The second complete chloroplast genome sequence of *Nymphaea alba* L. (Nymphaeaceae) to investigate inner-species variations

Jongsun Park\(^{a,b}\), Yongsung Kim\(^{a,b}\), Woohan Kwon\(^{a,b}\), Suhwan Nam\(^{c}\), and Myeoung Joon Song\(^d\)

\(^a\)InfoBoss Co., Ltd., Seoul, Republic of Korea; \(^b\)InfoBoss Research Center, Seoul, Republic of Korea; \(^c\)Baekdudaegan National Arboretum, Bonghwa-gun, Gyeongsangbuk-do, Republic of Korea; \(^d\)Nymph, Research Association of Aquatic Plants, Gwangju-si, Gyeonggi-do, Republic of Korea

**ABSTRACT**

*Nymphaea alba* L., which belongs to Nymphaeaceae, has been used as gardening plants, foods, and medicines. In this study, we completed whole chloroplast genome of *N. alba* of which length is 151,925 bp consisting four subregions: 90,001 bp of large single copy (LSC) and 19,562 bp of small single copy (SSC) regions are separated by 25,181 bp of inverted repeat (IR) regions. One hundred and twenty-eight genes (84 protein-coding genes, 8 rRNAs, and 36 tRNAs) were annotated successfully. The overall GC content of the chloroplast genome is 39.2% and those in the LSC, SSC, and IR regions are 37.8%, 34.4%, and 43.4%, respectively. Eleven single nucleotide polymorphisms and six insertion and deletions are identified between two *N. alba* chloroplast genomes. This genome will be used for developing molecular markers and for understating micro-evolutionary events of this species.

*Nymphaea alba* L., called as European white water lily, is a native species in North Africa, temperate and tropical Asia, and Europe. It has been used as gardening plants because of beautiful flowers. *Nymphaea alba* var. *rubra*, a hybrid of *N. alba* and *N. odorata* (Dkhari et al. 2011), has red flowers. In addition, it’s cooked root and seeds have utilized as foods for several thousand years (Mabey and Blamey 1972; Chiej 1984). *N. alba* contains several toxic alkaloids, such as nupharin and nymphaeine, which can affect nerve system (Wrobel 1967).

Its chloroplast genome was sequenced in 2004 to understand basal angiosperm families (Goremykin et al. 2004). Consequently, five more chloroplast genomes of *Nymphaea* have been sequenced. Here we sequenced another individual of *N. alba* deposited in the InfoBoss Cyber Herbarium (IN; IBS-00010) to dissect inner-species variations on the chloroplast genome. Total DNA was extracted from fresh leaves of *N. alba* by using a DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). Genome sequencing was performed using HiSeq2000 at Macrogen Inc., Seoul, Korea and de novo assembly was conducted by Velvet 1.2.10 (Zerbino and Birney 2008). Assembled sequences were confirmed by sequence alignment by BWA 0.7.17 (Li 2013) and SAMtools 1.9 (Li et al. 2009). Geneious R11 11.1.5 (Biomatters Ltd., Auckland, New Zealand) was used for annotation using *N. alba* chloroplast genome (NC_006050; Goremykin et al. 2004).

The chloroplast genome of *N. alba* (Genbank accession is MK040444) is 151,925 bp in length and has four subregions: 90,001 bp of large single copy (LSC) and 19,562 bp of small single copy (SSC) regions are separated by 25,181 bp of inverted repeat (IR). It contains 130 genes (85 protein-coding genes, 8 rRNAs, and 37 tRNAs). The overall GC content of *N. alba* is 39.2% and those in the LSC, SSC, and IR regions are 37.8%, 34.4%, and 43.4%, respectively.

Based on the alignment of two chloroplast genomes of *N. alba*, 11 single nucleotide polymorphisms (SNPs) and 6 insertion and deletions are identified. Three SNPs (27.3%) and one insertion (TTT; 16.7%) in our chloroplast genome are found inside *tRNA-Arg* (UCU), *psbC*, *rps4*, and *rpoC2* genes, respectively. Interestingly, four SNPs near to 3’ end of *psbA* gene are surrounded by 14 bp IRs. Number of sequence variations in *N. alba* reflects low level of genetic diversity in comparison to other studies (Young et al. 2011; Myszczynski et al. 2017).

Seven *Nymphaea* and three neighbor complete chloroplast genomes in Nymphaeaceae were used for constructing phylogenetic trees using maximum likelihood (bootstrap repeat is 1,000) and neighbor joining (bootstrap repeat is 10,000) methods with MAFFT 7.388 (Katoh and Standley 2013) and MEGA X (Kumar et al. 2018). The trees present that *N. alba* is clustered with *N. maxicana* belonging to subgenus Nymphaea. (Figure 1). In addition, *Euryale ferox* was clustered together with other *Nymphaea* chloroplast genomes, agreeing with previous studies (Figure 1). Taken together, this chloroplast genome shows inner-species sequence variations in *N. alba*, which will be used for developing molecular markers as well as for understanding micro-evolutionary events of this species.
Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by both InfoBoss Research Grant (IBG-0007).

ORCID

Jongsun Park http://orcid.org/0000-0003-0786-4701
Yongsung Kim http://orcid.org/0000-0002-5349-9226
Woochan Kwon http://orcid.org/0000-0001-8665-8097

References

Chiej R. 1984. The Macdonald encyclopedia of medicinal plants. Macdonald: Macdonald & Co (Publishers) Ltd.

Dikhar J, Kumaria S, Tandon S. 2011. Nymphaea alba var. rubra is a hybrid of N. alba and N. odorata as evidenced by molecular analysis. Annales Botanici Fennici. 48(4):317–325.

Goremykin VV, Hirsch-Ernst KI, Wölfli S, Hellwig FH. 2004. The chloroplast genome of Nymphaea alba: whole-genome analyses and the problem of identifying the most basal angiosperm. Mol Biol Evol. 21(7):1445–1454.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 35(6):1547–1549.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25(16):2078–2079.

Mabey R, Blamey M. 1972. Food for free. London: Collins.

Myszczyński K, Bączkiewicz A, Buczowska K, Ślipiko M, Szczecińska M, Sawicki J. 2017. The extraordinary variation of the organellar genomes of the Aneura pinguis revealed advanced cryptic speciation of the early land plants. Sci Rep. 7(1):9804.

Wrobel J. 1967. Nuphar alkaloids. In: R. H. F. Manske (Ed.) The alkaloids: chemistry and physiology. Elsevier; p. 441–465.

Young HA, Lanzatella CL, Sarath G, Tobias CM. 2011. Chloroplast genome variation in upland and lowland switchgrass. PLoS One. 6(8):e23980.

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