

The complete plastome sequence of the endangered orchid *Cymbidium macrorhizon* (Orchidaceae)

Young-Kee Kim^a, Myoung Hai Kwak^b, Myong Gi Chung^c, Hoe-Won Kim^a, Sangjin Jo^a, Jung-Yeon Sohn^a, Se-Hwan Cheon^a and Ki-Joong Kim^a

^aDivision of Life Sciences, Korea University, Seoul, Korea; ^bDepartment of Plant Resources, National Institute of Biological Resources, Incheon, Korea; ^cDivision of Life Science and the Research Institute of Natural Science, Gyeongsang National University, Jinju, Korea

ABSTRACT

In this study, we determined the complete chloroplast sequence of *Cymbidium macrorhizon* Lindl. (Orchidaceae) (NCBI acc. no. KY354040), an endangered plant species protected by the national law of Korea. The gene order and number in the *C. macrorhizon* plastome were similar to a typical Orchid plastome. The complete plastome was 149,859 bp in length and consisted of a large single copy region of 85,187 bp and a small single copy region of 13,766 bp; these were separated by two inverted repeats of 25,453 bp. The plastome contained 103 genes of which 69 were protein-coding genes, 30 were tRNA genes and four were rRNA genes. Fourteen genes contained one intron and two genes (*clpP*, and *ycf3*) had two introns. The AT content of the *C. macrorhizon* plastome was 60.0% and a total of 62 simple sequence repeat regions were identified in the plastome. Phylogenetic analysis also identified *C. lancifolium* as a closely related sister to *C. macrorhizon*, suggesting that the hemisaprophytic nature of *C. macrorhizon* is derived recently from a common leafy ancestor.

ARTICLE HISTORY

Received 16 September 2017
Accepted 6 October 2017

KEYWORDS

Plastome; *Cymbidium macrorhizon*; orchidaceae; hemisaprophyte; endangered species

Cymbidium macrorhizon Lindl., a terrestrial orchid in the genus *Cymbidium*, is native to tropical and subtropical Asia and northern Australia (Lee 2011). In nature, *C. macrorhizon* is very rare and has been designated as an endangered and protected species in Korea. The plants are less than 30 cm tall and, as they do not have any leaves, have been described as a hemisaprophytic orchid species (Yukawa and Stern 2002). The genus *Cymbidium* belongs to the subfamily Epidendroideae of the family Orchidaceae (Chase et al. 2016). To develop genetic markers for endangered *C. macrorhizon*, we obtained and analyzed its plastome sequence.

Cymbidium macrorhizon plant material was collected from its natural habitat of Wando-gun, Jeollanam-do in Korea under a collection permit from the environmental protection authority of the Korea Government. As this is a nationally endangered species, we were permitted to collect only one individual. Therefore, a representative specimen was not deposited in herbarium. But, extracted DNAs were deposited in the Plant DNA Bank in Korea (PDBK 2016-1500). A whole plant was ground into powder in liquid nitrogen and total DNA was extracted using a G-spinTM II Plant Genomic DNA extraction kit (iNtRON, Seongnam, Korea). The complete plastome sequence was generated using Illumina MiSeq (San Diego, CA) and assembled after trimming by Geneious 6.1.8 (Kearse et al. 2012). Average sequence coverage of 248 times the plastome size was obtained. Annotations were performed using the National Center for Biotechnology Information

(NCBI) BLAST and tRNAscan-SE programmes (Lowe and Eddy 1997). The complete plastome sequence was submitted to the NCBI database under the accession number KY354040.

The gene order and number in the *C. macrorhizon* plastome were similar to a typical angiosperm (Shinozaki et al. 1986; Kim and Lee 2004; Yi and Kim 2012) with the exception of the *ndh* gene family. A typical plant plastome contains 11 *ndh* genes, the *C. macrorhizon* plastome, however, contained only a single gene, *ndhC*. Ten *ndh* gene losses are not unique to *C. macrorhizon* and the losses occur commonly in other Orchidaceae plastomes (Chang et al. 2006; Wu et al. 2010; Lin et al. 2015). Therefore, the *C. macrorhizon* plastome contained 103 unique genes including 69 protein-coding genes, 30 tRNA genes and four rRNA genes. Fourteen genes had a single intron while the *clpP* and *ycf3* genes had two introns. The length of the complete plastome of *C. macrorhizon* was 149,859 bp; this was composed of a large single copy region of 85,187 bp, a small single copy region of 13,766 bp, and two inverted repeats of 25,453 bp. The *C. macrorhizon* plastome was approximately 10 kb shorter than a typical angiosperm plastome because of the losses of *ndh* gene family. The average AT content of the plastome was 60.0%. We identified a total of 62 simple sequence repeat (SSR) loci including 48 mono-SSRs, 11 di-SSRs and three tri-SSRs spread throughout the plastome.

Phylogenetic analyses were performed on a dataset that included 78 protein-coding genes (excluding *ycf1*) and four

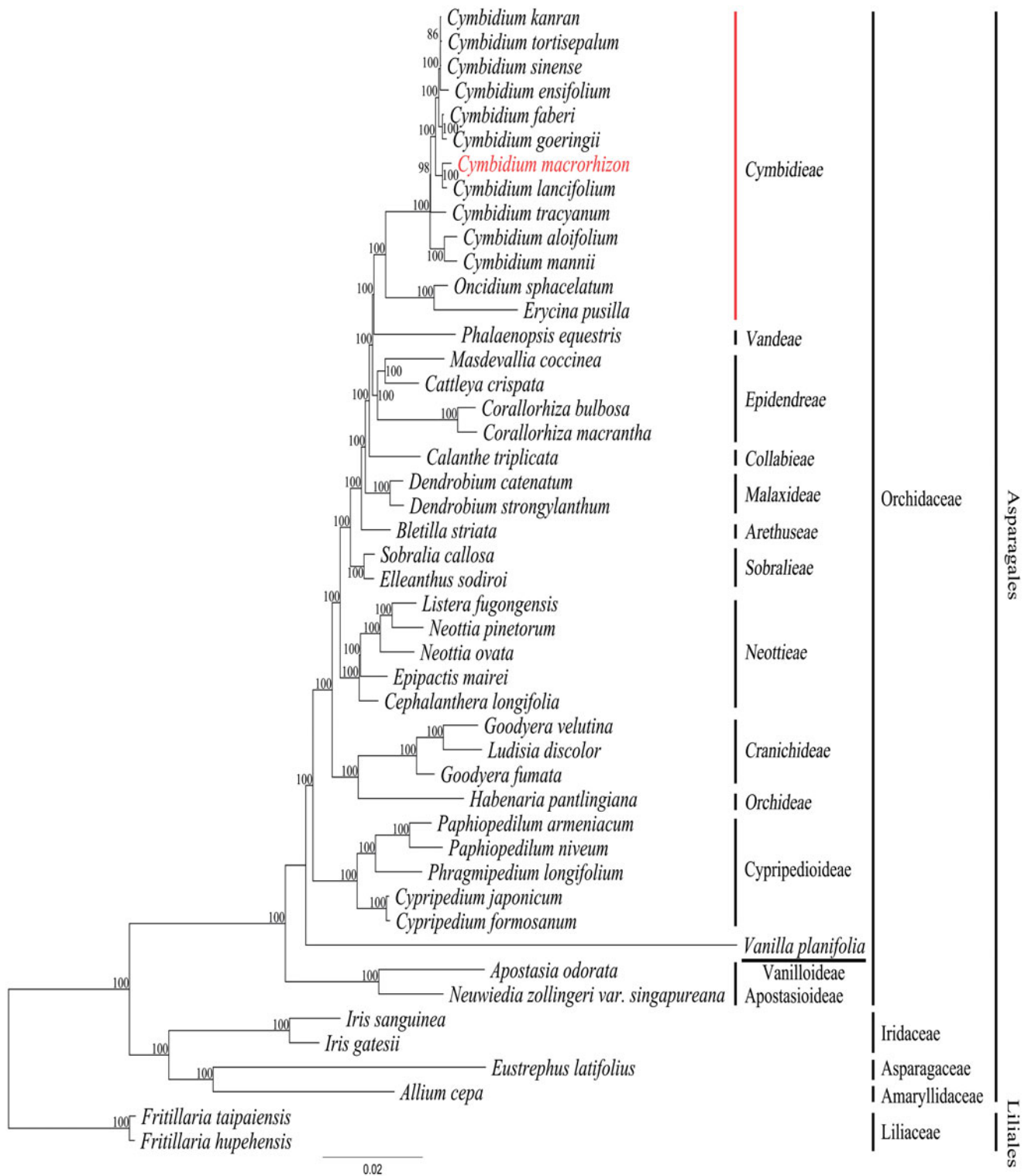


Figure 1. Chloroplast phylogenetic tree of Asparagales. A maximum likelihood tree (ML) inferred from analysis of alignment data containing 78 protein coding genes and four rRNA genes in 47 plastome sequences. The number above or below each node indicate bootstrap value. Genbank accession numbers of taxa are shown below, *Allium cepa* (NC024813), *Apostasia odorata* (NC030722), *Bletilla striata* (NC028422), *Calanthe triplicata* (NC024544), *Cattleya crispata* (NC026568), *Cephalanthera longifolia* (NC030704), *Corallorhiza bulbosa* (NC025659), *C. macrantha* (NC025660), *C. aloifolium* (NC021429), *C. ensifolium* (NC028525), *C. faberi* (NC027743), *C. goeringii* (NC028524), *C. kanran* (NC029711), *C. lancifolium* (NC029712), *C. manii* (NC021433), *C. macrorhizon* (KY354040), *C. sinense* (NC021430), *C. tortisepalum* (NC021431), *C. tracyanum* (NC021432), *Cypripedium formosanum* (NC026772), *C. japonicum* (NC027227), *Dendrobium catenatum* (NC024019), *D. strongylanthum* (NC027691), *Elleanthus sodiroi* (NC027266), *Epipactis mairei* (NC030705), *Erycina pusilla* (NC018114), *Eustrephus latifolius* (NC025305), *Fritillaria hupehensis* (NC024736), *F. taipaiensis* (NC023247), *Goodyera fumata* (NC026773), *G. velutina* (NC029365), *Habenaria pantlingiana* (NC026775), *Iris gatesii* (NC024936), *I. sanguinea* (NC029227), *Listera fugongensis* (NC030711), *Ludisia discolor* (NC030540), *Masdevallia coccinea* (NC026541), *Neottia ovata* (NC030712), *N. pinetorum* (NC030710), *Neuwiedia zollingeri* var. *singaporeana* (KM244735), *Oncidium sphacelatum* (NC028148), *Paphiopedilum armeniacum* (NC026779), *P. niveum* (NC026776), *Phalaenopsis equestris* (NC017609), *Phragmipedium longifolium* (NC028149), *Sobralia callosa* (NC028147), *Vanilla planifolia* (NC026778).

rRNA genes extracted from 46 taxa in the NCBI database and *C. macrorhizon*. The gaps for lost genes were treated as missing bases. The 82 gene sequences were aligned with MUSCLE in Geneious 6.1.8; the aligned data matrix consisted of a total of 70,692 bp. This alignment was used for phylogenetic analysis using RAxML version 7.7.1 (Stamatakis et al. 2008). An ML tree was obtained with an ML estimation value of -307032.696860 . The sister group relationship between *C. macrorhizon* and *C. lancifolium* was supported by 100% bootstrap value (Figure 1). The *C. macrorhizon*–*C. lancifolium* clade nested firmly within the *Cymbidium* group. While *C. macrorhizon* has been described as a saprophytic orchid species (Yukawa and Stern 2002), our plastome data showed an almost complete plastome gene content. The sister group relationship between the saprophytic *C. macrorhizon* and the leafy *C. lancifolium*, identified in our phylogenetic tree, suggest that the saprophytic nature are recently derived in the *C. macrorhizon* lineage (Figure 1). This tree also identified *Oncidium sphacelatum* and *Erycina pusilla* as sister genera to the monophyletic *Cymbidium* genus.

Acknowledgements

The endangered plant material was collected under the proper permit from the environmental protection authority of the Korean Government. DNA material from *C. macrorhizon* was deposited in the Plant DNA Bank of Korea (PDBK acc. no. 2016-1500).

Disclosure statement

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

Funding

This work was supported by a genome research grant for endangered species in Korea from the National Institute of Biological Resources [NIBR 201603203]. Additional funds were provided by the National Research Foundation of Korea [NRF-2015M3A9B8030588 and NRF-2015M3A9B8047398] to KJK.

References

- Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, et al. 2006. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol.* 23:279–291.
- Chase M, Christenhusz M, Fay M, Byng J, Judd W, Soltis D, Mabberley D, Sennikov A, Soltis P, Stevens P. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc.* 181:1–20.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz 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: 1647–1649.
- Kim K-J, Lee H-L. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11:247–261.
- Lee NS. 2011. Illustrated flora of Korean orchids. Seoul (South Korea): Ewha Womans University Press (in Korean).
- Lin CS, Chen JJ, Huang YT, Chan MT, Daniell H, Chang WJ, Hsu CT, Liao DC, Wu FH, Lin SY, et al. 2015. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci Rep.* 5:9040.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5:2043.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol.* 57:758–771.
- Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, Daniell H, Duvall MR, Lin CS. 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in *Oncidiinae*. *BMC Plant Biol.* 10:68.
- Yi DK, Kim KJ. 2012. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS One.* 7:e35872.
- Yukawa T, Stern WL. 2002. Comparative vegetative anatomy and systematics of *Cymbidium* (Cymbidieae: Orchidaceae). *Bot J Linn Soc.* 138:383–419.