Complete chloroplast genome of a valuable medicinal plant, Huperzia serrata (Lycopodiaceae), and comparison with its congener

Zhi-You Guo1,2,3,5, Hong-Rui Zhang2,4,5, Nawal Shrestha4, and Xian-Chun Zhang2,6

1State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, People’s Republic of China; 2College of Biological Sciences and Agriculture, Qianan Normal University for Nationalities, Guizhou 558000, People’s Republic of China; and 3University of Chinese Academy of Sciences, Beijing 100049, People’s Republic of China

Methods and Results: The whole chloroplast genome of H. serrata was sequenced using an Illumina platform and assembled with Geneious version R9.0.5. The genome size of H. serrata was 154,176 bp, with 36.3% GC content. The complete chloroplast genome contained 120 unique genes, including 86 coding genes, four rRNA genes, and 30 tRNA genes. Comparison with the chloroplast genome of H. lucidula revealed three highly variable regions (rps16-chlB, ycf2-trnR, and ycf1) between these two species and 252 mutation events including 27 insertion/deletion polymorphisms and 225 single-nucleotide polymorphisms (SNPs). Ninety-two SNPs were identified in the gene-coding regions. In addition, 18 microsatellite sites were found, which can potentially be used in phylogeographic studies.

Conclusions: The complete chloroplast genome of H. serrata is reported here, and will be a valuable genome resource for further phylogenetic, evolutionary, and medical studies of medicinal plants in the genus Huperzia.

Key words: Huperzia serrata; lycophytes; Lycopodiaceae; mutation; next-generation sequencing.

The structure of chloroplast genomes in land plants is generally highly conserved in terms of gene order, organization, and content, which makes them suitable for characterizing genetic relationships among species (Bock, 2007). Portions of these genomes have also been widely used by many plant taxonomists as effective DNA barcoding tools. Most of the chloroplast genomes of land plants have a pair of inverted repeats (IRs), separated by one large single copy region (LSC) and one small single copy region (SSC) (Jansen et al., 2005). However, variations occur in certain lineages, and these variations have been proven to be useful in identifying some critical events during the evolution of land plants (Dong et al., 2014; Song et al., 2015). One typical example is the 30-Kb inversion (from trnC to ycf2) detected in the LSC region from bryophytes and lycophytes to other land plants, supporting the hypothesis that lycophytes are a sister clade to all other extant vascular plants (Raubeson and Jansen, 1992).

Compared with those on seed plants, studies on chloroplast genomes of ferns and lycophytes have been relatively sparse (Lu et al., 2015). The North American fir moss Huperzia lucidula (Michx.) Trevis. (Lycopodiaceae) was the first lycophyte species with a complete chloroplast genome sequence (Wolf et al., 2005; GenBank accession no. NC_006861). Because H. lucidula belongs to a significant sister clade of all extant vascular plants, sequencing its complete chloroplast genome facilitates the exploration of the relationships between lycophytes and other vascular plants. Both the rearrangement structure of the chloroplast genome and the phylogenomic analyses of 73 protein-coding genes supported the hypothesis that lycophytes were a sister to both extant fern and seed plant lineages (Wolf et al., 2005).

However, the phylogenetic relationships within this family and particularly within the genus Huperzia Bernh. (ca. 55 species) are still unclear because of insufficient phylogenetic data (Zhang and Iwatsuki, 2013). Here we describe the complete chloroplast genome sequence of a valuable species (H. serrata (Thunb.) Trevis.) within this genus and compare it to existing chloroplast genome data of H. lucidula to better understand the mutation patterns in chloroplast genomes of Huperzia. Both H. lucidula and H. serrata belong to Huperzia sect. Serratae (Rothm.) Holub and form a clade based on matK sequences showing a close phylogenetic relationship (Zhang, 2004; Ji et al., 2007). Furthermore, H. serrata is an important medicinal plant containing huperzine A, which several studies have found to be effective in the treatment of Alzheimer’s disease (Tang, 1996; Wang et al., 1998; Guo et al., 2005). Thus,
TABLE 1. Summary of Huperzia serrata and H. lucidula chloroplast features.

| Feature             | H. lucidula   | H. serrata   |
|---------------------|---------------|--------------|
| Total cpDNA size    | 154,373       | 154,176      |
| LSC                 | 104,088       | 104,080      |
| SSC                 | 19,657        | 19,658       |
| IR                  | 30,628        | 30,438       |
| Total GC content (%)| 36.3          | 36.3         |
| LSC                 | 34.4          | 34.4         |
| SSC                 | 32.8          | 32.8         |
| IR                  | 44.9          | 45.0         |
| Total no. of genes  | 119           | 120          |
| Protein encoding    | 86            | 86           |
| rRNA                | 29            | 30           |
| tRNA                | 4             | 4            |

Note: IR = inverted repeat; LSC = large single copy; SSC = small single copy.

this draft genome may not only facilitate investigations into genetic variation but also elucidate relationships within the genus to guide further exploration of compounds in closely related species.

METHODS AND RESULTS

Specimens of H. serrata were collected from Helong, Jilin Province, northeastern China. A voucher specimen (X. C. Zhang 6972) has been deposited in the Herbarium of the Institute of Botany, Chinese Academy of Sciences (PE). Total DNA was extracted with a modified cetyltrimethylammonium bromide (CTAB) method (Li et al., 2013). The DNAs were sheared into ~350-bp fragments using the Covaris M220 focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA). The NEBNext DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, Massachusetts, USA) was used for library construction. Paired-end reads of 2 × 150 bp then were generated using an Illumina HiSeq PE150 (Illumina, San Diego, California, USA). A total of 9,391,796 paired-end sequence reads of 150 bp were generated, of which 406,164 reads belong to the chloroplast genome. The chloroplast genome data were extracted using H. lucidula as a reference and assembled de novo with Geneious version R9.0.5 (Kearse et al., 2012). The first de novo assembly generated eight contigs, and the eight contigs were then extended by mapping raw reads to the contigs several times until all contigs were merged into one whole sequence of 138,841 bp. The four ends of IR regions were located through BLAST with the whole sequence itself to assemble into the complete chloroplast genome sequence. The annotation of all the genes encoding proteins, rRNAs, and tRNAs was constructed with Dual Organellar GenoMe Annotator (DOGMA; Wyman et al., 2004) and was uploaded to GenBank.

The complete chloroplast genome sequence of H. serrata (GenBank accession no. KX426071) was 154,176 bp long, 197 bp shorter than that of H. lucidula (154,373 bp; GenBank accession no. NC_006861). Both genomes had GC content of 36.3% (Table 1). A pair of IRs of 30,438 bp was separated by an LSC and a SSC of 104,080 bp and 19,658 bp, respectively, in H. serrata. The complete chloroplast genome contained 120 putative unique genes, including 86 coding genes, four rRNA genes, and 30 tRNA genes. The gene map of H. serrata is shown in Fig. 1. Based on our preliminary analysis, we found that 15 genes have one predicted intron (10 coding genes and five rRNA genes) and two coding genes have two introns (clpP and ycf3). Compared with H. lucidula, we found that the gene order and features are almost identical in genomes of H. serrata. Because comparisons between H. lucidula and other land plants have already been conducted in previous studies (Wolff et al., 2005), we did not repeat the work again. However, some unusual features also existed: an extra rRNA trnG-GAU between rrn16 and trnA-UGC in the IR region, and an intron within ycf66 in the LSC region were first annotated in H. serrata. These three genes were also annotated in the chloroplast genome sequence of another lycophyte plant, Isoetes flac-}

CONCLUSIONS

The complete genome sequence of H. serrata enables us to evaluate the genome-wide mutational events within the genus Huperzia. The genome arrangement, gene order, gene size, and GC content of H. serrata and H. lucidula are almost identical.
Three divergence hotspots (\textit{rps16-chlB}, \textit{ycf12-trnR}, and \textit{ycfI}), 18 SSRs, 27 indels, and 225 SNPs across the whole genome were identified and could provide useful phylogenetic and phylogeographic information for closely related species. Moreover, conserved primers could be designed for the highly variable regions in \textit{Huperzia} based on these two complete chloroplast genomes.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{gene_map.png}
\caption{Gene map of the \textit{Huperzia serrata} chloroplast reference genome. Genes outside of the outer circle are transcribed clockwise, whereas genes inside the outer circle are transcribed counterclockwise. The colored bars indicate different functional groups. The dashed darker gray area in the inner circle denotes GC content while the lighter gray area shows the AT content of the genome. IR = inverted repeat; LSC = large single copy; SSC = small single copy.}
\end{figure}

\section*{LITERATURE CITED}
Bock, R. 2007. Structure, function, and inheritance of plastid genomes. In R. Bock [ed.], Cell and molecular biology of plastids, 29–63. Springer, Berlin, Germany.
Dong, W., H. Liu, C. Xu, Y. Zuo, Z. Chen, and S. Zhou. 2014. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. \textit{BMC Genetics} 15: 138.
Guo, S.-L., P.-L. Li, F. Fang, H. Huang, and C.-G. Cheng. 2005. FTIR spectra-principal component analysis of phenetic relationships of *Huperzia serrata* and its closely related species. *Guangxue Yu Guangpu Fenxi* 25: 693–697.

Jansen, R. K., L. A. Raubeson, J. L. Boore, T. W. Chumley, R. C. Habelle, S. K. Wyman, A. J. Alvey, et al. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology* 395: 348–384.

Ji, S., S. Pan, J. Wang, and K. Huo. 2007. Phylogeny relationship and molecular identification of ten *Huperzia* species (Huperziaceae) based on *matK* sequences. *China Journal of Medicinal Plants* 32: 1971–1975.

Karol, K. G., K. Arumuganathan, J. L. Boore, A. M. Duffy, K. D. Everett, J. D. Hall, S. K. Hansen, et al. 2010. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: Implications for plastid phylogeny and plastid genome evolution of early land plant lineages. *BMC Evolutionary Biology* 10: 321.

Kato, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

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

Kuo, L.-Y., F.-W. Li, W.-L. Chou, and C.-N. Wang. 2011. First insights into fern *matK* phylogeny. *Molecular Phylogenetics and Evolution* 59: 556–566.

Li, F.-W., L.-Y. Kuo, C. J. Rothfels, A. Ebihara, W.-L. Chou, M. D. Windham, and K. M. Pryer. 2011. *rbcL* and *matK* earn two thumbs up as the core DNA barcode for ferns. *PLoS ONE* 6: e26597.

Li, J. S., Wang, J. Yu, L. Wang, and S. Zhou. 2013. A modified CTAB protocol for plant DNA extraction. *Chinese Bulletin of Botany* 48: 72–78.

Librado, P., and J. Rozas. 2009. *DnaSP v5: A software for comprehensive analysis of DNA polymorphism data*. *Bioinformatics (Oxford, England)* 25: 1451–1452.

Lohse, M., O. Drischel, and R. Bock. 2007. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* 52: 267–274.

Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955–964.

Li, J. M., N. Zhang, X. Y. Du, J. Wen, and D. Z. Li. 2015. Chloroplast phylogenomics resolves key relationships in ferns. *Journal of Systematics and Evolution* 53: 448–457.

Oldenkott, B., K. Yamauchi, S. Tsuei-Tsukinoki, K. Knie, and V. Knoop. 2014. Chloroplast RNA editing going extreme: More than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte Selaginella uncinata. *RNA* 20: 1499–1506.

Raubeson, L. A., and R. K. Jansen. 1992. Chloroplast DNA evidence for the ancient evolutionary split in vascular land plants. *Science* 255: 1697–1699.

Schattner, P., A. N. Brooks, and T. M. Lowe. 2005. The tRNAScan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* 33: W686–W689.

Song, Y., W. Dong, B. Liu, C. Xu, X. Yao, J. Gao, and R. T. Corlett. 2015. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Frontiers in Plant Science* 6: 662.

Stinner, D., and I. Eusvly. 2007. Imperfect SSR Finder, Version 1.0. USDA-ARS, Northwest Irrigation and Soils Research Lab, Kimberly, Idaho, USA. Website http://ssr.nwisrl.ars.usda.gov/ [accessed 14 October 2016].

Tang, X.-C. 1996. Huperzine A (shuangyiping): A promising drug for Alzheimer’s disease. *Acta Pharmacologica Sinica* 17: 481–484.

Tsuei, S., K. Ueda, T. Nishiyama, M. Hasebe, S. Yoshikawa, A. Kinagaya, T. Nishimura, and K. Yamauchi. 2007. The chloroplast genome from a lycophyte (microphyllophyte), Selaginella uncinata, has a unique structure. *Trends in Genetics* 23: 510–515.

**Table 2.** Location of simple sequence repeats in *Huperzia serrata*.

| No. | Start | End | Location | Region | Motif | No. of repeats |
|-----|-------|-----|----------|--------|-------|----------------|
| 1   | 11,242| 11,255| petB     | Intron | TA    | 7              |
| 2   | 40,488| 40,497| trnF-trnL| Intron  | AT    | 5              |
| 3   | 43,483| 43,492| rps4-trnS| Intron  | AT    | 5              |
| 4   | 56,437| 56,454| psbD-trnE| Intron  | AT    | 9              |
| 5   | 70,509| 70,522| trnK-rps16| Intron | AT    | 7              |
| 6   | 74,093| 74,104| trnQ-psbK| Intron  | TA    | 6              |
| 7   | 84,161| 84,178| atp1-rps2| Intron  | TA    | 9              |
| 8   | 85,147| 85,161| rps2-rpoC2| Intron | TAG   | 5              |
| 9   | 87,843| 87,878| rpoC2    | CDS    | TGCCTCATC | 4          |
| 10  | 93,017| 93,031| rpoC1    | CDS    | TCT   | 5              |
| 11  | 97,444| 97,457| trnC-pefN| Intron  | TA    | 5              |
| 12  | 99,074| 99,097| psbM-trnL| Intron  | TAT   | 8              |
| 13  | 99,938| 99,947| trnL-nhB | Intron  | TA    | 5              |
| 14  | 100,289| 100,298| trnL-nhB| Intron  | AT    | 5              |
| 15  | 121,906| 121,917| chlN-ycfI| Intron  | TAA   | 4              |
| 16  | 127,339| 127,350| ycf1-ycf| Intron  | ACT   | 4              |
| 17  | 130,250| 130,261| ndhA     | Intron  | AT    | 4              |
| 18  | 138,671| 138,685| rpl21-ndhF| Intron | TAA   | 5              |

Note: CDS = coding DNA sequence.
Table 3. Location of indels in the genomes of *Huperzia serrata* and *H. lucidula*.

| No. | Position | Location | Region   | Motif  | Size (bp) | Direction |
|-----|----------|----------|----------|--------|-----------|-----------|
| 1   | 5862     | rpl14-rps8 | Intergenic | A      | 1         | Insertion |
| 2   | 7064     | rpl36-rps11 | Intergenic | G      | 1         | Insertion |
| 3   | 9953     | petD      | Intron   | A      | 1         | Insertion |
| 4   | 20,555–20,556 | rpl20-rps18 | Intergenic | GG     | 2         | Insertion |
| 5   | 30,794   | psaA-accD | Intergenic | T      | 1         | Deletion  |
| 6   | 42,082–42,083 | trnL-UNA | Exon     | CC     | 2         | Insertion |
| 7   | 42,758   | psaI-accD | Intergenic | T      | 1         | Deletion  |
| 8   | 44,361–44,363 | ycf3 | Exon     | CC     | 2         | Deletion  |
| 9   | 44,986–44,989 | ycf3 | Intron   | T      | 3         | Insertion |
| 10  | 51,669–51,673 | psaB-rps14 | Intergenic | GGGG   | 5         | Deletion  |
| 11  | 51,749   | psaB-rps14 | Intergenic | G      | 1         | Deletion  |
| 12  | 58,365   | rpl20-rps18 | Intergenic | GG     | 2         | Deletion  |
| 13  | 73,875   | chlB-trnQ | Intergenic | T      | 1         | Insertion |
| 14  | 75,019   | rpl36-rps11 | Intergenic | A      | 1         | Insertion |
| 15  | 76,144   | psaA-accD | Intergenic | T      | 1         | Insertion |
| 16  | 77,128   | ycf12-trnR | Intergenic | A      | 1         | Insertion |
| 17  | 81,577   | atpF      | Intron   | A      | 1         | Deletion  |
| 18  | 85,338–85,361 | rpoC2 | CDS      | TCGGTTGCTTCCACACAGTTCCC | 24 | Deletion |
| 19  | 85,557–85,586 | rpoC2 | CDS      | TATCTAGCTTCCACTACTAGTTCTTCT | 30 | Deletion |
| 20  | 88,781–88,906 | rpoC2 | CDS      | TCATCGTCTGTCAGCTTCTTCTTCTTAAGAACAGAATCATTGATTTTTTATTATTATTATT | 126 | Deletion |

Note: CDS = coding DNA sequence.

*The plastome of *Huperzia lucidula* was used as a reference.*

Inversion, transpositions and many gene losses. *Journal of Plant Research* 120: 281–290.

**Wang, L., X. Ji, Q. Wong, and J. Yang.** 1998. Observations of 36 cases on Alzheimer’s disease treatment using Huperzine A. *Journal of Nantong Medical College* 18: 486–488.

**Wolf, P. G., K. G. Karkoci, D. F. Mandoli, J. Kuehl, K. Arumuganathan, M. W. Ellis, B. D. Meslier, et al.** 2005. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopsidaceae). *Gene* 350: 117–128.

**Wyman, S. K., R. K. Jansen, and J. L. Boore.** 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics (Oxford, England)* 20: 3252–3255.

**Zhang, L. B., and K. Iwatsuki.** 2013. Lycopodiaceae. In *Z. Y. Wu, P. H. Raven, and D. Y. Hong [eds.], Flora of China, vol. 2–3, 37–66. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.*

**Zhang, L. B.** 2004. Huperziaceae. In *C. Y. Wu [ed.], Flora Reipublicae Popularis Sinicae, vol. 6, 86–219. Science Press, Beijing, China.*

http://www.bioone.org/loi/apps
### Table 4. Comparisons of mutations, number of transitions (Ts) and transversions (Tv), and number of synonymous (S) and nonsynonymous (N) substitutions per gene of *Huperzia serrata* and *H. lucidula*.

| Gene type              | Gene   | Ts | Tv | S   | N   |
|------------------------|--------|----|----|-----|-----|
| Photosynthetic apparatus | petB   | 1  | 0  | 0   | 1   |
|                        | petD   | 1  | 0  | 0   | 1   |
|                        | petN   | 1  | 0  | 1   | 0   |
|                        | psaA   | 0  | 1  | 1   | 0   |
|                        | psaB   | 3  | 0  | 1   | 2   |
|                        | psbB   | 3  | 0  | 1   | 2   |
|                        | psbD   | 1  | 0  | 1   | 0   |
| Photosynthetic metabolism | atpA   | 1  | 0  | 0   | 1   |
|                        | atpB   | 3  | 0  | 1   | 2   |
|                        | atpE   | 1  | 0  | 1   | 0   |
|                        | atpH   | 2  | 0  | 2   | 0   |
|                        | atpI   | 2  | 0  | 0   | 2   |
|                        | ndhA   | 0  | 1  | 0   | 1   |
|                        | ndhB   | 1  | 0  | 0   | 1   |
|                        | ndhC   | 1  | 0  | 1   | 0   |
|                        | ndhF   | 2  | 2  | 0   | 4   |
|                        | ndhG   | 1  | 0  | 0   | 1   |
|                        | ndhH   | 1  | 0  | 1   | 0   |
|                        | ndhK   | 1  | 0  | 1   | 0   |
|                        | rbcL   | 2  | 0  | 2   | 0   |
| Gene expression        | rpl21  | 1  | 0  | 1   | 0   |
|                        | rpoB   | 5  | 0  | 2   | 3   |
|                        | rpoC1  | 2  | 0  | 2   | 0   |
|                        | rpoC2  | 6  | 4  | 4   | 6   |
|                        | rps11  | 1  | 0  | 1   | 0   |
|                        | rps12  | 1  | 0  | 1   | 0   |
|                        | rps7   | 2  | 0  | 0   | 2   |
|                        | rps8   | 1  | 0  | 1   | 0   |
|                        | accD   | 2  | 0  | 0   | 2   |
|                        | clpP   | 1  | 0  | 0   | 1   |
|                        | matK   | 1  | 1  | 0   | 2   |
| Other genes            | chlB   | 2  | 0  | 2   | 0   |
|                        | chlL   | 2  | 0  | 0   | 2   |
|                        | chlN   | 2  | 0  | 0   | 2   |
|                        | ycf1   | 13 | 3  | 6   | 10  |
|                        | ycf10  | 1  | 0  | 0   | 1   |
|                        | ycf2   | 4  | 5  | 2   | 7   |
| Total                  |       | 75 | 17 | 36  | 56  |