Characterization and SSR Identify of the Complete Chloroplast Genome of Paphiopedilum concolor (Orchidaceae)

Wenping Hua and Chen Chen

ABSTRACT

Paphiopedilum concolor (Orchidaceae) is endemic to China with highly ornamental and horticultural value. However, its whole chloroplast (cp) genome has not been reported yet. Here, the first nucleotide sequence of P. concolor was identified using Illumina sequencing technology. The circular genome is 166,744 bp in size and comprises a pair of inverted repeat (IR) regions of 15,694 bp, a large single-copy (LSC) region of 139,351 bp, and a small single-copy (SSC) region of 11,729 bp. The total GC content is 34.6%, whereas the corresponding values of the LSC, SSC, and IR regions are 33.7%, 26.0%, and 41.6%, respectively. The chloroplast genome contains 137 genes, including 91 protein-coding genes, eight ribosomal RNA genes, and 38 transfer RNA genes. Moreover, we identified a total of 129 microsatellites (SSRs) within the P. concolor cp genome. These newly developed genomic resources will greatly important to the further conservation genetic studies of Paphiopedilum species.

Paphiopedilum concolor (Orchidaceae), popularly known as slipper orchids, native to southeast Burma, southwest China, Thailand, and Laos. Due to destruction of their survival space and human over-exploitation for horticulture (Liu et al. 2006), P. concolor is identified as endangered class by the IUCN (International Union for Conservation of Nature, http://www.iucnredlist.org/). To facilitate its genetic research and contribute to its conservation, we assembled the cp genome of P. concolor with shotgun genome sequences obtained with the Illumina HiSeq platform. Furthermore, the microsatellite markers of P. concolor were analyzed for application of the cp sequences. These newly developed genomic data here will greatly contribute to the genetic studies of Paphiopedilum in the future.

DNA samples were extracted from the fresh leaves that were collected from a single individual of P. concolor at Qipan mountain in Kunming, Yunnan Province (25°02′35N, 102°31′52E) with CTAB method (Doyle and Doyle 1987) and stored in our lab. The whole genome shotgun sequencing of P. concolor was performed by Beijing Gene Institution (BGI, Shenzhen, China) using the Illumina HiSeq 2000 platform (Illumina, Hayward, CA). Total 67.5 M 125 bp raw reads were retrieved and trimmed using CLC Genomics Workbench v8.0 (CLC Bio, Aarhus, Denmark). A subset of 34.9 M trimmed reads were used for reconstructing the chloroplast genome by MITObim v1.8 (Hahn et al. 2013), with that of its congener Paphiopedilum armeniacum (GenBank: NC_026779.1) as the initial reference genome. A total of 156,484,607 individual chloroplast reads yielded an average coverage of 606.4-fold. The chloroplast genome was annotated in GENEIOUS R11 (Biomatters Ltd., Auckland, New Zealand).

The chloroplast genome of P. concolor is a double-stranded circular DNA molecule with 166,744 bp in size (MH191340). It comprises a pair of inverted repeat (IR) regions of 15,694 bp each, separated by a large single-copy (LSC) region of 139,351 bp, and a small single-copy (SSC) region of 11,729 bp. The total GC content is 34.6%, whereas the corresponding values of the LSC, SSC, and IR regions are 33.7%, 26.0% and 41.6%, respectively. This chloroplast genome harbors 137 functional genes, including 91 protein-coding genes (PCGs), 38 tRNA genes, and eight rRNA genes. 35 PCGs, 17 tRNA genes, and four rRNA genes are located in the forward strand, whereas others are located in the reverse strand. Among them, 38 genes are involved in photosynthesis, six genes are in substance metabolism, and 29 genes are related with self-replication. Three genes that are ndhA and ndhE are pseudogenized or lost in its cpDNA when compared with that of a close species, P. dianthum and P. armeniacum (Liu et al. 2006; Hou et al. 2017). Phylogenetic tree of 38 species within the family Orchidaceae based on the Maximum-Likelihood analysis of the 45 shared PCGs sequences using 500 bootstrap replicates.
Figure 1. Phylogenetic of 38 species within the family Orchidaceae based on the Maximum-Likelihood analysis of 45 shared PCGs sequences using 500 bootstrap replicates. The analyzed species and corresponding Genbank accession numbers are as follows: Apostasia odorata (NC_030722.1), Apostasia wallichii (NC_036260.1), Bletilla ochracea (NC_029483.1), Bletilla striata (NC_028422.1), Cattleya crispata (NC_026568.1), Cattleya liliiflora (NC_032083.1), Cymbidium aloifolium (NC_021429.1), Cymbidium ensifolium (NC_028525.1), Cypripedium formosanum (NC_026772.1), Cypripedium macranthos (NC_024421.1), Dendrobium aphyllum (NC_035322.1), Dendrobium chrysanthum (NC_035336.1), Dendrobium crepidatum (NC_035331.1), Epipactis mairei (NC_030705.1), Epipactis veratrifolia (NC_030708.1), Gastrochilus fuscopunctatus (NC_035830.1), Gastrochilus japonicus (NC_035833.1), Goodyera fumata (NC_026773.1), Goodyera procera (NC_029363.1), Habenaria pantalingiana (NC_026775.1), Habenaria radiata (NC_035634.1), Masdevallia coccinea (NC_026541.1), Masdevallia picturata (NC_026777.1), Neofinetia falcata (NC_036372.1), Neofinetia richardsoniana (NC_036373.1), Paphiopedilum armeniacum (NC_026779.1), Paphiopedilum niveum (NC_026776.1), Phalaenopsis equestris (NC_017609.1), Phalaenopsis hybrid cultivar (NC_025593.1), Phragmipedium longifolium (NC_028149.1), Pleione bulbocodioides (NC_036342.1), Sobralia callosa (NC_028147.1), Thrixspermum japonicum (NC_035831.1), Vanilla aphylloidea (NC_035320.1), and Vanilla planifolia (NC_026778.1).
using MEGA 6.0 (Tamura et al. 2013), which showed the position of *P. concolor* was situated as the sister of *P. niveum* in Orchidaceae (Figure 1).

Additionally, a total of 129 microsatellites (SSRs) were identified in the *P. Concolor* cp genome using MISA. Among them, 97 were located in the LSC regions, whereas 12 and eight were found in the IR and SSC regions, respectively. Among these SSRs, 112 were mononucleotides, seven were dinucleotides, and 10 were tetranucleotides, and more than half were composed of A or T bases. Our findings will provide a foundation for further investigation of chloroplast genome evolution in *Paphiopedilum*.

**Disclosure statement**

The authors report no conflicts of interest, and alone are responsible for the content and writing of the paper.

**Funding**

This work was supported by the Project Supported by Natural Science Basic Research Plan in Shaanxi Province of China [Program No. 2018JQ3033], the National Natural Science Foundation of China [Grant No. 31800259], the Fundamental Research Funds from Shaanxi Province of Education [No. 17JK0182, No. 2015JK1177], and Key Research Plan of Shaanxi Science and Technology Department on Agricultural Research [No. 2017NY-142].

**References**

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Hou N, Wang G, Zhu Y, Wang L, Xu J. 2017. The complete chloroplast genome of the rare and endangered herb *Paphiopedilum dianthum* (Asparagales: Orchidaceae). Conserv Genet Res. 2017:1–4.

Liu ZJ, Liu KW, Chen LJ, Lei SP, Li LQ, Shi XC, Huang LQ. 2006. Conservation ecology of endangered species *Paphiopedilum armeniacum* (Orchidaceae). Acta Ecologica Sinica. 26:2791–2799.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. Nucleic Acids Res. 41: e129–e129.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30: 2725–2729.