2007). The chloroplast DNA was purified from fresh juvenile sporophytes according to the reported method (Fu et al., 2008).

2.2. Chloroplast genome sequencing, assembling and annotation

The shotgun libraries and PCR strategies were used to generate the complete chloroplast genome sequence of *S. japonica*. The shotgun libraries were constructed with fragments (1.6–4 kb) of nebulized, purified plastid DNA cloned into pUC19 (Fermentas). Sequencing reactions were performed and analyzed on ABI3730xl automated sequencers.

In total, 2216 fragments were sequenced with average fragment length of 969 base pairs. The chloroplast genome sequences were assembled with the Phred–Phrap–Consed package (Gordon et al., 1998). The initial assembly of random clone sequences produced four large contigs. Contigs were united by reverse sequencing of selected clones. The missing sequences in the gaps were generated by multiple PCR reactions with reference to the reported *E. siliculosus* and *F. vesiculosus* chloroplast genome sequences (Le Corguillé et al., 2009). The whole sequence was verified by sequencing PCR fragments randomly selected based on the assembled genome sequence. All sequences of selected PCR fragments were identical to the original sequences from the assembled chloroplast genome. Every base of the chloroplast genome has minimum quality with Phred value of at least 20 and was confirmed in both directions by a minimum of three reads. The final genome coverage was approximately sixteen-fold with 2216 reads used for the genome assembling. The genome sequence has been deposited in GenBank with the accession number JQ405663. The *S. japonica* chloroplast physical map of the circular genome was drawn using GenomVx (http://wolfe.gen.tcd.ie/GenomeVx/).

Protein-coding and ribosomal RNA genes were identified by BLAST searches (Altschul et al., 1997) of the nonredundant databases at the National Center for Biotechnology Information. The tRNA genes were found using tRNAscan-SE (Lowe and Eddy, 1997).

2.3. Phylogenetic analysis

The phylogenetic analysis was performed with DNA sequences from the chloroplast genome of *S. japonica*, *E. siliculosus*, *F. vesiculosus*, *Vaucheria litorea* (Xanthophyte), *Heterosigma akashiwo* (Raphidophyte) and the pinnate diatom *Fistulifera* sp. strain JPCC DA0580 plus the 15 algal sequences and two reference Cyanobacterium genomes (Khan et al., 2007; Le Corguillé et al., 2009). The phylogenetic analyses were performed with a total of two Cyanobacteria and 21 chloroplast genomes, including four complete genomes from red algae, twelve from heterokont, haptophyte and cryptophyte species, four from green species and one from glaucophytic species.

Two datasets were constructed for phylogenetic analysis (Table 1). The first dataset included 40 chloroplast protein-coding sequences selected from 44 plastid-encoded proteins (Martin et al., 2002; Khan et al., 2007; Le Corguillé et al., 2009). The second dataset used 26 protein coding sequences from 21 out of 23 species excluding *V. litorea* and *Cyanidium caldarum*. The amino acid composition bias of the two datasets was investigated using Tree-Puzzle (Strimmer and von Haeseler, 1996). The concatenated protein sequences were aligned with ClustalX (Larkin et al., 2007) and MEGA5 (Tamura et al., 2011). Each alignment was further optimized using GBlocks under stringent settings (Castresana, 2000). Phylogenetic analyses of the aligned concatenated protein data were carried out on 7140 and 5746 amino acids with unambiguous positions from the respective 40- and 26-protein datasets (Table 1).

Maximum likelihood (ML) analysis for constructing phylogenetic trees was performed using PhyML (Guindon and Gascuel, 2003) and MEGA5 programs (Tamura et al., 2011). PhyML was used under model cpREV with a gamma distribution approximated by 5 categories to model site rate heterogeneity. MEGA5 was performed under the selected model cpREV with a gamma distribution approximated by 5 categories to model site rate heterogeneity.

The neighbor-joining (NJ) analysis was performed under the model JTT with a gamma distribution approximated by 5 categories to model site rate heterogeneity using MEGA5 (Tamura et al., 2011).

For both the ML and NJ analyses, a statistical support for individual branch was investigated by 100 bootstrapping replicates.

Bayesian analyses were performed using MrBayes 3.2 (Huelsenbeck and Rouquist, 2001). MrBayes was performed using the cpREV model for amino acid sequence evolution, including four γ-distributed rate categories and invariant sites under the co-variation model. Two runs with four Markov chains each were run for 100,000–300,000 generations until the average standard deviation of split frequencies was below 0.01 and sampled every hundredth generation with a burn-in period that corresponded to 25% of samples.

CONSEL (Shimodaira and Hasegawa, 2001) was used to statistically test the topologies of trees with approximately unbiased (AU) and Shimodaira–Hasegawa (SH) analyses. Site likelihoods for each phylogenetic tree were calculated using Tree-Puzzle (Strimmer and von Haeseler, 1996).

3. Results

3.1. Features of S. japonica chloroplast genome

The general features of *S. japonica* chloroplast genome (cpDNA) were compared with the chloroplast genomes of *E. siliculosus* and *F. vesiculosus*. All the three chloroplast genomes are divided into large (LSC) and small single copy (SSC) regions separated by two inverted repeats (IRs) (Fig. 1; Table 2). In *S. japonica* and *F. vesiculosus*, the cpDNA IR contains two ribosomal operons encoding the 16S, 23S and 5S rRNA and two tRNAs, however in *E. siliculosus*, it contains additional coding