Aquaculture and Fisheries xxx (2017) 1–8

Contents lists available at ScienceDirect

Aquaculture and Fisheries

journal homepage: www.keaipublishing.com/en/journals/aquaculture-and-fisheries/

Complete sequence of chloroplast genome from *Sargassum vachellianum* (Sargassaceae, Phaeophyceae): Genome structure and comparative analysis

Yan-Hui Bi, Jia-Li Li, Zhi-Gang Zhou*

College of Fisheries and Life Science, Shanghai Ocean University, 999 HuchengHuan Road, Pudong New District, Shanghai 201306, China

ARTICLE INFO

Article history:
Received 14 March 2017
Received in revised form
21 May 2017
Accepted 22 June 2017
Available online xxx

Keywords:
Brown alga
Chloroplast genome
Phaeophyta
Sargassaceae
Sargassum vachellianum

ABSTRACT

Sargassum vachellianum is an ecologically important brown alga. It is China-specific and mainly inhabits in rocky intertidal zones in southeast coastal waters of China. In this study, we sequenced its circular complete chloroplast genome (cpDNA) and compared it with cpDNAs from *S. vachellianum*, *S. horneri* and *S. thunbergii*. The complete *S. vachellianum* cpDNA was 124,582 bp in length and consisted of a pair of inverted repeats (IRs) of 5435 bp, a large single copy (LSC) region of 73,721 bp and a small single copy (SSC) region of 39,991 bp. Totally 160 genes were predicted, including 132 protein-coding genes, four ribosomal RNA genes and 24 tRNA genes, and the coding sequences contributed 77.48% of the whole genome. In addition, 25 SSR loci and 28 highly variable regions were identified from the *S. vachellianum* cpDNA, which might be used as candidates for developing DNA barcode markers of *Sargassum* species. The phylogenetic tree based on datasets of all the plastid-encoded proteins demonstrated that species of *S.* subgenus *Bactrophycus* were firstly combined and then clustered with *S. vachellianum*, which belongs to *S.* subgen. *Sargassum*. The results indicate that the chloroplast genomes are good resources for developing new DNA markers for taxonomy, and also as tools for evolutionary research of closely related species in future studies.

1. Introduction

Chloroplasts are unique organelles of plants and algae. They originated from an endosymbiotic event that occurred more than 1.2 billion years ago, in which a photosynthetic ancestor of present-day cyanobacteria was captured by a mitochondriate host cell. Most of the genetic information of the endosymbiot was lost or was transferred into the nucleus of the host, and the internalized cyanobacterium was transformed into a cell organelle chloroplast. Therefore, compared with nuclear genomes with horizontal gene transfer during evolution, chloroplast DNA is considered a valuable source of genetic markers for phylogenetic analyses since it lacks recombination, has low rates of nucleotide substitution, and is usually uniparentally inherited (Korpelainen, 2004; Provan, Powell, & Hollingsworth, 2001; Ravi, Khurana, Tyagi, & Khurana, 2008; Wolfe, Li, & Sharp, 1987). In addition, other advantages such as its small size, high copy number, simple structure, as well as its conserved gene content and gene arrangement also makes them good candidates for taxonomy and evolutionary research. Many studies have been conducted using chloroplast genomes as phylogenetic (Jansen *et al.*, 2007; Li & Zhang, 2010; Moore, Bell, Soltis, & Soltis, 2007; Moore, Soltis, Bell, Burleigh, & Soltis, 2010; Parks, Cronn, & Liston, 2009) and DNA barcoding markers (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005) to study the phylogenetic relationships of organisms, identify particular DNA samples, as well as analyze the mechanism of molecular phylogeny and the divergence of species (Kim & Archibald, 2009, pp. 1–39).

Over the last five years, the number of fully sequenced chloroplast genomes has increased rapidly owing to next-generation sequencing techniques, which have made genome sequencing time-saving, and low-cost (Shendure & Ji, 2008). In recent years, chloroplast genomes have also been identified in some brown algae. To date, 15 cpDNAs have been sequenced in seven brown algae, *Saccharina japonica* (Wang *et al.*, 2013), *Ectocarpus siliculosus*(Le Corguillé *et al.*, 2009), *Undaria pinnatifida* (Zhang *et al.*, 2015a), *Costaria costata* (Zhang *et al.*, 2015b), *Fucus vesiculosus* (Le

Abbreviations: LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat; NJ, neighbor-joining; SSR, simple sequence repeats.

* Corresponding author.
E-mail addresses: yhbi@shou.edu.cn, zhouzg01@163.com (Z.-G. Zhou).

http://dx.doi.org/10.1016/j.aaf.2017.06.006

Corguillé *et al.*, 2009), *Sargassum horneri* (Liu & Pang, 2016), and *S. thunbergii* (KU500638.1).

S. vachellianum is an ecologically important brown alga (Chromalveolata). It belongs to Sargassaceae family in Phaeophyta and inhabits the subtidal zone of the coasts of China. In this paper, we report the sequencing of the complete chloroplast genome and conduct comparative analyses with the other two available *Sargassum* chloroplast genomes for screening the highly variable regions. In addition, chloroplast simple sequence repeats (SSR) in *S. vachellianum* were also investigated. These results will provide good candidates for developing *Sargassum* genus-specific DNA markers for taxonomy as well as determining more detailed relationships among *Sargassum* species.

2. Materials and methods

2.1. Sample collection and DNA extraction

S. vachellianum adult plants were sampled from Gouqi Island, Zhejiang Province, China (30°43′ N, 122°45′ E) in April, 2014. The collected samples were taken back to the laboratory in coolers (5–8 °C) within 24 h. Approximately 0.3 g of fresh root was cleaned with filtered seawater and ground to fine powder in liquid nitrogen. Total genomic DNA was extracted using a plant genomic DNA kit (Tiangen Biotech, Beijing, China) following the manufacturer's instructions. The DNA concentration was measured in a spectrophotometer at 260 nm according to standard methods (Sambrook).

2.2. Chloroplast genome sequencing, assembling and annotation

The short-insert libraries were constructed with 5 μg of purified DNA according to the manufacturer's instructions (Illumina Inc., San Diego, CA, USA). DNA library construction and sequencing were performed at the Beijing Genomics Institute (BGI) in Shenzhen, China.

The raw sequence reads were screened for plastid-related sequences by using the BLAST software with plastid genomes of *S. japonica* (Wang *et al.*, 2013), *E. siliculosus*, and *F. vesiculosus* (Le Corguillé *et al.*, 2009) as reference sequences. Then the paired-end plastid-related reads were merged by FLASH 1.2.9 (Magoc & Salzberg, 2011) to establish longer reads. For genome assembly scaffolds were generated by assembler and were concatenated in GENEIOUS R8.1 software (Auckland, New Zealand).

The complete sequence was annotated using DOGMA software (Wyman, Jansen, & Boore, 2004). The predicted annotations were verified by alignment against the available chloroplast genomic sequences using BLAST and start and stop sites of the annotated protein-coding genes were manually inspected. The tRNA genes were confirmed using tRNAscan-SE (ver. 1.23) (Lowe & Eddy, 1997). The circular chloroplast genome map was generated by the OGDRAW program (Lohse, Drechsel, Kahlau, & Bock, 2013).

2.3. Comparative genome analysis

Except for *S. vachellianum* cpDNA, there are two other complete chloroplast genomes reported in *Sargassum*, *S. horneri* (Liu & Pang, 2016) and *S. thunbergii*. Translocations and inversions were analyzed by pair-wise comparisons among these three *Sargassum* chloroplast genomes using PipMaker (Schwartz *et al.*, 2000). Multiple alignments were conducted using VISTA (Mayor *et al.*, 2000) with *S. vachellianum* chloroplast genome sequence as the reference.

The results of the chloroplast genome annotation varied with the different annotation methods or software, for this reason the three *Sargassum* cpDNAs were initially analyzed with the same method. Then, each predicted genes (≥150 bp) and the intergenetic spacer regions (≥150 bp) were aligned among the three species separately using ClustalX1.83 software (Chenna *et al.*, 2003). The corresponding sequence similarities were calculated using BioEdit (Hall, 1999). The highly variable regions in the IR/SC borders were found even between closely related genera of the same family (Kim & Lee, 2004), the border positions as well as the corresponding sequences among these three *Sargassum* and *F. vesiculosus* (NC_016735.1) (Le Corguillé *et al.*, 2009)chloroplast genomes were also compared.

2.4. Phylogenetic analysis

Phylogenetic analysis among *S. vachellianum*, *S. horneri* (Liu & Pang, 2016), *S. thunbergii* (KU500638.1), and *F. vesiculosus* (Le Corguillé *et al.*, 2009) chloroplast genomes was conducted with *E. siliculosus* (Le Corguillé *et al.*, 2009) and *S. japonica* (Wang *et al.*, 2013) cp DNAs as outgroup. All the amino acid sequences of the plastid-encoded proteins were extracted for evaluation of the phylogenetic relationships. Sequence alignments were processed using MEGA 6.06 software (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). The evolutionary distances were computed based on the Kimura two parameter model (Kimura, 1980). A neighbor-joining (NJ) tree was constructed with 1000 bootstrap replicates using MEGA 6.06 (Tamura et al., 2013).

3. Results and discussions

3.1. Genome features

The circular chloroplast genome of *S. vachellianum* (GenBank accession number: KT188823) is 124,582 bp in length (Fig. 1), which is similar to the available *Sargassum* chloroplast genomes (124,068-124,592 bp). It contained two inverted repeat regions (IR) of 5435 bp separated by one small copy region (SSC) of 39,991 bp and one large single copy region (LSC) of 73,721 bp. A total of 160 unique genes were predicted (Table 1), which included 132 protein-coding genes, 4 ribosomal and 24 tRNA genes. Two tRNA genes (*trnI-GAU* and *trnA-UGC*) and all the rRNA genes were located in the IRs. No introns were identified in the plastid genes of *S. vachellianum*. The coding sequences composed the major part of the *S. vachellianum* chloroplast genome, and contributed 77.48% of the total genomic sequence. The remaining 22.52% corresponded to the spacer regions, which were 28,051 bp in length. The average length of the spacer regions was 178.7 bp, which was between that of the other two *Sargassum* chloroplast genomes (175.3–180.5 bp). The overall GC and AT content were 30.44% and 69.56%, respectively.

In general, chloroplast genomes could be classified into three types according to their IRs. Chloroplast genomes lacking IRs were defined as group I, those containing IRs were defined as group II, and chloroplast genomes with tandem repeats were defined as group III (Sugiura, 1992). Since the *S. vachellianum* chloroplast genome contained two IRs and was classified as a group II chloroplast DNA, in common with all other reported brown algae chloroplast genomes (Le Corguillé *et al.*, 2009; Liu & Pang, 2016; Wang *et al.*, 2013; Zhang *et al.*, 2015a, 2015b).

3.2. Simple sequence repeats

In this study, simple sequence repeats (SSRs) containing more than six repeat motifs were screened in the *S. vachellianum* chloroplast genome. As a result, totally 25 SSR loci were identified (Table 2), including two types of perfect SSRs and one type of compound SSR. Perfect SSRs of single nucleotide repeats were five A stretches (10, 11 bases), ten T stretches (10, 11 bases), and one G stretch (11 bases), but no C stretches. Perfect SSRs of two base

Please cite this article in press as: Bi, Y.-H., et al., Complete sequence of chloroplast genome from *Sargassum vachellianum* (Sargassaceae, Phaeophyceae): Genome structure and comparative analysis, Aquaculture and Fisheries (2017), http://dx.doi.org/10.1016/j.aaf.2017.06.006