MITOCHONDRIAL DNA PART B: RESOURCES
2025, VOL. 10, NO. 3, 192–196
https://doi.org/10.1080/23802359.2025.2460782

Taylor & Francis
Taylor & Francis Group

PLASTOME REPORT

OPEN ACCESS Check for updates

Two complete chloroplast genomes of *Ceratophyllum*, an aquatic genus with unresolved phylogenetic position

Shuangyan Ru[a,b], Zhigang Wu[a], Huijun Wang[a,b], Qi Li[a] and Tao Li[a]

[a]Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China; [b]University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Ceratophyllum is an aquatic genus noted for its enigmatic position in the angiosperm phylogenetic tree. In this study, we assembled and annotated the chloroplast genomes of two species. The chloroplast genome length of *Ceratophyllum platyacanthum* subsp. *oryzetorum* (Kom.) (V.Komarov, 1988) and *Ceratophyllum submersum* L. (Carl Linnaeus, 1763) is 156,168 bp and 156,554 bp, respectively. The chloroplast genomes of *Ceratophyllum* encode 113 unique genes, including 79 protein-coding genes, four rRNA genes, and 30 tRNA genes. The assembly of these two chloroplast genomes not only contributes to our understanding of the genome of *Ceratophyllum* but also provides valuable insights for the evolutionary study of angiosperms.

ARTICLE HISTORY
Received 2 July 2024
Accepted 26 January 2025

KEYWORDS
Ceratophyllum platyacanthum subsp. *oryzetorum* (Kom.); *Ceratophyllum submersum* L.; chloroplast genome; angiosperms

Introduction

Ceratophyllum is a group of perennial submerged plants with about seven species worldwide and is a crucial component of aquatic ecosystems (Dilcher and Wang, 2009). Ceratophyllales is one of the five subclades of Mesangiospermae, with *Ceratophyllum* as the only extant genus in this order (Dilcher and Wang, 2009; Yang et al., 2020). Ceratophyllales exhibits special morphological and molecular characteristics, such as the absence of roots and special flower development, resulting in its phylogenetic placement that remains a topic of debate (Iwamoto et al., 2003; Dilcher and Wang, 2009; Iwamoto et al., 2015). Based on fossil evidence, Les classified it as a relatively basal lineage within angiosperms (Les, 1986). Based on other evidence, Ceratophyllales was once placed as a sister group to Chloranthales, monocots, and eudicots (Zanis et al., 2002; Moore et al., 2007; Maia et al., 2014; Zeng et al., 2014; Li et al., 2019; Guo et al., 2023; Hu et al., 2023). These research results have drawn the attention of many researchers in plant phylogenetics to the unresolved phylogenetic placement of Ceratophyllales.

Currently, only two complete *Ceratophyllum* chloroplast genomes are available on GenBank (*C. demersum* L., EF614270.1, and AM712908.1), which impedes the resolution of the systematic backbone of angiosperm. In this study, we assembled and annotated the complete chloroplast genomes of *C. platyacanthum* subsp. *oryzetorum* Kom. and *C. submersum* L., and it could provide additional chloroplast genome data resources for other researchers' further study. Our primary objective is to characterize the chloroplast genomes of these two *Ceratophyllum* species and investigate their placement in the angiosperm phylogeny, with the hope of contributing to the refinement of the angiosperm phylogeny framework.

Materials and methods

Materials, DNA extraction, and genome sequencing

Leaves of *C. platyacanthum* subsp. *oryzetorum* Kom. were collected from East Lake, Wuhan (30.3418°N, 114.2217°E) (Figure 1). The collected plant was successfully subjected to tissue culture and the tissue-cultured seedlings were deposited in the plant tissue culture room, at the Laboratory of Algal Genomics, Institute of Hydrobiology (http://www.ihb.ac.cn/, contact Qi Li, liqi@ihb.ac.cn) under the voucher number LAG_Cor_201507.

Total genomic DNA was extracted using the Hi-DNAsecure Plant Kit (DP350, Tiangen Biotech Co., Ltd, Beijing). Then, the purified DNA was utilized to prepare libraries with the VAHTS Universal Plus DNA Library Prep Kit (Vazyme Biotech, Nanjing) before being sequenced on the Illumina HiSeq2000 platform. In total, we obtained ~8 Gb of raw reads (PE150). The data of *C. submersum* L. was downloaded from GenBank (ERR5529394) (PE100). All raw reads were filtered using Trimmomatic (v. 0.39) (Bolger et al., 2014) with the parameters recommended by the authors.

CONTACT Qi Li liqi@ihb.ac.cn Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2025.2460782.

Figure 1. *C. platyacanthum* subsp. *oryzetorum* Kom. in the environment. The plants were collected from East Lake, Wuhan (30.3418°N, 114.2217°E) in May 2018 and this picture was taken in the laboratory.

Chloroplast genome assembly and annotation

Both chloroplast genomes were assembled using SPAdes (v3.15.5) (Bankevich et al., 2012) with different *k*-mer values: 127 for *C. platyacanthum* subsp. *oryzetorum* Kom. and 71 for *C. submersum* L. The unconnected nodes were manually removed using Bandage (v0.8.1) (Wick et al., 2015). The coverage depth map was derived from alignment data processed with the BWA (v.0.7.17-r1188), coverage calculations performed using the depth function of SAMtools (v.1.9), and visualization achieved through ggplot2 in R (Li and Durbin, 2009; 1000 Genome Project Data Processing Subgroup, 2009**).** The annotation was conducted using GeSeq (Tillich et al., 2017), utilizing the published chloroplast genomes of *C. demersum* L. as references. Additionally, all tRNAs were validated by tRNAscan-SE with default settings (Lowe and Eddy, 1997). All annotations were manually verified and edited with the aid of Geneious Prime (R9.0.2) (Kearse et al., 2012). The cis(trans)-spilicing gene structures were visualized using the CPGView (Liu et al., 2023). Finally, both genomes were visualized using OGDRAW (Greiner et al., 2019). Additionally, genomic hotspot analysis was performed using mVISTA under the Shuffle-LAGAN mode (Frazer et al., 2004). Nucleotide diversity (*Pi*) was calculated using DNAsp (v.6.12.03) (200 bp step size, 600 bp sliding window length) (Rozas et al., 2017).

Repeat and IR boundary analysis

Simple sequence repeats (SSRs) were identified using MISA (Beier et al., 2017) with specific parameters of 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats, respectively. Four types of dispersed repeats (forward, palindromic, reverse, and complement) were identified using REPuter (Kurtz et al., 2001) (Hamming Distance = 3; Minimal Repeat Size = 30 bp). Tandem repeats (TE) within plastomes were detected using Tandem Repeats Finder (v.4.09.1) (Benson, 1999). Additionally, the contraction and expansion of IR boundaries were compared using IRscope (Amiryousefi et al., 2018).

Phylogenetic analysis

To explore the phylogenetic relationship between five subclades in Mesangiospermae, we downloaded complete chloroplast genomes of 19 Mesangiospermae species from GenBank, along with two records from basal angiosperms (ANA) as outgroup (Table S1). We extracted 73 protein-coding genes with no more than two missing taxa. These protein sequences were aligned using MAFFT (v.7.508) (Katoh and Standley, 2013), and then the corresponding CDS sequences were mapped onto the alignments by pal2nal (Suyama et al., 2006). The concatenated supermatrix was used to conduct ML analyses using IQ-TREE (v.2.2.0.3) (Minh et al., 2020) with 1000 bootstrap replicates. The substitution model was set to GTR+F+I+G4. The tree was visualized with Figtree (v.1.4.3).

Results

The length of the chloroplast genomes of *C. platyacanthum* subsp. *oryzetorum* Kom. and *C. submersum* L. was 156,168 bp (~2111.36x in depth) and 156,554 bp (~230.45x), respectively (Figure 2 and Tables S1 and S2). *Ceratophyllum* chloroplast genomes encode 113 unique genes belonging to three categories, including 79 protein-coding genes, four rRNA genes, and 30 tRNA genes (Table S2 and S3). 18 genes are duplicated in IRs or at their boundaries, comprising seven protein-coding genes, four rRNA genes, and seven tRNA genes (Figure 2; Tables S2 and S3). 13 cis-splicing genes and one trans-splicing gene were identified in either of the two chloroplast genomes (Figure S2 and S3). Both mVISTA and *Pi* window analyses demonstrate that the IR regions are more conserved than the single-copy regions (Figures S4 and S5).

The numbers and types of repeat sequences in *Ceratophyllum* are highly conserved (Table S1). Among them, *C. submersum* L. possesses the highest number of repeats, followed by *C. demersum* L. (EF614270.1), *C. demersum* L. (AM712908.1), and *C. platyacanthum* subsp. *oryzetorum* Kom. Although the chloroplast genomes of *Ceratophyllum* are highly conserved in structure and genetics, there are still slight differences in the boundaries of IRs (Figure S6).

According to the phylogenetic analysis, Ceratophyllales and eudicots form a sister relationship, supported by a bootstrap (BS) value of 99, whereas monocots appear as the sister lineage to the clade comprising Ceratophyllales and eudicots (Figure 3). Within Ceratophyllales, *C. submersum* L. occupies an independent branch, while *C. platyacanthum* subsp.

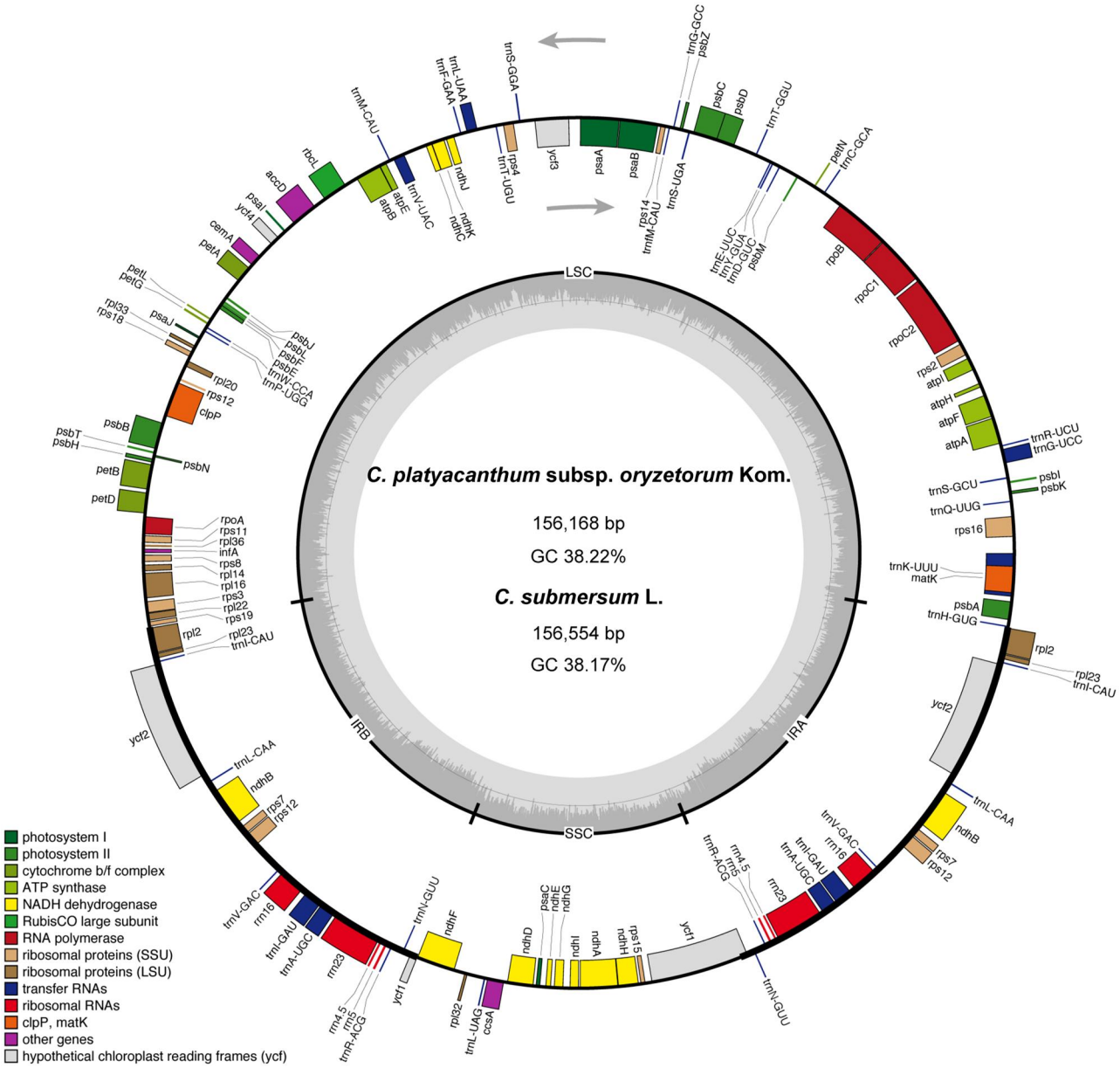

Figure 2. Circular map of the chloroplast genome of *C. platyacanthum* subsp. *oryzetorum* Kom. and *C. submersum* L. Genes depicted inside and outside the circle are transcribed counterclockwise and clockwise, respectively. Various colors denote genes with distinct functions. The dark and light grey regions in the dashed area within the inner circle represent the GC and AT contents, respectively. LSC: large single-copy region; IR: inverted repeat; SSC: small single-copy region.

oryzetorum Kom. clusters with *C. demersum* L., indicating that *C. demersum* L. is not monophyletic.

Discussion and conclusion

In this study, the chloroplast genomes of *C. platyacanthum* subsp. *oryzetorum* Kom. and *C. submersum* L. were successfully assembled and annotated. Phylogenetic analysis indicates that *Ceratophyllum* is sister to eudicots and together they are sister to monocots. In addition, *C. platyacanthum* subsp. *oryzetorum* Kom. exhibits a closer relationship with *C. demersum* L. compared to *C. submersum* L.

Previous studies have attempted to infer the systematic evolutionary history of angiosperms using various molecular markers, including chloroplast, mitochondrial, and nuclear DNA. However, the relationships among the five subclades of Mesangiospermae differ across studies (Li et al., 2021; Hu et al., 2023). Hu et al. (2023), based on 98 plastomes, and Li et al. (2021), who conducted the most extensive plastid phylogenomic analysis of angiosperms to date, both support *Ceratophyllum* as the sister group to eudicots. This places *Ceratophyllum* as a higher evolutionary branch within angiosperms, contrary to earlier conclusions based on morphology and structure (Les, 1986). However, the classification studies of *Ceratophyllum* have primarily relied on morphological and anatomical characteristics. In the future, future efforts should prioritize collecting additional *Ceratophyllum* samples to ascertain its systematic position while simultaneously confirming the species diversity at the molecular level.

Figure 3. Maximum likelihood phylogeny inferred from IQ-TREE analysis of the concatenated supermatrix of 73 plastid genes. Branch lengths are represented in terms of substitutions per site. All nodes exhibit maximal support (IQ-TREE UFBoot = 100) except where noted by red-colored numbers. In the Maximum likelihood species tree, a total of 21 species are included, with the chloroplast genomes of 19 species, excluding *C. platyacanthum* subsp. *oryzetorum* Kom. and *C. submersum* L., downloaded from NCBI. Each species is labeled with its GenBank ID after its name. These 19 species represent several major branches of angiosperms: ANA (*Nymphaea jamesoniana* NC_031826.2 and *Kadsura heteroclita* NC_057266.1), Ceratophyllales (*C. demersum* L. EF614270.1 and *C. demersum* L. AM712908.1), Eudicots (*Arabidopsis thaliana* NC_000932.1, *Populus adenopoda* NC_032368.1, *Prunus zippeliana* NC_043926.1, *Vitis vinifera* NC_007957.1 and *Solanum lycopersicum* NC_007898.3), Monocots (*Carex agglomerata* MT795185.1, *Oryza sativa* NC_031333.1, *Musa yunnanensis* NC_056834.1, *Dendrobium findlayanum* NC_058618.1 and *Acorus americanus* NC_010093.1), Magnoliids (*Piper laetispicum* NC_042254.1, *Chimonanthus nitens* NC_042745.1, *Cinnamomum tenuipile* NC_057069.1, *Annona reticulata* NC_052009.1 and *Magnolia delavayi* NC_053643.1), as well as Chloranthales (*Sarcandra glabra* NC_039621.1 and *Chloranthus erectus* NC_039627.1).

Ethical approval

The species used in this work is not protected by regulations or laws. An ethical review by the Statement Animal Experiment Committee was therefore not required.

Authors' contributions

Tao Li and Qi Li conceived and supervised the project. Zhigang Wu and Huijun Wang helped collect the samples, perform the experiments, and revise the paper. Shuangyan Ru analyzed the data and wrote the paper. All of the authors have read and approved the final manuscript and have agreed 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 National Key Research and Development Program of China [Grant No. 2020YFA0907402] and the Strategic Priority Research Program of the Chinese Academy of Sciences [Grant No. XDB31000000].

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI (https://www.ncbi.nlm.nih.gov/) under the accession number PP692210 (*C. platyacanthum* subsp. *oryzetorum* Kom.) and PP692211 (*C. submersum* L.). The associated BioProject, Bio-Sample, and SRA numbers are PRJNA1100514, SAMN40971558, and SRR28682666, respectively.

References

Amiryousefi A, Hyvönen J, Poczai P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 34(17):3030–3031. doi:10.1093/bioinformatics/bty220.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477. doi:10.1089/cmb.2012.0021.

Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics. 33(16):2583–2585. doi:10.1093/bioinformatics/btx198.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27(2):573–580. doi:10.1093/nar/27.2.573.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Dilcher DL, Wang H. 2009. An early cretaceous fruit with affinities to *Ceratophyllaceae*. Am J Bot. 96(12):2256–2269. doi:10.3732/ajb.0900049.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32(Web Server issue):W273–W279. doi:10.1093/nar/gkh458.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47(W1):W59–W64. doi:10.1093/nar/gkz238.

Guo C, Luo Y, Gao L-M, Yi T-S, Li H-T, Yang J-B, Li D-Z. 2023. Phylogenomics and the flowering plant tree of life. J Integr Plant Biol. 65(2):299–323. doi:10.1111/jipb.13415.

Hu H, et al. 2023. Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. J Integr Plant Biol. 65(2):299–323.

Iwamoto A, Izumidate R, Ronse De Craene LP. 2015. Floral anatomy and vegetative development in Ceratophyllum demersum: a morphological picture of an "unsolved" plant. Am J Bot. 102(10):1578–1589. doi:10.3732/ajb.1500124.

Iwamoto A, Shimizu A, Ohba H. 2003. Floral development and phyllotactic variation in Ceratophyllum demersum (Ceratophyllaceae). Am J Bot. 90(8):1124–1130. doi:10.3732/ajb.90.8.1124.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780. doi:10.1093/molbev/mst010.

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(12): 1647–1649. doi:10.1093/bioinformatics/bts199.

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. doi:10.1093/nar/29.22.4633.

Les DH. 1986. The evolution of achene morphology in Ceratophyllum (Ceratophyllaceae), I. Fruit-spine variation and relationships *of C. demersum, C. submersum, and C. apiculatum*. Systematic Botany. 11(4): 549–558. doi:10.2307/2419033.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics . 25(14):1754–1760. doi:10.1093/bioinformatics/btp324.

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. doi:10.1093/bioinformatics/btp352.

Li H-T, Luo Y, Gan L, Ma P-F, Gao L-M, Yang J-B, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, et al. 2021. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 19(1):232. doi:10.1186/s12915-021-01166-2.

Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants. 5(5):461–470. doi:10.1038/s41477-019-0421-0.

Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Resour. 23(3):694–704. doi:10.1111/1755-0998.13729.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5): 955–964. doi:10.1093/nar/25.5.955.

Maia VH, Gitzendanner MA, Soltis PS, Wong GK-S, Soltis DE. 2014. Angiosperm phylogeny based on 18S/26S rDNA sequence data: constructing a large data set using next-generation sequence data. Int J Plant Sci. 175(6):613–650. doi:10.1086/676675.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534. doi:10.1093/molbev/msaa015.

Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci U S A. 104(49):19363–19368. doi:10.1073/pnas.0708072104.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 34(12):3299–3302. doi:10.1093/molbev/msx248.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34SI(Web Server issue):W609–W612. doi:10.1093/nar/gkl315.

Tillich M, Lehwark 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. doi:10.1093/nar/gkx391.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 31(20):3350–3352. doi:10.1093/bioinformatics/btv383.

Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, Ma T, Zhang L, Shen X, Meng F, et al. 2020. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. Nat Plants. 6(3):215–222. doi:10.1038/s41477-020-0594-6.

Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ. 2002. The root of the angiosperms revisited. Proc Natl Acad Sci U S A. 99(10):6848–6853. doi:10.1073/pnas.092136399.

Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nat Commun. 5(1):4956. doi:10.1038/ncomms5956.