



Short communication

The complete chloroplast genome sequence of *Abies nephrolepis* (Pinaceae: Abietoideae)

Dong-Keun Yi, Kyung Choi, Minjung Joo, Jong Cheol Yang, Feruza U. Mustafina, Jun-Soo Han, Dong Chan Son, Kae Sun Chang, Chang Ho Shin, You-Mi Lee*

Division of Forest Biodiversity and Herbarium, Korea National Arboretum, Pocheon 11186, South Korea

ARTICLE INFO

Article history:

Received 3 March 2016

Received in revised form

15 March 2016

Accepted 16 March 2016

Available online 31 March 2016

Keywords:

Abies

chloroplast genome

Khinghan fir

NGS sequencing

ABSTRACT

The plant chloroplast (cp) genome has maintained a relatively conserved structure and gene content throughout evolution. Cp genome sequences have been used widely for resolving evolutionary and phylogenetic issues at various taxonomic levels of plants. Here, we report the complete cp genome of *Abies nephrolepis*. The *A. nephrolepis* cp genome is 121,336 base pairs (bp) in length including a pair of short inverted repeat regions (IRa and IRb) of 139 bp each separated by a small single copy (SSC) region of 54,323 bp (SSC) and a large single copy region of 66,735 bp (LSC). It contains 114 genes, 68 of which are protein coding genes, 35 tRNA and four rRNA genes, six open reading frames, and one pseudogene. Seventeen repeat units and 64 simple sequence repeats (SSR) have been detected in *A. nephrolepis* cp genome. Large IR sequences locate in 42-kb inversion points (1186 bp). The *A. nephrolepis* cp genome is identical to *Abies koreana*'s which is closely related to taxa. Pairwise comparison between two cp genomes revealed 140 polymorphic sites in each. Complete cp genome sequence of *A. nephrolepis* has a significant potential to provide information on the evolutionary pattern of Abietoideae and valuable data for development of DNA markers for easy identification and classification.

Copyright © 2016, National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA). Production and hosting by Elsevier. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The chloroplasts (cp) are specialized intracellular organelles for photosynthesis. The cp genomes have maintained a relatively conserved structure and gene content throughout evolution (Shinozaki et al 1986; Douglas 1998). The complete cp genomes of more than 830 accessions have been reported from seed plants until now National Center for Biotechnology Information (NCBI) Organelle Genome Resources, (<http://www.ncbi.nlm.nih.gov/genomes/>). The majority of spermatophytes contain 90–110 unique genes within 120–170 kb of circular chromosome (Sugiura 1992). The gene contents and the polycistronic transcription units of cp genome are generally conserved among seed plant species with the exception of nonphotosynthetic parasitic plants. The gene order of the cp genome is also relatively

conserved; however, it was frequently modified by inversion mutations. Various structural modifications including loss of inverted repeat (IR) or short IR regions, rearrangement of gene order and loss of some genes are reported for gymnosperm cp genomes (Wu et al 2009; Lin et al 2010; Wu et al 2011; Wu and Chaw 2014).

The cp genome sequences have been widely used for resolving evolutionary and phylogenetic issues at various taxonomic levels of plants (Raubeson and Jansen 1992; Downie et al 1996; Jansen et al 2007). Phylogenetic analyses and taxonomic systematizations are performed by a number of cp markers (Kress and Erickson 2007; Erickson et al 2008; Group CPW 2009; Hollingsworth et al 2011). Complete cp genome sequencing provides valuable information for determination of suitable cp DNA markers for plant species classification. Moreover, acquisition of genome sequence nowadays does not demand much time and funds due to availability and accessibility of sequencing technologies.

Firs (genus *Abies*, family Pinaceae) include about 48 species of evergreen coniferous trees (Suyama et al 2000; Farjon 2010; Xiang et al 2015). *Abies nephrolepis* (Traits. ex Maxim.) Maxim. is

* Corresponding author. Tel.: +82 0315401001.

E-mail address: sl1962@korea.kr (Y.-M. Lee).

Peer review under responsibility of National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA).

commonly known as the Khinghan fir and is an important timber tree in Northeast Asia. *A. nephrolepis* shows close relationships with *Abies koreana*, *Abies sachalinensis*, *Abies sibirica*, and *Abies veitchii*. Complete cp genome sequence of *Abies koreana* has already been published (Yi et al 2015). Thus, analysis of cp genome structure of *A. nephrolepis* and close related species, as well as sequence comparison, will contribute to a better understanding of the evolutionary mode of the cp genome and marker development for using in *Abies* genus classification.

Materials and methods

Plants materials and DNA extraction

Approximately 5 g of fresh leaves of *Abies nephrolepis* (Trautv. ex Maxim.) Maxim. were collected from a single individual in the natural forest habitat at the southeast part of South Korea. Voucher specimen and DNA sample (KHB1343154) were deposited in Korea National Arboretum (KH). Total genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle 1987). Extracted DNA was quantified using NanoDrop spectrophotometers (Nanodrop 2000; Thermo Scientific, Wilmington, DE, USA).

Sequencing and chloroplast genome assembly

Chloroplast genome sequences were analyzed using Illumina HiSeq 2000 (Illumina, San Diego, CA, USA). A total number of 35,665,348 reads were analyzed to generate 3,602,200,148 base pairs (bp) of sequence. Acquisition of precise sequence data allowed low quality reads (< Q20) to be filtered out from raw data. The filtered sequences were assembled using the Bowtie2 software (ver. 2.2.3, <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>; Langmead and Salzberg 2012) with reference sequence of *A. koreana*. A total number of 433,280 reads were mapped in reference sequence with an average coverage of 346.7X. Gaps or uncertain sites were filled and resequenced via direct sequencing of Polymerase Chain Reaction (PCR) products amplified using primers designed from assembled sequences and previously published primers (Tsumura et al 2000). The sequenced fragments were assembled using Geneious 8.0.3 (Biomatters, Auckland, NZ, <http://www.geneious.com>; Kearse et al 2012).

Chloroplast gene annotation

Gene annotations were performed using the Basic Local Alignment Search Tool (BLAST; BLASTN, PHI-BLAST and BLASTX), Open Reading Frame (ORF) finder program from the National Center for Biotechnology Information and DOGMA (Wyman et al 2004). The nomenclature of cp genes was used according to Chloroplast Genome Database (<http://chloroplast.cbio.psu.edu>) and previous published cp genomes (Cui et al 2006). A circular map of cp genomes was drawn by OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/>; Lohse et al 2007).

Sequence analyses

A-T contents were evaluated by MEGA6 (version 6.06, <http://www.megasoftware.net>; Tamura et al 2013). Repeating sequences were analyzed by using REPuter (<https://bibiserv2.cebitec.uni-bielefeld.de/reputer>; Kurtz et al 2001) and Tandem Repeats Finder, ver. 4.07b (<https://tandem.bu.edu/trf/trf.html>; Benson 1999). Simple sequence repeat (SSR) loci were identified by Simple Sequence Repeats Extractor (<http://www.aridolan.com/ssr/ssr.aspx>). All SSR regions were PCR amplified and

resequenced manually to mitigate errors from the Next Generation Sequencing (NGS) sequencing procedure. Protein coding genes sequences were aligned using MUSCLE (Edgar 2004) and adjusted by hand. Nucleotide diversity and ratio of divergence at nonsynonymous and synonymous sites (dN/dS) value were analyzed using DnaSP (version 5.10.01, <http://www.ub.edu/dnasp>; Librado and Rozas 2009) and MEGA6. Mauve whole genome aligner (Darling et al 2004) was used for alignment of complete cp genomes.

Results

General features of the *A. nephrolepis* cp genome

The complete cp genome of *A. nephrolepis* is 121,336 bp in length (Figure 1). Its sequence is deposited in GenBank with the accession number KT834974. The *A. nephrolepis* cp genome includes a pair of short inverted repeat regions (IRa and IRb) of 139 bp each separated by a small single copy region of 54,323 bp (SSC) and a large single copy region of 66,735 bp (LSC). It contains 114 individual genes including 68 protein encoding genes, 35 tRNA and four rRNA genes, six open reading frames, and one pseudogene. *A. nephrolepis* is identical to *A. koreana* in cp genome content and structure (Yi et al 2015). Fifty-three protein coding and 16 tRNA genes, three open reading frames, and one pseudogene are located in the LSC region, while 15 protein coding and 17 tRNA genes, four rRNA, and three open reading frames are sited in the SSC region. Only one tRNA gene (*trnI*-CAU) is duplicated and found in IR regions. The cp genome of *A. nephrolepis* lacks NADH dehydrogenase genes (*ndh* genes) as reported for the cp genome of other representatives of the family Pinaceae (Lin et al 2010; Wu et al 2011). Twelve genes contain one intron, *rps12* and *ycf3* genes have two introns. The major part (56.0%) of the *A. nephrolepis* cp genome consists of gene-coding regions (50.1% protein and 5.9% RNA coding regions), whereas the intergenic spacers (including 15 introns) comprise 44.0%. The overall A-T content of *A. nephrolepis* cp genome is 61.7%, which is slightly lower than in the cp genome of other conifers and generally in angiosperms (Shinozaki et al 1986; Wu et al 2009; Lin et al 2010; Wu et al 2011; Yi and Kim 2012; Wu and Chaw 2014). The A-T content in noncoding regions (64.6%) is higher than in coding regions (59.5%). IR regions contain 58.3% of A-T while LSC and SSC regions have 62.6% and 60.8%, respectively (Table 1).

Examination of repeat unit and SSRs in *A. nephrolepis* cp genome

Seventeen repeat units of at least 22 bp length are identified in the *A. nephrolepis* cp genome. One 1186-bp length repeat is novel for the IR region. It is described for cp genomes for some representative of *Abies* and *Tsuga* genera (Tsumura et al 2000). The other 16 repeat units include six direct, six dispersed and four palindromic repeats (Table 2). Sixty-four SSRs with the length longer than 10 bp are identified. Thirty-six mononucleotide SSRs contain 3 Cs, 1 Gs, and 32 As or Ts. Fourteen dinucleotide SSRs contain ATs or TAs and none of GAs. The other SSRs are identified as dinucleotides or trinucleotides, 10 tetranucleotides (3 repeats were duplicated in di-nucleotide) and two pentanucleotides (Table S1).

Comparative analysis of cp genomes

A. nephrolepis is identical to *A. koreana* in cp genome structure, gene order, and its content. Pairwise sequence comparison between their cp genomes revealed high identity (99.9%) and

Download English Version:

<https://daneshyari.com/en/article/4394959>

Download Persian Version:

<https://daneshyari.com/article/4394959>

[Daneshyari.com](https://daneshyari.com)