MITOCHONDRIAL DNA PART B: RESOURCES
2025, VOL. 10, NO. 1, 88–93
https://doi.org/10.1080/23802359.2024.2449090

Taylor & Francis
Taylor & Francis Group

MITOGENOME REPORT

OPEN ACCESS Check for updates

The mitochondrial genome of *Carex pseudochinensis* H. Lév. & Vaniot, an endemic sedge in Korea

Jieun Lee, Sang Chul Choi and Sangtae Kim

Department of Biology, Sungshin Women's University, Seoul, Republic of Korea

ABSTRACT

Carex pseudochinensis H. Lév. & Vaniot is an endemic species in Korea and is included in the clade of section *Paludosae* in the recent classification system. We present the complete mitochondrial genome sequence of *C. pseudochinensis* based on the POLAP pipeline with both long- and short-read sequences. The mitochondrial genome is 997,628 bp in length, containing two large regions of 536.94 and 419.04 kbp, respectively, and a pair of direct repeat regions of about 20.25 kbp. The genome contains 57 genes, including 31 protein-coding genes, 20 tRNAs, and 6 rRNAs. Phylogenetic analysis based on mitochondrial proteomes, including those from ten species of related taxa, confirmed a close phylogenetic relationship between *C. breviculmis* and *C. pseudochinensis*.

ARTICLE HISTORY
Received 26 September 2024
Accepted 27 December 2024

KEYWORDS
Cyperaceae; Oxford nanopore; Poales; POLAP; phylogenomics

Introduction

Carex L. (Cyperaceae) comprises approximately 2,000 species, representing the fifth-largest genus among angiosperms (Smith and Faulkner 1976). As an endemic *Carex* to Korea, *Carex pseudochinensis* H. Lév. & Vaniot (1902) is morphologically distinguished from other Korean sedges by its reddish-brown basal sheath of the stem and ovate perigynium having gradually narrowed beak (Korea National Arboretum 2016). The morphology of this species has traditionally placed it in section *Anomalae*, a name derived from the Latin word 'anomalous,' reflecting irregular or distinctive morphological features (Dai and Koyama 2010a, 2010b; Roalson et al. 2021). However, according to the recent classification systems (Roalson et al. 2021), this species has been reclassified to the section *Paludosae* based on molecular phylogenetic data (Jiménez-Mejías et al. 2016; Villaverde et al. 2020). Despite its endemic status, *C. pseudochinensis* has received limited scientific attention. Recent studies have reported its distribution in Heoninlleung, Seoul (Kim et al. 2010) and emphasized its critical role in maintaining biodiversity within the herbaceous plant communities of Mt. Cheongoksan, Korea (Son et al. 2014). This lack of research is also reflected in the absence of genetic and genomic data, which leaves significant gaps in our understanding of the evolution of the genus. In particular, reporting on the mitochondrial genome of an unstudied taxon will improve our understanding of its speciation, hybridization, and genome structure evolution. To date, the only reported mitochondrial genome in the genus is that of *C. breviculmis* (NC_068626). Given the considerable size of the mitochondrial genome of *C. breviculmis* (1.5 Mb) and the reported polymorphism in plant mitochondrial genomes (Gualberto et al. 2014), assembling the mitochondrial genome of *C. pseudochinensis* presents a significant challenge.

This study presents the complete mitochondrial genome of *C. pseudochinensis* based on the Plant Organelle Long-read Assembly Pipeline (POLAP) with both long- and short-read sequences. As the first report of the mitochondrial genome of this endemic species, it will significantly contribute to taxonomic and genetic studies relevant to future conservation, propagation, and ecosystem restoration of the species.

Materials and methods

We collected *C. pseudochinensis* at Wolchon-ri, Gunbuk-myeon, Haman-gun, Korea (N35.306179°, E128.320333°), and a voucher specimen deposited at the Sungshin Women's University Herbarium (SWU0036913, *Y. Cho, s. n.*; Sangtae Kim, amborella@sungshin.ac.kr) (Figure 1). The species was identified based on the recently published Korean Cyperaceae manuals, Cho et al. (2016) and Park et al. (2016). We employed a recently developed high-molecular-weight (HMW) DNA extraction method (Kang et al. 2023) for effective third-generation genome sequencing. Long-read sequencing data were obtained using the MinION platform with the R9 version flow cells and SQK-LSK109 library preparation kit (Oxford Nanopore Technologies, Oxford). Short-read sequencing was performed using the Illumina NovaSeq 6000 S4 platform with TruSeq Nano DNA Kit (Illumina, San Diego).

CONTACT Sangtae Kim amborella@sungshin.ac.kr; Sang Chul Choi sangchulchoi@sungshin.ac.kr Department of Biology, Sungshin University, Seoul 01133, Korea

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2024.2449090.

Figure 1. The morphology of *C. pseudochinensis*. (A) Habitat in Gwangneung Forest, Pocheon, Korea (photo by Minkyung Jung). (B) A representative specimen of *C. pseudochinensis* (SWU0054300) showing reddish-brown basal sheaths of the stems. (C) Ovate perigynia with a gradually narrowed beak.

The mitochondrial genome of *C. pseudochinensis* was assembled using the long-read data and polished with short-read data based on the POLAP (https://github.com/goshng/polap). The pipeline includes the extraction of mitochondrial-originated seed contigs considering three factors: (1) mitochondrial gene annotation, (2) sequence depth, and (3) graph connectivity of contigs from the long-read whole-genome assembly using Flye (v2.9.2; Kolmogorov et al. 2019). We employed the GeSeq tool (v2.03; Tillich et al. 2017) to annotate the mitochondrial genome with the mitochondrial genome of *C. breviculmis* as a reference genome (Xu et al. 2023). A circular genome map of the mitochondrial DNA was created using the OGDRAW software (v1.3.1; Greiner et al. 2019).

We performed phylogenetic analyses using a mitochondrial proteome dataset comprising six Cyperaceae taxa, including *C. pseudochinensis*, and representatives of other Poales and eudicots as outgroups based on their phylogenetic relationships (Angiosperm Phylogeny Website v14; http://www.mobot.org/MOBOT/research/APweb/) (Table S1). The Orthofinder software (v2.5.5; Emms and Kelly 2019) was employed to cluster amino acid sequences into 33 homologous groups that were aligned using the MUSCLE algorithm (v3.8.1551; Edgar 2004). The alignment matrices of the amino acid sequences were used to infer the phylogeny of the ten species employing IQ-TREE (v2.3.6; Minh et al. 2020) with the partition model (Chernomor et al. 2016) and the ultrafast bootstrap option (Hoang et al. 2018). The resulting tree and a table detailing the presence and absence of genes were illustrated using the MEGA (v11.0.13; Kumar et al. 2018) and the ggtree R package (v3.6.0; Yu et al. 2017; R core Team, 2023), respectively. Pairwise sequence alignment of the mitochondrial genomes of *C. breviculmis* and *C. pseudochinensis* was performed using the progressiveMauve software (v2.4.0; Darling et al. 2010).

Results

The circular mitochondrial genome of *C. pseudochinensis* (GenBank accession number: PP465050) was assembled with

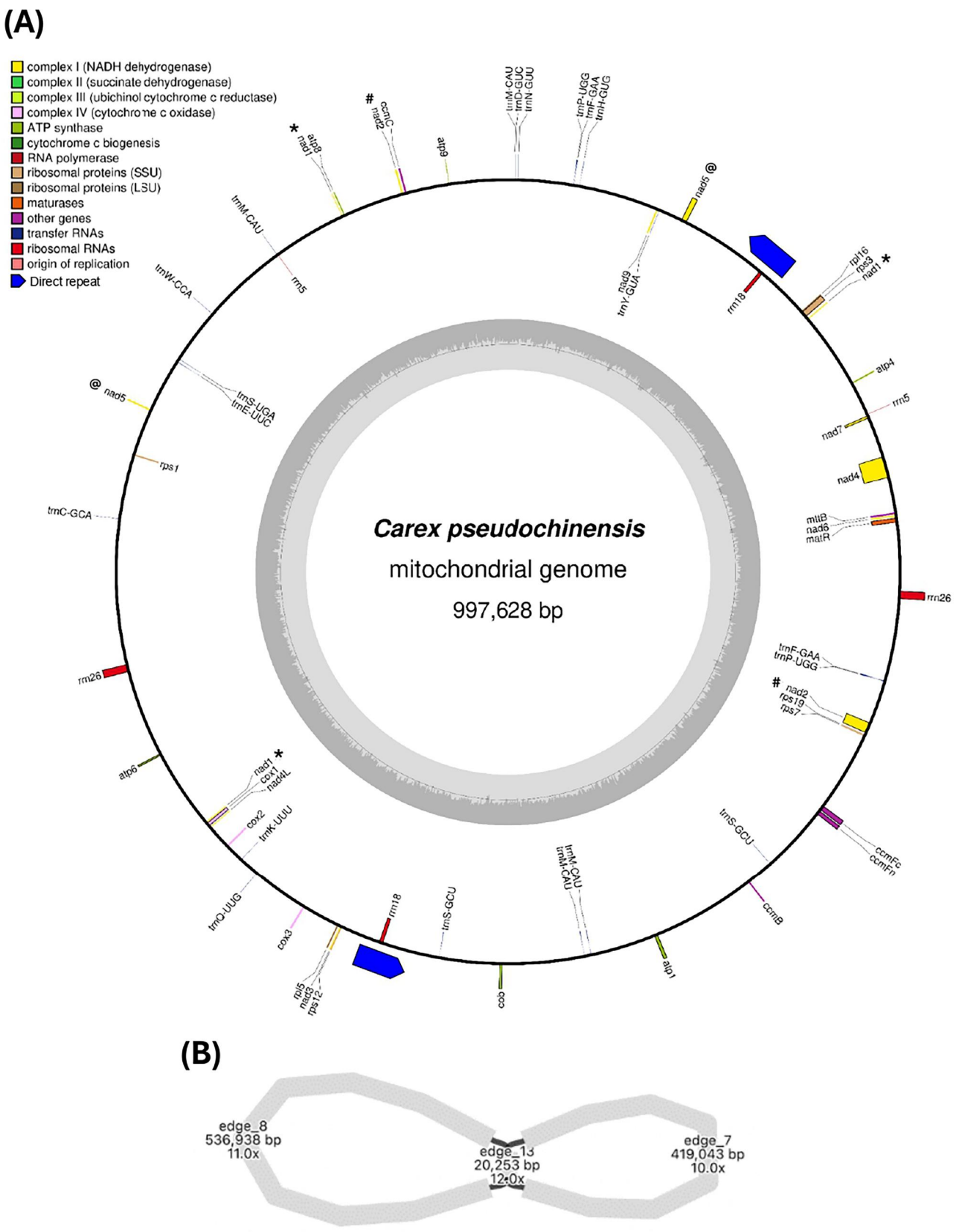

Figure 2. Circular genome map (A) and its Bandage graph (B) of the mitochondrial DNA of *Carex pseudochinensis*. Three trans-spliced genes, *nad1*, *nad2*, and *nad5*, are indicated by different special characters, *, #, and @, respectively. Blue arrows indicate direct repeats.

the genome sequencing datasets, which included approximately 42 million short reads (6.32 Gbp) and 343 thousand long reads (2.09 Gbp). The whole-genome assembly generated 423 contigs (Figure S1(A)). Among these, ten contigs were selected as seed contigs based on organelle gene annotation (Figure S1(B)). They were used as a reference for mapping and

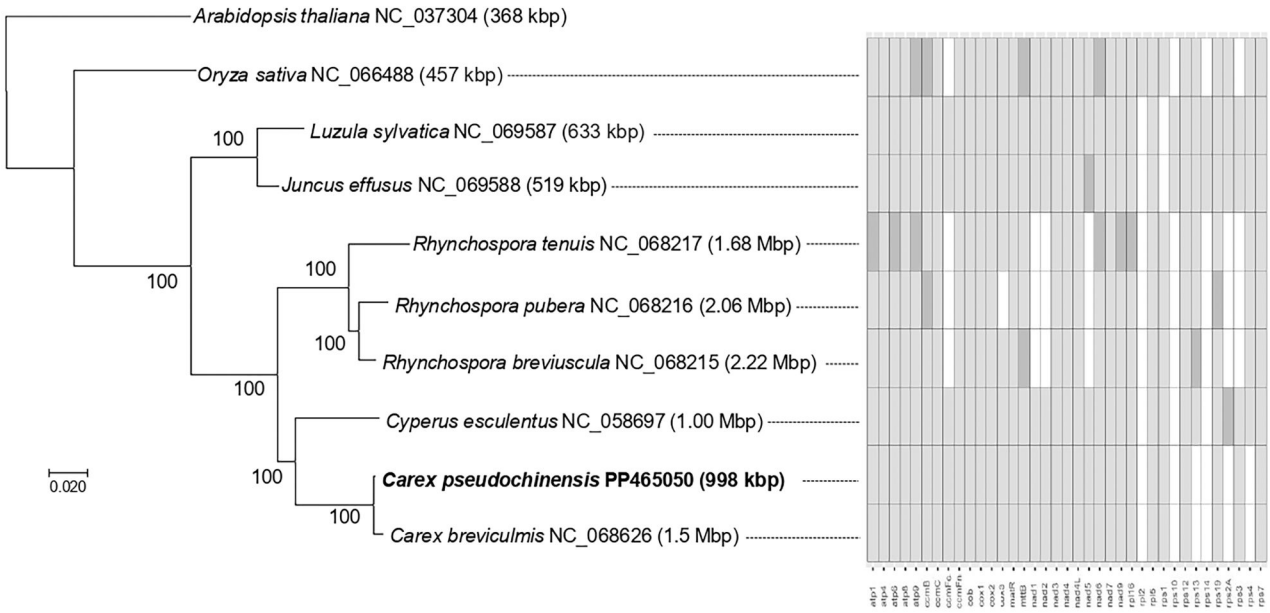

Figure 3. Maximum-likelihood tree based on amino acid sequences of 33 coding genes and presence (gray)/absence (white) table. Six Cyperaceae taxa [*Carex pseudochinensis* PP465050 (this study), *Carex breviculmis* NC_068626 (Xu et al. 2023), *Cyperus esculentus* NC_058697 (unpublished), *Rhynchospora breviuscula* NC_068215 (unpublished), *Rhynchospora pubera* NC_068216 (unpublished), *Rhynchospora tenuis* NC_068217 (unpublished)] and outgroups representing other Poales [*Juncus effusus* NC_069588 (unpublished), *Luzula sylvatica* NC_069587 (unpublished), *Oryza sativa* NC_066488 (Jiang et al. 2022)] and eudicots [*Arabidopsis thaliana* NC_037304 (Sloan et al. 2018)] are included. Numbers above the nodes indicate bootstrap support (1,000 replicates). Dark gray indicates two or more duplicated paralogous genes in each taxon, and light gray indicates a single gene. One of the duplicated genes was randomly selected for analysis. Genes shared by five or fewer taxa (*rpl2*, *rpl10*, *rpl14*, and *rps2A*) were excluded from the phylogenetic analysis. The best model for each gene is shown in Table S2.

selecting long-reads (Li 2018), which were used for organelle-genome assembly.

The organelle genome assembly generated using the Flye assembler (v2.9.2; Kolmogorov et al. 2019) in the POLAP pipeline yielded 13 contigs, four of which were found to contain mitochondrial genes based on the annotation using GeSeq (v2.03; Tillich et al. 2017) with the *C. breviculmis* mitochondrial genome of as a reference (Figure S1(C)). Of these, only three contigs (edges 7, 8, and 13) were linked and had high sequence coverage (>10×), while the other formed an independent contig (edge 6) and had low sequence coverage (4×). These three fragments were extracted as a circular DNA sequence using Bandage software (v0.8.1; Wick et al. 2015), and the resulting DNA was polished with short-read data (FMLRC v1.0.0; Mak et al. 2023; Zhou et al. 2023).

The assembled mitochondrial genome of *C. pseudochinensis* is 997,628 base pairs long, with a GC content of 41%, consisting of 28% A, 20% C, 20% G, and 32% T (Figure 2(A)). In the circular form of the mitochondrial genome, two small direct repeats (20.25 kbp) separate two large regions (536.94 and 419.04 kbp), as shown in the Bandage graph (Figure 2(B)). It contains 57 genes, including 31 protein-coding genes, six ribosomal RNAs, and 20 tRNAs. Three coding genes, NADH dehydrogenase subunits *nad1*, *nad2*, and *nad5*, were trans-spliced [Figure 2; detailed maps generated with PMGmap (Zhang et al. 2024) were in Figure S2]. *C. pseudochinensis* has the same number of trans-spliced elements in all three trans-spliced genes as *C. breviculmis* (*nad1*, 3; *nad2*, 4; *nad5*, 3). However, compared to *Carex*, *nad5* from *Cyperus esculentus* has five trans-spliced elements.

The phylogeny of nine reported monocot mitochondrial genomes was reconstructed along with that of *Arabidopsis thaliana* as an outgroup (Figure 3, Table S1). Among the six Cyperaceae species, *C. pseudochinensis* formed a clade with *C. breviculmis*. In contrast to the close phylogenetic relationship based on the protein-coding genes between these two species, the mitochondrial genome structures of these two species show remarkable differences (Figure S3).

Discussion and conclusion

The mitochondrial genome of *C. pseudochinensis* was successfully assembled using the POLAP pipeline. The complex structure and polymorphism of plant mitochondrial genomes, especially those of the Cyperaceae, present a challenge for the construction of 'master' genome sequences. Our initial attempt to assemble the *C. pseudochinensis* mitochondrial genome using the ptGAUL pipeline (v1.0.5; Zhou et al. 2023), which requires the closest genome, was unsuccessful when we used the *C. breviculmis* genome (Xu et al. 2023) as a reference (Figure S4), likely due to structural and sequence dissimilarities between the two species (Figure S3). It confirmed the prior observations that the mitochondrial genomes of Cyperaceae are challenging to assemble. The *C. pseudochinensis* mitochondrial genome generated in this study was verified by uniform sequencing coverage with short-read sequences using Minimap2 (v2.24; Li 2018) and Geneious Prime (v11.0.11; Kearse et al. 2012) (Figure S5). The mitochondrial genome of *C. pseudochinensis* will facilitate a deeper comprehension of the structural evolution of mitochondrial genomes in the genus *Carex* and provide insights into the delimitation of species boundaries.

Acknowledgments

We thank Minkyung Jung for providing the photo of *Carex pseudochinensis*.

Ethical approval

We declare that the material used in this study, *C. pseudochinensis*, is not classified as an endangered species by the Ministry of Environment of the Republic of Korea and, therefore, does not necessitate collection permits.

Author contributions

SK designed the research and revised the manuscript. JL collected the material, produced raw data, analyzed data, and prepared the initial draft of the manuscript. SCC analyzed data and prepared and revised the manuscript. All authors have reviewed and agreed upon the published version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Sungshin Women's University Research Grant of 2024 [H20240060].

ORCID

Jieun Lee http://orcid.org/0009-0003-6067-9641
Sang Chul Choi http://orcid.org/0000-0002-2340-8389
Sangtae Kim http://orcid.org/0000-0002-1821-4707

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov under accession no. PP465050. The associated BioProject, SRA (long reads and short reads), and Bio-Sample numbers are PRJNA1072305, SRR30757341, SRR30757340, and SAMN39743234, respectively.

References

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65(6): 997–1008. doi:10.1093/sysbio/syw037.

Cho YH, Kim JH, Park SH. 2016. Grasses and Sedges in South Korea. Seoul: Geobook (in Korean).

Dai LK, Koyama T. 2010a. *Carex* section Anomalae. J. Carey. In: wu ZY, Raven PH, Hong DY, eds. Flora of China. Acoraceae through Cyperaceae: cyperaceae. Beijing: science Press; St. Louis: missouri Botanical Garden Press. 23: p. 316–317.

Dai LK, Koyama T. 2010b. *Carex* section Paludosae. In: Don G, Wu ZY, Raven PH, Hong DY, editors. Flora of China. Acoraceae through Cyperaceae. Vol. 23. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press; p. 405–409.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLOS One. 5(6): e11147. doi:10.1371/journal.pone.0011147.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5(1):113. (doi:10.1186/1471-2105-5-113.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238. doi:10.1186/s13059-019-1832-y.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) v1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47(W1):W59–W64. doi:10.1093/nar/gkz238.

Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. 2014. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 100:107–120. doi:10.1016/j.biochi.2013.09.016.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522. doi:10.1093/molbev/msx281.

Jiang H, Lu Q, Qiu S, Yu H, Wang Z, Yu Z, Lu Y, Wang L, Xia F, Wu Y, et al. 2022. Fujian cytoplasmic male sterility and the fertility restorer gene OsRf19 provide a promising breeding system for hybrid rice. PNAS. 119(34):e2208759119.

Jiménez-Mejías P, Hahn M, Lueders K, Starr JR, Brown BH, Chouinard BN, Chung KS, Escudero M, Ford BA, Ford KA, et al. 2016. Megaphylogenetic specimen-level approaches to the *Carex* (Cyperaceae) phylogeny using ITS, ETS, and *matK* sequences: implications for classification. Syst Bot. 41(3):500–518. doi:10.1600/036364416X692497.

Kang M, Chanderbali A, Lee S, Soltis DE, Soltis PS, Kim S. 2023. High-molecular-weight DNA extraction for long-read sequencing of plant genomes: an optimization of standard methods. Appl Plant Sci. 11(3): e11528. doi:10.1002/aps3.11528.

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.

Kim KO, Hong SH, Lee YH, Na CS, Kang BH, Son YW. 2010. Distribution of vascular plants at the ecological landscape conservation area Heoninlleung in Seoul. Korean J Plant Resour. 23(1):60–78.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37(5):540–546. doi: 10.1038/s41587-019-0072-8.

Korea National Arboretum. 2016. Illustrated Cyperaceae of Korea. Vol. 4: Cyperaceae. Pocheon: Korea National Arboretum; p. 120–121.

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. doi:10.1093/molbev/msy096.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

Mak QXC, Wick RR, Holt JM, Wang JR. 2023. Polishing de novo Nanopore assemblies of bacteria and eukaryotes with FMLRC2. Mol Biol Evol. 40(3):msad048. doi:10.1093/molbev/msad048.

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.

Park SH, Lee YM, Kim HJ, Yang J, Jang CS, Lee KH, Lee J, Han JS, Kim HJ, Jeong K, et al. 2016. Illustrated Cyperaceae of Korea. Pocheon: Korea National Arboretum. (in Korean)

R Core Team. 2023. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roalson EH, Jiménez-Mejías P, Hipp AL, Benítez-Benítez C, Bruederle LP, Chung KS, Escudero M, Ford BA, Ford K, Gebauer S, Gehrke B, Hahn M, Hayat MQ, Hoffmann MH, Jin XF, Kim S, Larridon I, Léveillé-Bourret É, Lu YF, Luceño M, Maguilla E, Márquez-Corro JI, Martín-Bravo S, Masaki T, Míguez M, Naczi RFC, Reznicek AA, Spalink D, Starr JR, Uzma, Villaverde T, Waterway MJ, Wilson KL, Zhang SR. 2021. A framework infrageneric classification of *Carex* (Cyperaceae) and its organizing principles. J Sytemat Evolut. 59(4): 726–762. doi:10.1111/jse.12722.

Sloan DB, Wu Z, Sharbrough J. 2018. Correction of persistent errors in Arabidopsis reference mitochondrial genomes. Plant Cell. 30(3):525–527. doi:10.1105/tpc.18.00024.

Smith DL, Faulkner JS. 1976. The inflorescence of *Carex* and related genera. Bot Rev. 42(1):53–81. doi:10.1007/BF02860862.

Son H-J, Kim Y-S, Yun J-U, Chun K-W, Park W-G. 2014. The flora and vegetation structure of forest wetlands in Mt. Cheongok (Gyeongbuk Bonghwa). J Kor Forest Soci. 103(3):313–320. doi:10.14578/jkfs.2014.103.3.313.

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.

Villaverde T, Jiménez-Mejías P, Luceño M, Waterway MJ, Kim S, Lee B, Rincón-Barrado M, Hahn M, Maguilla E, Roalson EH, et al. 2020. The Global Carex Group. 2020. A new classification of *Carex* (Cyperaceae) subgenera supported by a HybSeq backbone phylogenetic tree. Bot J Linn Soc. 194(2):141–163.

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.

Xu S, Teng K, Zhang H, Wu J, Duan L, Zhang H, Wen H, Teng W, Yue Y, Fan X. 2023. The first complete mitochondrial genome of *Carex* (*C. breviculmis*): a significantly expanded genome with highly structural variations. Planta. 258(2):43. doi:10.1007/s00425-023-04169-1.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 8(1):28–36. doi:10.1111/2041-210X.12628.

Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, Liu C. 2024. Plant mitochondrial genome map (PMGmap): a software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. Mol Ecol Resour. 24(5):e13952.

Zhou W, Armijos CE, Lee C, Lu R, Wang J, Ruhlman TA, Jansen RK, Jones AM, Jones CD. 2023. Plastid genome assembly using long-read data. Mol Ecol Resour. 23(6):1442–1457. doi:10.1111/1755-0998.13787.