The complete chloroplast genome of *Abies ernestii* Rehder (Pinaceae) and its phylogenetic implications

Yi-Zhen Shao*, Zhi-Yuan Shi*, Zhao Wang*, Wei Wang*, Yun Chen* and Qian Wen*

*College of Life Sciences, Henan Agriculture University, Zhengzhou, China; College of Resources and Environmental Sciences, Henan Agriculture University, Zhengzhou, China

**ABSTRACT**

*Abies ernestii* Rehder is endemic to the montane regions of Southwest China. Till now, phylogenetic relationships between *A. ernestii* and other closely related species remain unclear. In this study, we first characterized the complete chloroplast (cp) genome of *A. ernestii*. The whole cp genome was 121,841 bp in size, including one hundred and thirteen genes. Results of comparative cp genome revealed that only *ycf1* and *ycf2* was characterized by a considerable variation. Our phylogenetic analyses supported the monophyly of the genus *Abies* and revealed a clear separation between *A. ernestii* and *A. chensiensis* Tiegh. This study highlights the significance of using cp genomes to examine species boundaries among closely related fir species.

**Introduction**

*Abies ernestii* Rehder is distributed across Southwest China and characterized by high economic and ecological importance (Farjon and Rushforth 1989; Farjon 1990). It grows in cold, moist forests of the Hengduan Mountain region at 2600–3800 m elevations, with annual precipitation of 1100–2200 mm (Kuan 1981). *Abies ernestii* differs from *Abies chensiensis* in the length of needles and color of cones (green or pale purplish), in addition to its distribution in the north and west of the Sichuan province, extending into extreme north-western Yunnan province (Liu 1971). However, the species delimitation between *A. ernestii* and *A. chensiensis* has been a continuous source of debate (Liu 1971; Farjon 2001). Recently, several studies have been conducted on the species delimitation between *A. ernestii* and *A. chensiensis* using morphological or molecular data (Suyama et al. 2000; Xiang et al. 2004, 2009, 2015; Liepelt et al. 2010; Aguirre-Planter et al. 2012). However, no robust phylogenetic relationships have been reported owing to the lack of high-resolution molecular markers (e.g. complete chloroplast [cp] genomes) (Shao and Xiang 2015; Xiang et al. 2018).

In this study, we first determined the cp genome of *A. ernestii* using next-generation sequencing technology. Thereafter, we compared it with other available genomes to explore their genetic divergence, develop molecular markers, and reconstruct phylogenetic relationships. This study evaluated the significant value of cp genomes to examine the species boundaries among closely related fir species.

**Materials and methods**

**Sample collection and DNA extraction**

The specimens and leaf materials of *A. ernestii* were collected by Xianchun Zhang from Wenchuan county, Sichuan province, China (30.55°N, 103.03°E). The voucher specimen was deposited in the herbarium of the Institute of Botany, CAS (PE) (http://pe.ibcas.ac.cn/, Qin Ban, herbarium2@ibcas.ac.cn) under the voucher number 5874. The leaf materials were first desiccated in silica gel, and the Ezup plant genomic DNA prep kit was used to extract the total DNA. The relevant cp genome sequence was submitted to the NCBI database (No. MH706707). The associated numbers were PRJNA790664 (BioProject), SRP351570 (SRA), and SAMN24219951 (Bio-Sample), respectively.

**Polymerase chain reaction amplification, sequencing, and assembly**

The extracted DNA was sequenced on the Illumina HiSeq X platform with libraries in 350 bp length. The clean reads were
approximately 10.2 million in 150 bp length. We aligned, assembled, and annotated the reads using CLC de novo assembler (v. beta 4.6, CLC Bio, Aarhus, Denmark), GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html), and tRNAscan-SE v1.3.1 (Schattner et al. 2005; Tillich et al. 2017). Few adapt-related reads were recognized in the raw reads. These reads were trimmed \( N > 10\% \) or \( Q \leq 5 \) to ensure high-quality data and then assigned to the genome sequence using Velvet (Zerbino and Birney 2008). The cp genome sequence was annotated in tRNAscan-SE and GeSeq (Schattner et al. 2005; Tillich et al. 2017). To match the gene predictions, we checked all the start/stop codons and intron/exon boundaries in Sequin 15.50 and Geneious 8.0.2 (Kearse et al. 2012; Lohse et al. 2013). Finally, we annotated the sequences by comparing them with the published genomes of \( A. \) chensiensis (MH706706 and MH047653).

**Figure 1.** The chloroplast genome map of \( A. \) ernestii. Genes shown outside the circle are transcribed clockwise, and genes inside are transcribed counter-clockwise. Genes belonging to different functional groups are color-coded. The darker grey in the inner corresponds to the GC content and the lighter grey to the AT content.

**Repeat sequence detection and comparative analysis of cp genomes**

The MISA software was used to check the simple repetitive DNA sequences (SSRs) in the cp genome of \( A. \) ernestii.
Subunits of cytochrome b/f complex peptide-encoding genes (68), tRNA genes (16), open reading frames (54,156 bp, respectively. It contained 113 genes, including 24 protein-coding genes, 16 tRNA genes, and 3 open reading frames. The SSC region and short single-copy (SSC) region were 67,157 and 16,102 bp, respectively. The large single-copy (LSC) region was characterized by a typical quadripartite structure, similar to the cp genomes in Pinaceae. The large single-copy (LSC) region was 102,901 bp in length. We aligned the sequences containing the hotspot repeats were 50 bp, and minimal repeat size was 30 bp. Using the REPuter website (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) (Kurtz et al. 2001). The parameters used in the analysis were as follows: the hamming distance was 3, maximum computed repeats were 50 bp, and minimal repeat size was 30 bp. Using the mVista program with Shuffle-LAGAN mode, we compared the whole cp genome of A. ernestii, A. chensiensis (MH706706 and MH047653), and Abies fargesii Franch. (MH706716) (Frazer et al. 2004). The repeat units were set to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively.

**Phylogenetic analyses**

Phylogenetic analysis was performed using the complete cp genomes. We aligned the sequences containing the hotspot mutation regions and chose the maximum-likelihood (ML) analysis as the method in this study. The analysis was performed using a rapid bootstrap analysis and 1000 rapid bootstrap search steps in the RAxML v.7.8 (Stamatakis 2014). The relevant bootstrap value under each node was obtained from FigTree v.2.2. We collected the complete plastomes of 16 fir species from the NCBI database (Figure 3, Table S1). Juniperus squamata Buchanan–Hamilton ex D. Don in the family Cupressaceae was used as an outgroup.

**Results and discussion**

**Chloroplast genome features of A. ernestii**

The complete cp genome of A. ernestii was 121,841 bp in length, and its G and C content was ~38.30%. This genome was characterized by a typical quadripartite structure, similar to the cp genomes in Pinaceae. The large single-copy (LSC) region and short single-copy (SSC) region were 67,157 and 54,156 bp, respectively. It contained 113 genes, including peptide-encoding genes (68), tRNA genes (16), open reading frames (6), and rRNA genes (4), respectively (Figure 1, Table S2). In the LSC region, there were 53 protein-coding genes, 16 tRNA genes, and 3 open reading frames. The SSC region owned all the 4 rRNA. The IR regions were 264 bp and featured trnT-CAU and trnT-GGU. Among the encoded genes, ycf3 and rpl2 had two introns, and 11 genes (atpF, trnV-UAC, trnA-UGC, trnL-UAA, trnK-UUU, trnG-GCC, rps12, petD, rpoC1, petB, and rpl16) had one intron (Table 1). The palindromic inverted repeat, including trnS, psaM, ycf12, and trnG, was located in 52-kb inversion points and 1180 bp in length. Such inverted repeat was identical to the published cp genomes of fir species (Dong et al. 2021). In addition, all ndh genes were lost as other cp genomes of Pinaceae, such as Keteleeria davidioida var. calcaraean (W. C. Cheng & L. K. Fu) Silba (Li et al. 2021). As proposed by Blazier et al. (2011), the ndh genes were initially discovered through their homology to complex I in the mitochondrial respiratory electron transport. The lack of ndh genes across various species in Pinaceae indicated its dispensability, suggesting that there might be alternative or supplementary electron transport pathways to achieve this function.

**Repeat sequence analysis**

The simple sequence repeats (SSRs) have been widely used in phylogenetic and phylogeographic analyses of highly polymorphic genetic materials (Kaur et al. 2015). In this study, we detected 67 SSRs in the cp genome of A. ernestii, and they included 42 mononucleotide repeats, 14 dinucleotide repeats, 2 trinucleotide repeats, 7 tetranucleotide repeats, and 2 pentanucleotides repeats (Figure S1). Compared with other SSRs, the mononucleotide repeats (62.68%) were the most abundant and they contributed more to the genetic variations. Most mononucleotide repeats were A or T, and three types of dinucleotide SSRs AG/CT/AT were identified. In addition, we detected two types of trinucleotide SSRs (AAT/ATT), AGG/GCG, and ACG/GCG.
twelve types of tetranucleotide SSRs (AAAG/CTTT/AAAT/ATTT/AACC/GGTT/ACCT/AGGT/AGAT/ATCT/ATCC/ATGG), and four types of pentanucleotides SSRs (AACAT/ATGTT/AATCG/ATTCG) (Figure S1). Furthermore, sixty-four long repeats were revealed, and they included 37 forward repeats, 14 tandem repeats, and 13 palindromic repeats (Figure S1).

**Comparative genome analysis**

To perform the comparative genome analysis, we used three closely related firs: *A. ernestii*, *A. chensiensis*, and *A. fargesii* (Figure 2). In the phylogenetic analysis, they clustered together as a monophyletic lineage, with an extremely high
bootstrap (BSML = 100) (Figure 3). Our results indicated that the IR region was more conservative than the SSC and LSC regions. Almost all the genetic variability was concentrated in the intergenic or noncoding region, corroborating the extremely low resolution among closely related fir species in previous phylogenetic studies (Figure 2) (Xiang et al. 2004, 2009, 2015). In the coding regions, only ycf1 and ycf2 were characterized by a considerable variation between A. ernestii and A. chensiensis (Figure 2). ycf1 had been proposed as the most variable coding region that was better than the existing plastid barcodes and could serve as a DNA barcode (Dong et al. 2015). Unfortunately, the cp markers widely used in the Abies phylogenetic studies, for example, matK, rpl16, rps12-rpl20, rps18, trnC-D, trnS-G, and trnT-F, were highly conservative among the three closely related species (Xiang et al. 2018).

**Phylogenetic analysis**

To infer the phylogenetic status and evolutionary relationship of A. ernestii, we selected 16 published cp genomes of fir species, with Juniperus squamata as an outgroup. Based on the best ML phylogram, 17 Abies species were supported as one monophyletic lineage (BSML = 100) (Figure 3). Abies balsamea (L.) Mill. was distributed in North America and formed a sister lineage with the firs from East Asia (BSML = 100). Among the East Asia firs, Abies beshanzuensis M. H. Wu was nested with others (BSML = 97). Abies ernestii, A. chensiensis, and A. fargesii clustered together as a monophyletic lineage, with an extremely high bootstrap (BSML = 100). Our results further indicated that A. ernestii and A. chensiensis were separated with an extremely high bootstrap (BSML = 100); thus, their current species status was well corroborated. The treatments of A. ernestii as a variety of A. chensiensis, or an identical subspecies of A. chensiensis were rejected (Handel-Mazzetti 1929; Dallimore and Jackson 1966). Such a robust phylogenetic relationship between A. ernestii and A. chensiensis has not been revealed previously (Xiang et al. 2004; Jaramillo-Correa et al. 2008; Semerikova et al. 2011, 2018).

This study provides new insights into the species delimitation and phylogenetic relationship in the genus Abies. Our results validated the reliability of using cp genome data to delimitate problematic groups at low taxonomic levels. These data provided important genetic resources for these ecologically important Abies species and further cp genome evolution research.

**Acknowledgments**

The authors thank Professor Xian-Chun Zhang (Institute of Botany, Chinese Academy of Sciences) for helping in the fieldwork.
Ethical approval

All procedures in this study involving plants followed the Plant Use and Collection guidelines of Henan Agriculture University and the Institute of Botany, Chinese Academy of Sciences (IBCAS).

Author contributions

Yi-Zhen Shao, Yun Chen, and Qian Wen contributed to the conception and design of the study, analysis, and interpretation of the data. Yi-Zhen Shao and Zhi-Yuan Shi wrote the first version of the manuscript. Zhao Wang, Wei Wang, Yun Chen, and Qian Wen critically reviewed and modified the article regarding its intellectual content. All authors read, discussed, and approved the final version, and all authors agree to be accountable for all aspects of the work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Key Project of Science and Technology Research of Henan Province, China [212102110155], and the National Natural Science Foundation of China [31600163].

Data availability statement

The genome sequence data supporting this study’s findings are available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/) under accession no. MH706707. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA790664, SRP351570, and SAMN24219951, respectively.

References

Aguirre-Planter É, Jaramillo-Correa JP, Gómez-Acevedo S, Khasa DP, Bousquet J, Eguiarte LE. 2012. Phylogeny, diversification rates and species boundaries of Mesoamerican firs (Abies, Pinaceae) in a genome-wide context. Mol Phylo Evo. 62(1):263–274.

Blazier JC, Guisinger MM, Jansen RK. 2011. Recent loss of plastid-encoded ndh genes within Erodium (Geraniaceae). Plant Mol Biol. 76(3–5): 263–272.

Dallimore W, Jackson AB. 1966. A handbook of Coniferae and Ginkgoceae. 4th ed. London (UK): Edward Arnold (Publishers) Ltd.

Farjon A. 1990. Pinaceae: drawings and descriptions of the genera Abies, Cedrus, Pseudolarix, Keteleeria, Nothotsuga, Tsuga, Cathaya, Pseudotsuga, Larix and Picea (Regnum Vegetabile 121). Königstein (Germany): Koeltz Scientific Books; p. 9–109.

Farjon A. 2001. World checklist and bibliography of conifers. 2nd ed. London (UK): The Royal Botanic Gardens; p. 109–134.

Farjon A, Rushforth KD. 1989. A classification of Abies Miller (Pinaceae). Notes of the Royal Botanical Gardens Edinburgh. 46(1):59–77.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32: W273–W279.

Fu ZX, Wang XY, Fan PZ, Tian XY, Shao YZ. 2019. The complete chloroplast genome of the endangered Pinaceae species Abies ziyuanensis and its phylogenetic implications. Mitochondrial DNA B. 4(1):137–138.

Guo ZN, Lu XD, Dong YB, Wu JC, Li GY. 2019. Next-generation sequencing yields the complete chloroplast genome of Abies fangjingshanensis, an Endangered species from South China.Mitochondrial DNA B. 4(1): 880–881.

Guo C, Xu Y. 2019. Next-generation sequencing yields the complete chloroplast genome of Abies fargesii. Mitochondrial DNA B. 4(1): 1273–1274.

Handel-Mazzetti H. 1929. Symbolae Sincae, vol. 7. J. Springer, Wien, 4–18 pp.

Jaramillo-Correa JP, Aguirre-Planter E, Khasa DP, Eguiarte LE, Piñero D, Furnier GR, Bousquet A. 2008. Ancestry and divergence of subtropical montane forest isolates: molecular biogeography of the genus Abies (Pinaceae) in southern México and Guatemala. Evolution. 17(10): 2476–2490.

Kaur S, Panesar PS, Bera MB, Kaur V. 2015. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. Crit Rev Food Sci Nutr. 55(1):41–49.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12): 1647–1649.

Kuan CT. 1981. Fundamental features of the distribution of coniferae in Sichuan. Acta Phytotaxonomica Sinica. 19(4):393–407.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29(22):4633–4642.

Li JR, Chen JR, Wang JK, Zheng WL. 2020. The complete chloroplast genome of Abies georgei Orr var. smithii, a species endemic to the Qinghai-Tibet Plateau, China. Mitochondrial DNA B. 53(3):2553–2554.

Li JJ, Cheng NY, Shi YC. 2021. The complete chloroplast genome of Keteleeria davidiana var. calcarea (Pinaceae), an endangered species endemic to China. Mitochondrial DNA B. 6(2):693–695.

Liepelt S, Mayland-Quellhorst E, Lahme M, Ziegenhagen B. 2010. Contrasting geographical patterns of ancient and modern genetic lineages in Mediterranean Abies species. Plant Syst Evol. 284(3–4): 141–151.

Liu TS. 1971. A monograph of the genus Abies. Taipei (Taiwan): Department of Forestry, College of Agriculture, National Taiwan University.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41:75–81.

Schattner P, Brooks AN, Lowe TM. 2005. RePuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29(22): 4633–4642.

Shao YZ, Hu JT, Fan PZ, Liu YY, Wang YH. 2018. The complete chloroplast genome sequence of Abies beshanzuensis: inferred from both nuclear and cytoplasmic markers. J Biogeogr. 38(2): 326–340.

Shao Y, Chen Y, Zhang X, Xiang Q. 2020. Species delimitation and phylogeography of Abies georgei var. smithii using AFLPS and mitochondrial DNA. J Biogeogr. 38(2): 326–340.

Shao YZ, Hu JT, Fan PZ, Liu YY, Wang YH. 2018. The complete chloroplast genome sequence of Abies beshanzuensis, a highly endangered fir species from south China. Mitochondrial DNA B. 3(2):923–924.

Shao YZ, Xiang QP. 2015. Species delimitation and phylogeography of Abies chensiensis complex: inferred from morphological and molecular data. Bot J Linn Soc. 177(2):175–188.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9): 1312–1313.
Su L, Zhao PF, Lu XF, Shao YZ. 2019. The complete chloroplast genome sequence of *Abies chensiensis* (Pinaceae). Mitochondrial DNA B Resour. 4(2):3262–3263.

Suyama Y, Yoshimaru H, Tsumura Y. 2000. Molecular phylogenetic position of Japanese *Abies* (Pinaceae) based on chloroplast DNA sequences. Mol Phylogenet Evol. 16(2):271–277.

Tillich M, Lehward P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1):W6–W11.

Wu S, Lu XF, Zhao PF, Shao YZ. 2019. Next-generation sequencing yields the complete chloroplast genome of *Abies balsamea*. Mitochondrial DNA B. 4(1):1445–1446.

Xiang QP, Wei R, Shao YZ, Yang ZY, Wang XQ, Zhang XC. 2015. Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. Mol Phylo and Evo. 82:1–14.

Xiang QP, Wei R, Zhu YM, Harris AJ, Zhang XC. 2018. New infrageneric classification of *Abies* in light of molecular phylogeny and high diversity in western North America. J Syst Evo. 56(5):562–572.

Xiang QP, Xiang QY, Guo YY, Zhang XC. 2009. Phylogeny of *Abies* (Pinaceae) inferred from nrITS sequence data. Taxon. 58(1):141–152.

Xiang QP, Xiang QY, Liston A, Zhang XC. 2004. Phylogenetic relationships in *Abies* (Pinaceae): evidence from PCR-RFLP of the nuclear ribosomal DNA internal transcribed spacer region. Bot J Linn Soc. 145(4):425–435.

Yi DK, Choi K, Joo M, Yang JC, Mustafina FU, Han J-S, Son DC, Chang KS, Shin CH, Lee Y-M. 2016. The complete chloroplast genome sequence of *Abies nephrolepis* (Pinaceae: Abietoideae). J Asia Pac Biodivers. 9(2):245–249.

Yi DK, Yang JC, So S, Joo MJ, Kim DK, Shin CH, Lee YM, Choi K. 2015. The complete plastid genome sequence of *Abies koreana* (Pinaceae: Abietoideae). Mitochondrial DNA. 27(4):1.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18(5):821–829.

Zhang YY, Xiang RC, Dong NL, Liu YY, Shao YZ. 2020. Next-generation sequencing yields the complete chloroplast genome of *Abies yunnanboshiensis*, an endangered species from South China. Mitochondrial DNA B Resour. 5(4):3821–3822.