Structure evolution and diversification of the organelle genomes in *Carex* (Cyperaceae)

Jieun Lee

Department of Biology

The Graduate School of Sungshin University

Structure evolution and diversification of the organelle genomes in *Carex* (Cyperaceae)

A Master's Thesis

Submitted to the

Graduate School of Sungshin University

in partial fulfillment of the requirements

for the degree of Master of Biology

Jieun Lee

11, 2024

This is to certify that we have examined the

Master's Thesis Dissertation of

Jieun Lee

Submitted to Department of Biology

Approved as to style and content:

Thesis Advisor .....

Committee Chairman .....

Committee Member 이재원

The Graduate School of Sungshin University

ABSTRACT

Structure evolution and diversification of the organelle genomes in *Carex* (Cyperaceae)

Jieun Lee

Department of Biology

Graduate School of

Sungshin University

This study analyzes the chloroplast genomes of 10 *Carex* species and the mitochondrial genome of *C. pseudochinensis* to investigate the structural evolution and phylogenetic relationships within *Carex*, the largest genus in the Cyperaceae. Chapter 1 discusses the need to integrate morphological and molecular phylogenetic classifications to

address the complex taxonomy of *Carex*. With approximately 2,000 species, *Carex* presents classification challenges due to high morphological similarity among species, which limits the effectiveness of traditional classification methods. As molecular phylogenetic approaches gain recognition for their objectivity, this study emphasizes the need for a combined approach that incorporates both morphological and molecular data. In particular, the *Carex* genome contains numerous repetitive sequences that cause frequent structural rearrangements, complicating genome assembly and analysis. To accurately analyze these complex structural variations, the use of long-read sequencing technology, such as next generation sequencing (NGS), is essential. Therefore, this study employs NGS data to assemble *Carex* plant organelle genomes and compares structural variations among *Carex* and related taxa within Poales. Chapter 2 reveals that *Carex* chloroplast genomes exhibit low GC content and frequent rearrangements due to numerous short and long repeat sequences, resulting in genomic instability. Synteny analysis identified structural variations among *Carex*, while phylogenomic analysis suggests a potential association between these structural variations and certain morphological traits, such as the presence of bisexual spikes. Chapter 3 describes the assembly of the mitochondrial genome of the

Korean endemic species *C. pseudochinensis* using POLAP and its structural comparison with *C. breviculmis*, confirming the high structural plasticity of the *Carex* mitochondrial genome. These findings provide essential foundational data for future studies on *Carex* mitochondrial genomes and contribute to a deeper understanding of evolutionary adaptation and structural diversity within Cyperaceae.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The Cyperaceae (sedges), a family within the Poales, encompasses approximately 88 genera and over 5,600 species globally, making it the third-largest family among monocots (Fig. 1.1; The Angiosperm Phylogeny Group, 2016; Sokoloff et al., 2018; Govaerts et al., 2021). With its extensive range of habitats and geographical distribution, Cyperaceae holds a significant position in ecological and evolutionary studies (Simpson & Inglis, 2001; Spalink et al., 2016, 2018). The family is particularly notable for its unique adaptive traits, such as the independent origins of C4 photosynthesis pathways and the presence of holocentric chromosomes, enabling rapid adaptation to diverse environments (Besnard et al., 2009; Melters et al., 2012; Márquez-Corro et al., 2019, 2021).

Carex L., the largest genus within Cyperaceae, consists of more than 2,000 species worldwide and has been a focal point of taxonomic studies due to its remarkable morphological diversity (Hipp, 2007). Taxonomic investigations in Northeast Asia date back to early studies by Kükenthal (1909, 1935, 1936), Kreczetovicz (1935), Ohwi (1936, 1944), Akiyama (1955), and Koyama (1961, 1962). Kükenthal (1909) proposed

a comprehensive classification framework for Caricoideae, including *Carex*, based on characteristics such as the arrangement of pistillate and staminate spikelets, the presence or absence of pedicels, and the shape of the perigynium. These morphological markers became foundational for later taxonomic studies. Subsequent researchers, including Egorova (1999), Dai et al. (2010), and Hoshino et al. (2011), refined this framework to establish a modern classification system for *Carex*.

Cyperaceae is characterized by its unique inflorescence structures, which differ significantly from conventional floral arrangements and are considered adaptive responses to diverse environmental pressures. Most species within Cyperaceae exhibit unisexual flowers arranged as staminate or pistillate spikelets. Key taxonomic features include the perigynium and achene morphology, as well as leaves arranged in three ranks and stems with triangular cross-sections (Mohlenbrock & Nelson, 1999; Yang, 2014). However, convergent evolution and phenotypic variability pose challenges to taxonomic differentiation, necessitating molecular phylogenetic approaches (Naczi, 2009).

The Global *Carex* Group (2016) represents one of the earliest molecular studies on *Carex*. This study used Sanger sequencing to

analyze nuclear ribosomal ITS and ETS regions, along with the plastid *matK* gene, and included a vast dataset of 996 species, representing 50.23% of the genus. While the study proposed phylogenetic hypotheses largely consistent with previous morphological classifications, it failed to resolve certain relationships, such as those involving the Caricoid clade.

Villaverde et al. (2020) adopted the HybSeq (anchored hybrid enrichment sequencing) approach to provide greater clarity in resolving phylogenetic relationships within *Carex*. Using a Cyperaceae-specific HybSeq bait kit, they analyzed 308 nuclear exons, 543 nuclear introns, and 66 plastid exons. Villaverde et al. (2020) included 88 species of *Carex* and a total of 100 taxa, leading to the identification of six major lineages: *Siderostictae*, *Schoenoxiphium*, *Unispicate*, *Uncinia*, *Vignea*, and Core *Carex*.

Finally, Roalson et al. (2021) synthesized findings from The Global *Carex* Group (2016), Villaverde et al. (2020) and considered morphological data to propose a new classification framework, dividing *Carex* into six subgenera: *Siderosticta*, *Psyllophorae*, *Euthyceras*, *Uncinia*, *Vignea*, and *Carex* (Fig. 1.2).

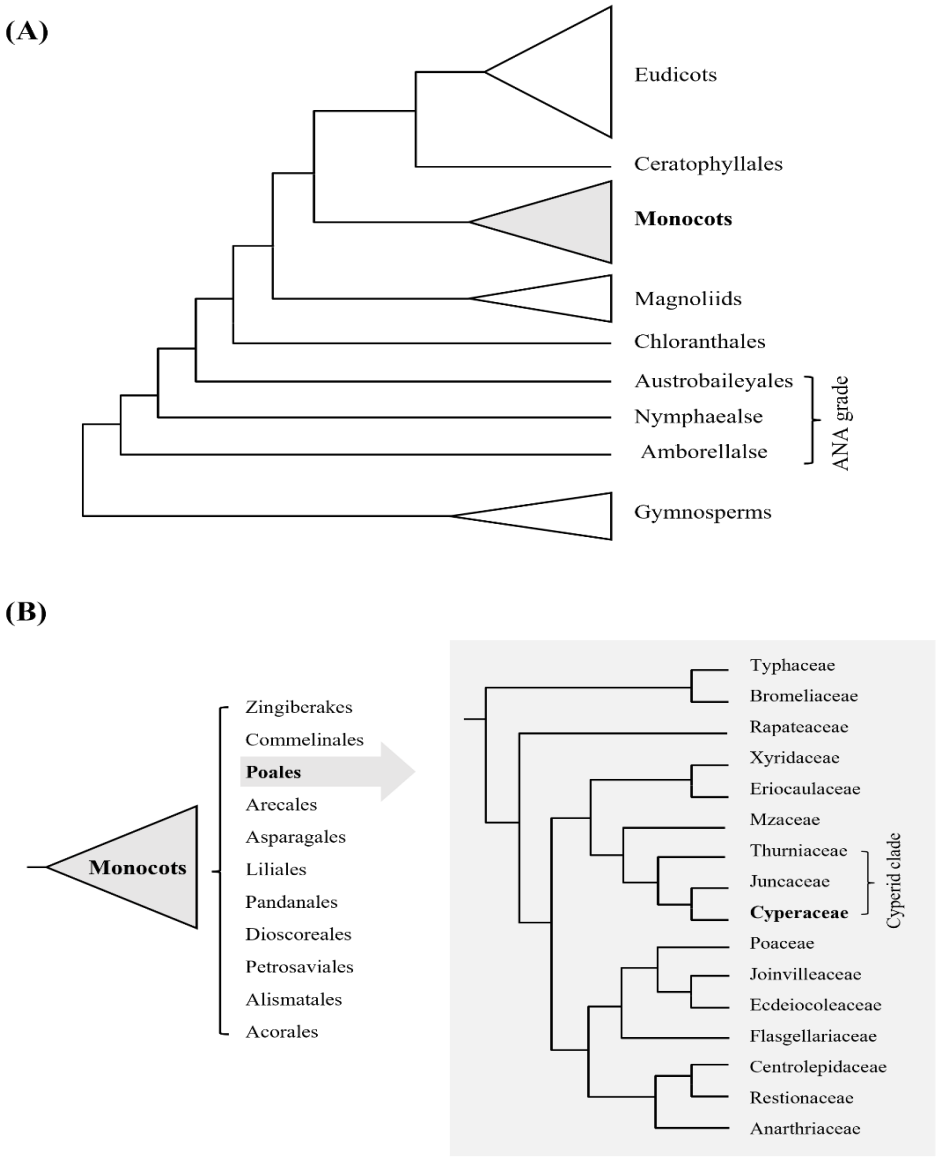


Fig. 1.1. Position of monocots in the angiosperm phylogenetic tree (A), and position of Cyperaceae in Poales (B). Topologies are summarized from The Angiosperm Phylogeny Group (2016) and Sokoloff et al. (2018).

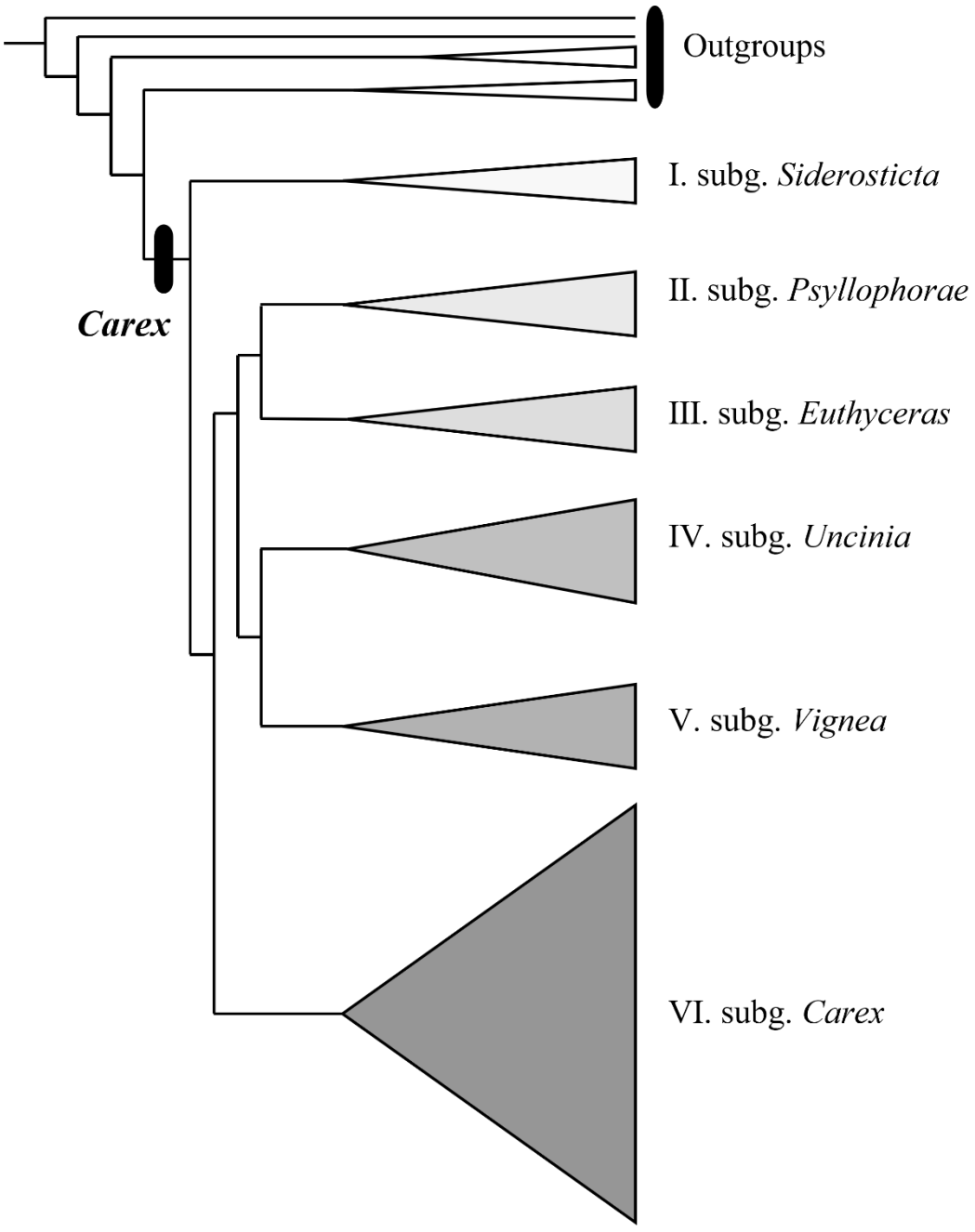


Fig. 1.2. New classification system of *Carex* based on the phylogenetic tree. Topology is summarized from Roalson et al. (2021).

In this study, the chloroplast genomes of 10 *Carex* species and the mitochondrial genome of one species were assembled and analyzed using next-generation sequencing (NGS) technology (Table 1.1). The chloroplast genomes of *Carex* are characterized by an exceptionally high abundance of repetitive sequences, which pose significant challenges during genome assembly (Wu et al., 2024). To overcome these challenges, long-read data were employed to ensure accuracy in assembly. By analyzing syntenic blocks—conserved homologous regions within the genome that include coding sequences, introns, and non-coding regions—this study integrates structural variation into phylogenetic analyses. This approach offers a broader framework for understanding evolutionary relationships within *Carex*, expanding upon previous studies that primarily focused on specific genes or intergenic regions, such as ITS and ETS. By encompassing genome-wide structural changes and phylogenetic patterns, this methodology advances research on chloroplast genomes in *Carex*, providing deeper insights into Cyperaceae and Poales.

The unique genomic features observed in *Carex* chloroplast genomes may also extend to their mitochondrial genomes. However, mitochondrial genome studies in *Carex* remain extremely limited, with

only a single species analyzed to date (Xu et al., 2023b). Seeking to bridge this research gap, this study newly analyzes the mitochondrial genome of *Carex* and explores its structural features and evolutionary relationships within Poales. These efforts aim to expand our understanding of mitochondrial genome evolution and promote further research on *Carex* and related taxa within Poales.

Table 1.1 Information on the 11 *Carex* species used in this study

Index	*Carex* subgenus [a]	Species	Collector No. / Voucher No.
1	I. subg. *Siderosticta*	*Carex siderosticta* Hance	*Y. Cho s. n.*, SWU0036866
2	V. subg. *Vignea*	*Carex paxii* Kük.	*Y. Cho s. n.*, SWU0036903
3	V. subg. *Vignea*	*Carex duriuscula* C.A.Mey	*Y. Oh & C. Lee s. n.*, SWU0001081
4	VI. subg. *Carex*	*Carex bostrychostigma* Maxim.	*H. Moon 2022-418*, SWU0060300
5	VI. subg. *Carex*	*Carex capricornis* Meinsh. ex Maxim.	*Y. Cho 2022-055,* SWU0054311
6	VI. subg. *Carex*	*Carex dickinsii* Franch. & Sav.	*Y. Cho 2022-037*, SWU0054291
7	VI. subg. *Carex*	*Carex cinerascens* Kük.	*Y. Cho 2022-045*, SWU0054302
8	VI. subg. *Carex*	*Carex laticeps* C.B.Clarke ex Franch.	*J. –E. Lee 2021-002*, SWU0039760
9	VI. subg. *Carex*	*Carex breviculmis* R.Br.	*S. Kim 2019-005*, SWU0029244
10	VI. subg. *Carex*	*Carex peiktusani* Kom.	*Y. Cho s. n.*, SWU0036865
11	VI. subg. *Carex*	*Carex pseudochinensis* H. Lév. & Vaniot [b]	*Y. Cho s. n.*, SWU0036913

a : Subgeneric classification follows Roalson et al. (2021)
b : Korean endemic species
Index 1-10: Species included in the chloroplast genome study
Index 11: Species included in the mitochondrial genome study

2. *Carex* chloroplast genome

2.1. Introduction

Chloroplasts are organelles found in plants and algae, playing a central role in the conversion of solar energy into chemical energy through photosynthesis. This function is facilitated by chlorophyll and other photosynthetic pigments, which absorb light energy and drive the photosynthetic process (Sugiura, 2003; López-Juez & Pyke, 2005). In addition to their role in photosynthesis, chloroplasts possess their own DNA, supporting the theory of endosymbiosis alongside mitochondria (Keeling, 2010). They also contribute to various metabolic pathways, including nitrogen metabolism, amino acid biosynthesis, and lipid synthesis, further emphasizing their importance in plant physiology (Daniell et al., 2016).

The chloroplast genome is generally considered more stable than nuclear and mitochondrial genomes due to its relatively small size (approximately 120–160 kb), circular structure, and quadripartite organization. This organization consists of a large single-copy region (LSC), a small single-copy region (SSC), and two identical inverted repeats (IRa, IRb) (Wicke et al., 2011; Mower & Vickrey, 2018).

Notably, the IR regions house ribosomal RNA genes (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*), approximately seven protein-coding genes (*rpl2*, *rpl23*, *ndhB*, *rps7*, *rps12*, *ycf2*, and part of *ycf1*), and several tRNA genes. These components promote homologous recombination for DNA repair and regulate gene expression, contributing to the overall stability of the chloroplast genome (Mower et al., 2018; Wu et al., 2024). IR is known as a key factor in enhancing the stability of chloroplast genomes, but it does not guarantee stability in all cases. For instance, studies on legumes and cedars have shown that IR loss induces genome rearrangements, whereas cases in families like Oleaceae and Geraniaceae (e.g., *Erodium and Pelargonium*) have reported genome rearrangements even in the presence of IRs (Palmer & Thompson, 1982; Lee et al., 2007; Blazier et al., 2016; Chumley et al., 2006; Weng et al., 2014; Zhu et al., 2016).

Instead, repetitive sequences such as short dispersed repeats (SDR) are identified as the main cause of genome instability. Chloroplast genomes with high levels of repetitive sequences are prone to increased illegitimate recombination, resulting in genome rearrangements, IR boundary shifts, and gene duplications or deletions. *Carex* chloroplast genome is characterized by a high abundance of both short dispersed repeats (SDR) and numerous long repeat sequences, contributing to a

complexity comparable to that of mitochondrial or nuclear genomes (Xu et al., 2023a). For instance, the chloroplast genome of *Carex siderosticta* contains 2,718 dispersed repeat sequences, many of which exceed 1 kb in length (Wu et al., 2024). These long repeat sequences provide homologous sequences for recombination-dependent replication, potentially inducing chloroplast heteroplasmy and significantly impacting chloroplast genome diversity and evolution. This suggests that repeat sequences are a major cause of structural instability in the chloroplast genome (Saxena et al., 2014).

In this study, 10 *Carex* species native to Korea were selected based on the latest classification system (Table 1.1; Index 1-10), and their chloroplast genomes were sequenced using NGS technology. Additionally, structural variation patterns in the chloroplast genomes of Cyperaceae within Poales were examined, and a synteny analysis was incorporated into the phylogenetic analysis to explore the evolutionary traits and genomic diversity within *Carex* and its related taxa.

2.2. Matrerials and methods

2.2.1. Plant materials

In this study, 10 *Carex* species native to Korea (Table 1.1; Index 1–10) were selected based on the new classification system established by Roalson et al. (2021), and voucher specimens were deposited at the Sungshin Women's University Herbarium (Figs. 2.1–2.10). Species identification was conducted using the recently published Korean Cyperaceae manuals (Cho et al., 2016).

Fig. 2.1. The voucher specimen of *C. siderosticta* used in this study (SWU0036866)

Fig. 2.2. The voucher specimen of *C. paxii* used in this study (SWU0036903)

Fig. 2.3. The voucher specimen of *C. duriuscula* used in this study (SWU0001081)

Fig. 2.4. The voucher specimen of *C. bostrychostigma* used in this study (SWU0060300)

Fig. 2.5. The voucher specimen of *C. capricornis* used in this study (SWU0054311)

Fig. 2.6. The voucher specimen of *C. dickinsii* used in this study (SWU0054291)

Fig. 2.7. The voucher specimen of *C. cinerasecens* used in this study (SWU0054302)

Fig. 2.8. The voucher specimen of *C. laticeps* used in this study (SWU0039760)

Fig. 2.9. The voucher specimen of *C. breviculmis* used in this study (SWU0029244)

Fig. 2.10. The voucher specimen of *C. peiktusani* used in this study (SWU0036865)

2.2.2. DNA extraction and sequencing

DNA was extracted using the High-Molecular-Weight (HMW) DNA extraction method (Kang et al., 2023), optimized for obtaining long DNA fragment. The concentration and quality of the extracted DNA were measured using Nanodrop (Thermo Fisher Scientific, Massachusetts) and Qubit (Introgen, Massachusetts) systems, and the DNA fragment length was verified using the FEMTO Pulse system (Agilent Technologies).

Long-read sequencing data were generated using the Nanopore MinION instrument with R9 version flow cells and the SQK-LSK109 library preparation kit (Oxford Nanopore Technologies, Oxford). Short-read sequencing data were generated on either the Illumina platform using the NovaSeq 6000 S4 instrument with the TruSeq Nano DNA Kit (Illumina, San Diego) or the BGISEQ platform using the MGISEQ-2000RS instrument with the MGIEasy DNA Library Prep Kit (MGI).

2.2.3. Chloroplast genome assembly and polishing

The chloroplast genome assembly of *Carex* were performed using the ptGAUL pipeline (v1.0.5; Zhou et al., 2023) with long-read sequencing data. The -g option, which specifies the *Carex* chloroplast

genome size, was set to 21000, while all other options remained at their default settings. For cases where assembly with the ptGAUL pipeline proved challenging, reassembly was conducted using the Plant Organelle-Genome Long-Read Assembly Pipeline (POLAP) (https://github.com/goshng/polap).

Following the chloroplast genome assembly using either the ptGAUL or POLAP, circular DNA sequences were extracted from the assembly graph using Bandage software (v0.8.1; Wick et al., 2015). To improve the accuracy and completeness of chloroplast genomes, the extracted draft sequences were polished with short-read data using FMLRC (v1.0.0; Mak et al., 2023; Zhou et al., 2023).

2.2.4. Chloroplast genome annotation

Chloroplast genome annotation were performed using the GeSeq tool (v2.03; Tillich et al., 2017), referencing chloroplast genome sequences from 16 *Carex* species downloaded from NCBI on February 7, 2024 (Table 2.1). Protein-coding genes were annotated using translated BLAT (Kent, 2002), while rRNA and tRNA genes were annotated with nucleotide BLAT (Kent, 2002). tRNA genes were additionally annotated using tRNAscan-SE (v2.0.7; Chan, 2019). For transfer RNA and

transfer-messenger RNA gene annotations, ARAGORN (v1.2.38; Laslett & Canback, 2004) was utilized with the setting ‘Bacterial/Plant Chloroplast.’ CDS and rRNA genes were annotated with HMMER profile searches (v3.4; Finn et al., 2011), and certain genes received further validation using the Chloë tool (v0.1.0; https://github.com/ian-small/Chloe.jl).

Table 2.1 Reference genomes used for annotation, including chloroplast genome sequences from 16 *Carex* species, downloaded from NCBI on February 7, 2024

Family	GenBank no.	Species
Cyperaceae	MT795185.1	*Carex agglomerata* C.B.Clarke
Cyperaceae	OL674104.1	*Carex alatauensis* S.R.Zhang
Cyperaceae	OP764679.1	*Carex breviculmis* R.Br.
Cyperaceae	OL674105.1	*Carex capillifolia* (Decne.) S.R.Zhang
Cyperaceae	OK539703.1	*Carex gibba* Wahlenb.
Cyperaceae	ON964521.1	*Carex giraldiana* Kük.
Cyperaceae	OL674106.1	*Carex kokanica* (Regel) S.R.Zhang
Cyperaceae	MZ846224.1	*Carex laevissima* Nakai
Cyperaceae	ON920464.1	*Carex lithophila* Turcz.
Cyperaceae	OL674107.1	*Carex microglochin* Wahlenb.
Cyperaceae	MZ962720.1	*Carex myosuroides* Vill.
Cyperaceae	KU238086.1	*Carex neurocarpa* Maxim.
Cyperaceae	OL674108.1	*Carex sargentiana* (Hemsl.) S.R.Zhang
Cyperaceae	ON920465.1	*Carex siderosticta* Hance
Cyperaceae	OK539704.1	*Carex sp.* SCSB-B-000526
Cyperaceae	ON682441.1	*Carex littledalei* (C.B.Clarke) S.R.Zhang

2.2.5. Assembly validation

To verify the completeness and accuracy of the chloroplast genome sequences, long-read and short-read data were mapped to the assembled sequences using SAMtools (Li et al., 2009), and the mapping results were visualized with Geneious Prime (v2024.0.5; Kearse et al., 2012).

2.2.6. Identification of repetitive sequences

The repeat sequences in the chloroplast genomes of *Carex* were analyzed using the web-based version of REPuter software (Kurtz et al., 2001). For the chloroplast genome, one copy of the inverted repeat (IR) region was removed before repeat analysis to avoid redundancy. REPuter identified forward, reverse, complement, and palindromic repeats within chloroplast genomes, with the maximum number of repeats set to 5,000, the upper limit of the software. The minimum repeat length was set to 30 bp, and the Hamming distance was configured to 3, ensuring over 90% sequence identity.

2.2.7. Comparison of the chloroplast genome structure in Poales

In Poales, Typhaceae occupies an early branching position, while Cyperaceae is positioned later in the phylogeny, indicating a more derived evolutionary status. Cyperaceae is also closely related to Juncaceae and is grouped with several other families within the 'Cyperid clade' (Fig. 1.1B). Based on this phylogenetic positioning, pairwise dot-plot analysis was performed using UGENE (v47.0; Okonechnikov et al., 2012) to directly visualize structural differences between chloroplast genomes and explore evolutionary changes from Typhaceae to Cyperaceae within Poales. Dot-plot analysis was performed with a minimum repeat length of 100 bp and a repeat identity threshold of 90%. To further explore the structural relationships and conserved sequences across multiple genomes, multiple alignments were performed using progressiveMAUVE (v.1.1.3; Darling et al., 2010), embedded in Geneious Prime. Prior to alignment, one inverted repeat (IR) region was removed from each chloroplast genome to avoid redundancy. The Match Seed weight was set to the default value of 15, and the Min LCB weight was adjusted to 1500 to capture major structural similarities and differences. Representative species from Typhaceae, Eriocaulaceae, and Poaceae were selected, along with three species from Juncaceae, the sister group

of Cyperaceae, were included, resulting in a total of 16 species analyzed. Species selection was based on phylogenetic relationships provided by the Angiosperm Phylogeny Website v14 (http://www.mobot.org/MOBOT/research/APweb/) (Table 2.2).

Table 2.2. Representative species selected for chloroplast genome comparison in Poales (16 species)

Family	GenBank no.	Species
Typhaceae	NC_013823	*Typha latifolia* L.
Eriocaulaceae	NC_031333	*Eriocaulon buergerianum* Körn.
Poaceae	NC_042211	*Oryza sativa* L.
Juncaceae	NC_071363	*Juncus roemerianus* Scheele
Juncaceae	NC_071364	*Juncus validus* Coville
Juncaceae	PP790565	*Juncus fauriei* H.Lév. & Vaniot
Cyperaceae	PP790566	*Carex siderosticta* Hance
Cyperaceae	PP790556	*Carex bostrychostigma* Maxim.
Cyperaceae	PP790557	Carex breviculmis R.Br.
Cyperaceae	PP790558	*Carex capricornis* Meinsh. ex Maxim.
Cyperaceae	PP790559	*Carex cinerascens* Kük.
Cyperaceae	PP790560	*Carex dickinsii* Franch. & Sav.
Cyperaceae	PP790561	*Carex duriuscula* C.A.Mey
Cyperaceae	PP790562	*Carex laticeps* C.B.Clarke ex Franch.
Cyperaceae	PP790563	*Carex paxii* Kük.
Cyperaceae	PP790564	*Carex peiktusani* Kom.

2.2.8. Phylogenomic and synteny analysis of *Carex* chloroplast genomes

The amino acid sequences of the chloroplast genomes from a total of 16 species (Table 2.2), including six representative species within Poales, were clustered into 79 homologous groups using Orthofinder software (v2.5.5; Emms & Kelly, 2019). These clusters were subsequently aligned with the MUSCLE algorithm (v3.8.1551; Edgar, 2004).

The aligned matrix was used to infer the phylogeny of the 16 species with IQ-TREE (v2.3.6; Minh et al., 2020), applying the partition model (Chernomor et al., 2016) and ultrafast bootstrap options (Hoang et al., 2018). The final phylogenetic trees were visualized in MEGA (v11.0.13; Kumar et al., 2018), and the gene presence/absence matrix was generated using the ggtree R package (v3.6.0; Yu et al., 2017; R Core Team, 2023).

Additionally, 25 locally collinear block (LCBs) sequences were extracted using progressiveMauve (v.1.1.3; Darling et al., 2010), with *C. siderosticta* designated as the reference chloroplast genome. The minimum LCB length was set to 1000 bp, and the Min LCB weight was set to 1500. The LCB sequences were then aligned with MUSCLE (v3.8.1551; Edgar, 2004) and used to infer phylogenetic relationships of

the 16 species with IQ-TREE (v2.3.6; Minh et al., 2020), where the bootstrap value was set to 1000. The final phylogenetic trees were visualized in MEGA (v11.0.13; Kumar et al., 2018).

2.3. Results

2.3.1. DNA sequencing

In this study, 10 species of *Carex* were sequenced using Oxford Nanopore Technologies (ONT) and either Illumina or MGI platforms. ONT sequencing generated a total read base range of 0.08 to 77.41 Gbp, with variations in read counts across species. The maximum read length reached 292,724 bp, and N50 values ranged from 7,223 to 46,449 bp (Table 2.3). The Illumina/MGI datasets, obtained with a 500 bp insert size and generating 150 bp paired-end reads, provided high-quality short-read data, with an average total read base of approximately 4.26 Gbp, enhancing assembly accuracy.

Table 2.3. Basic statistics of DNA sequencing data

Species	Sequencing methods	Total read bases (Gbp)	Reads	Min length (bp)	Mean length (bp)	Max length (bp)	N50 (bp)
C. siderosticta	ONT	7.36	2262303059	1	7,111.30	214,344	18,158
	Illumina	2.26	15000000/15000000	N/A	N/A	N/A	N/A
C. bostrychostigma	ONT	1.77	26593	1	6,845.09	137,374	27,585
	MGI	5.1	34544790/34586660	N/A	N/A	N/A	N/A
C. breviculmis	ONT	46.6	9943944	1	4,686.25	219,088	13,036
	Illumina	3.01	20000000/20000000	N/A	N/A	N/A	N/A
C. capricornis	ONT	6.57	896161	1	7,326.36	164,765	12,780
	MGI	19.35	130895521/131133736	N/A	N/A	N/A	N/A
C. cinerascens	ONT	0.08	22492	1	3,490.23	111,847	16,042
	MGI	5.09	34205257/34239784	N/A	N/A	N/A	N/A
C. dickinsii	ONT	23.21	3534130	1	6,594.25	135,017	13,778
	Illumina	3.02	20000000/20000000	N/A	N/A	N/A	N/A
C. duriuscula	ONT	1.85	804988	14	2,361.19	122,679	7,223
	MGI	4.94	33141972/33287065	N/A	N/A	N/A	N/A
C. laticeps	ONT	13.06	163673	50	8,542.97	152,483	22,447
	Illumina	3.37	22363744/22363744	N/A	N/A	N/A	N/A
C. paxii	ONT	77.41	2117214	1	24,375	292,724	46,449
	Illumina	10.16	67271251/67271251	N/A	N/A	N/A	N/A
C. peiktusani	ONT	1.42	271930	18	5,408.90	136,888	11,585
	Illumina	3.56	23561702/23561702	N/A	N/A	N/A	N/A

* N/A: Illumina/MGI sequencing with a 500 bp insert size and generated 150 bp paired-end reads.

2.3.2. Chloroplast genome assembly and annotation

The assembled chloroplast genomes of *Carex* species ranged in size from 185,213 bp (*C. bostrychostigma*) to 241,458 bp (*C. peiktusani*), while the outgroup species exhibited sizes from 134,502 bp (*O. sativa*) to 196,852 bp (*J. roemerianus*) (Table 2.4). These size differences are attributed to variations in the lengths of the LSC (Large Single Copy), SSC (Small Single Copy), and IR (Inverted Repeat) regions among species.

The LSC region in *Carex* ranged from 97,568 bp (*C. bostrychostigma*) to 115,594 bp (*C. peiktusani*), while in the outgroup species, it varied from 80,547 bp (*O. sativa*) to 91,055 bp (*J. fauriei*). The SSC region showed relatively less variation, ranging from 8,491 bp (*C. paxii*) to 10,302 bp (*C. capricornis*) in *Carex* and from 2,046 bp (*J. validus*) to 19,652 bp (*T. latifolia*) in the outgroup species. The IR region in *Carex* ranged from 36,494 bp (*C. paxii*) to 57,823 bp (*C. peiktusani*), while in the outgroup, it varied from 20,804 bp (*O. sativa*) to 53,003 bp (*J. roemerianus*) (Table 2.4).

The GC content in the outgroup species ranged from 33.5% (*O. sativa*) to 36.6% (*T. latifolia*), while *Carex* species exhibited slightly lower GC content, ranging from 32.7% (*C. peiktusani*) to 34.2% (*C.*

cinerascens and *C. duriuscula*) (Table 2.4). This relatively lower GC content in *Carex* may be associated with the stability of the chloroplast genome, suggesting a potential for structural variation.

In terms of gene composition, *Carex* species contain a total of 132 to 148 genes, including 85 to 97 protein-coding genes (CDS), 37 to 46 tRNA genes, and 8 rRNA genes. In contrast, the outgroup species generally have a simpler gene composition (Table 2.4).

Circular maps of all assembled chloroplast genomes were generated using the OGDRAW software (v1.3.1; Greiner et al., 2019; Figs. 2.11-2.20).

Table 2.4. Statistical summary of features of the chloroplast genomes of 10 *Carex* and related taxa

	Typha latifolia (Typhaceae)	*Eriocaulon buergerianum* (Eirocaulaceae)	*Oryza sativa* (Poaceae)	*Juncus validus* (Juncaceae)	*Juncus roemerianus* (Juncaceae)	*Juncus fauriei* (Juncaceae)
GenBank accession no.	NC_013823	NC_042211	NC_031333	NC_071634	NC_071363	PP790565
Genome size (bp)	161,572	151,434	134,502	147,183	196,852	185,388
LSC length (bp)	89,140	81,534	80,547	87,125	82,944	91,055
SSC length (bp)	19,652	17,114	12,347	2,046	7,902	7,529
IR length (bp)	26,390	26,393	20,804	28,961	53,003	43,402
Gene	133	133	132	114	136	135
CDS	86	88	84	68	87	91
tRNA	38	37	40	38	41	36
rRNA	8	8	8	8	8	8
GC content (%)	36.6	35.7	33.5	34.7	35.2	35.9

Table 2.4. (Continued)

	Carex siderosticta	*Carex bostrychostigma*	*Carex breviculmis*	*Carex capricornis*	*Carex cinerascens*	*Carex dickinsii*	*Carex duriuscula*	*Carex laticeps*	*Carex paxii*	*Carex peiktusani*
GenBank accession no.	PP790566	PP790556	PP790557	PP790558	PP790559	PP790560	PP790561	PP790562	PP790563	PP790565
Genome size (bp)	195,239	185,213	213,647	205,950	212,327	208,924	188,436	210,898	185,458	241,458
LSC length (bp)	102,454	97,568	102,256	100,890	110,167	106,331	102,724	107,518	103,979	115,594
SSC length (bp)	9,011	8,971	8,863	10,302	10,198	10,287	8,616	8,784	8,491	10,218
IR length (bp)	41,887	39,337	51,264	47,379	45,981	46,153	38,548	47,298	36,494	57,823
Gene	141	140	145	143	148	145	141	140	132	145
CDS	96	92	93	92	96	93	93	86	85	97
tRNA	37	40	43	41	44	43	40	46	39	40
rRNA	8	8	8	8	8	8	8	8	8	8
GC content (%)	34.1	33.7	33.5	33.8	34.2	33.8	34.2	33.4	33.9	32.7

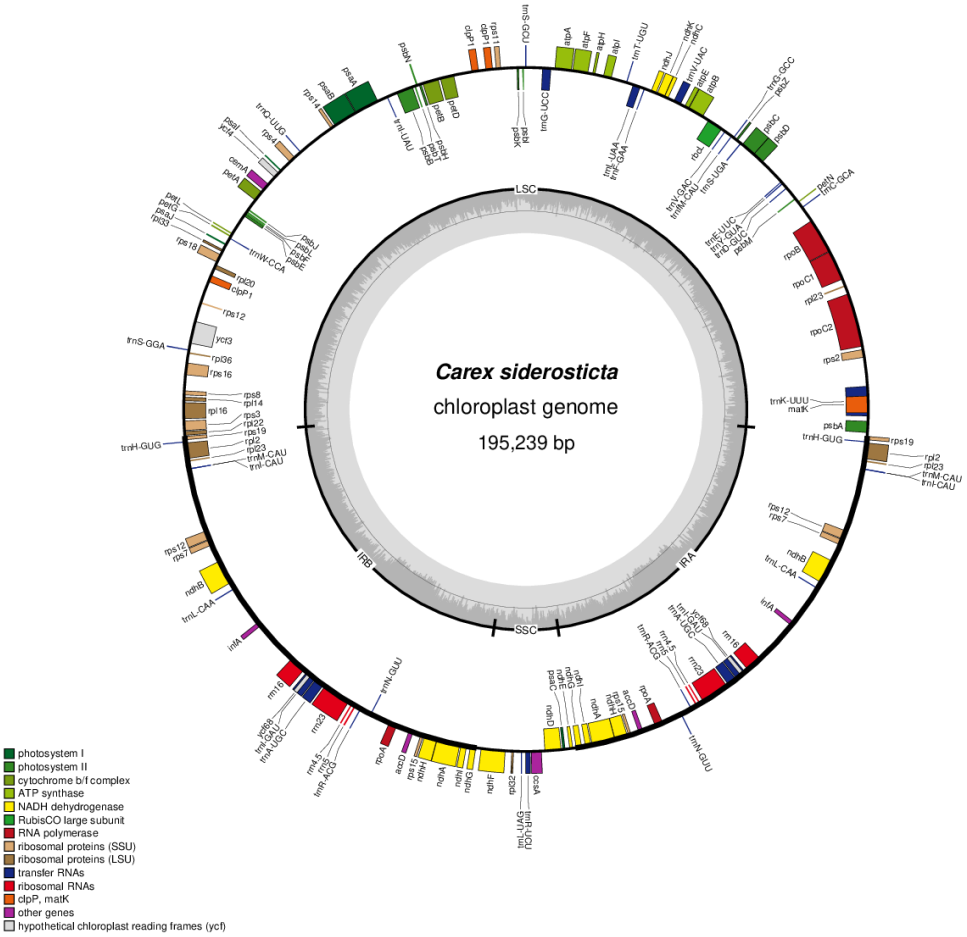


Fig. 2.11. Gene map of the chloroplast genome in *C. siderosticta* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

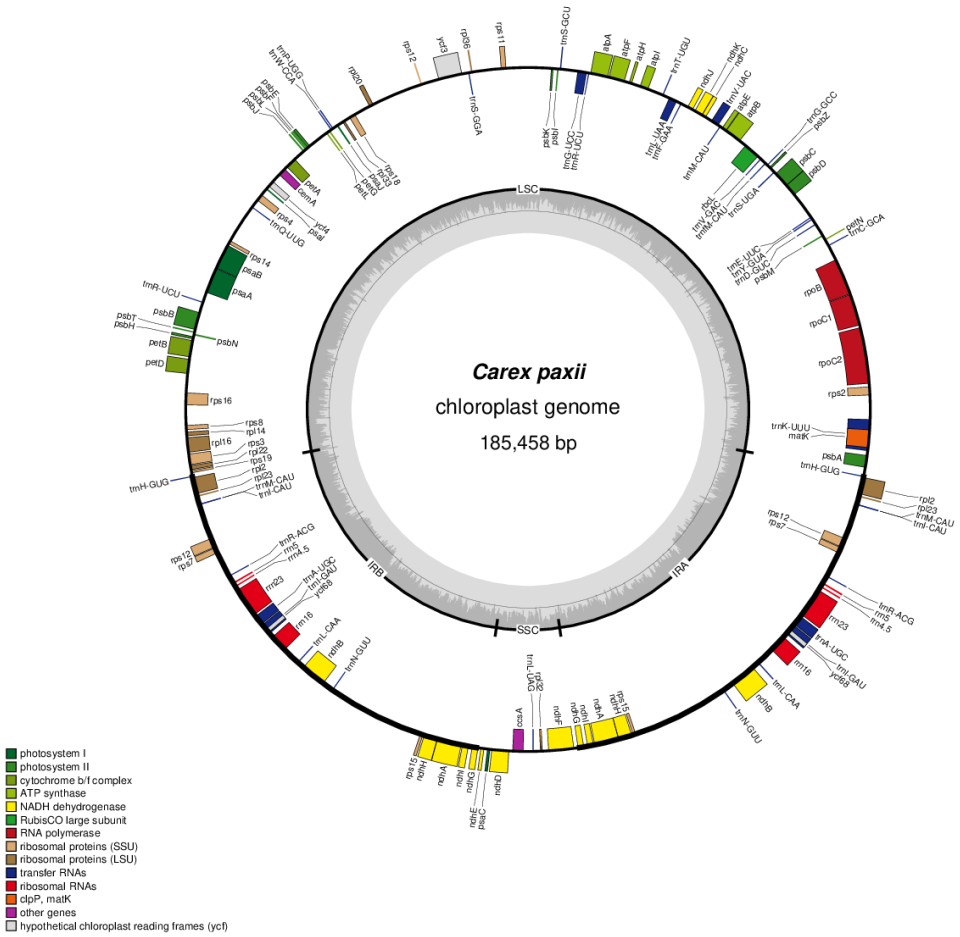


Fig. 2.12. Gene map of the chloroplast genome in *C. paxii* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

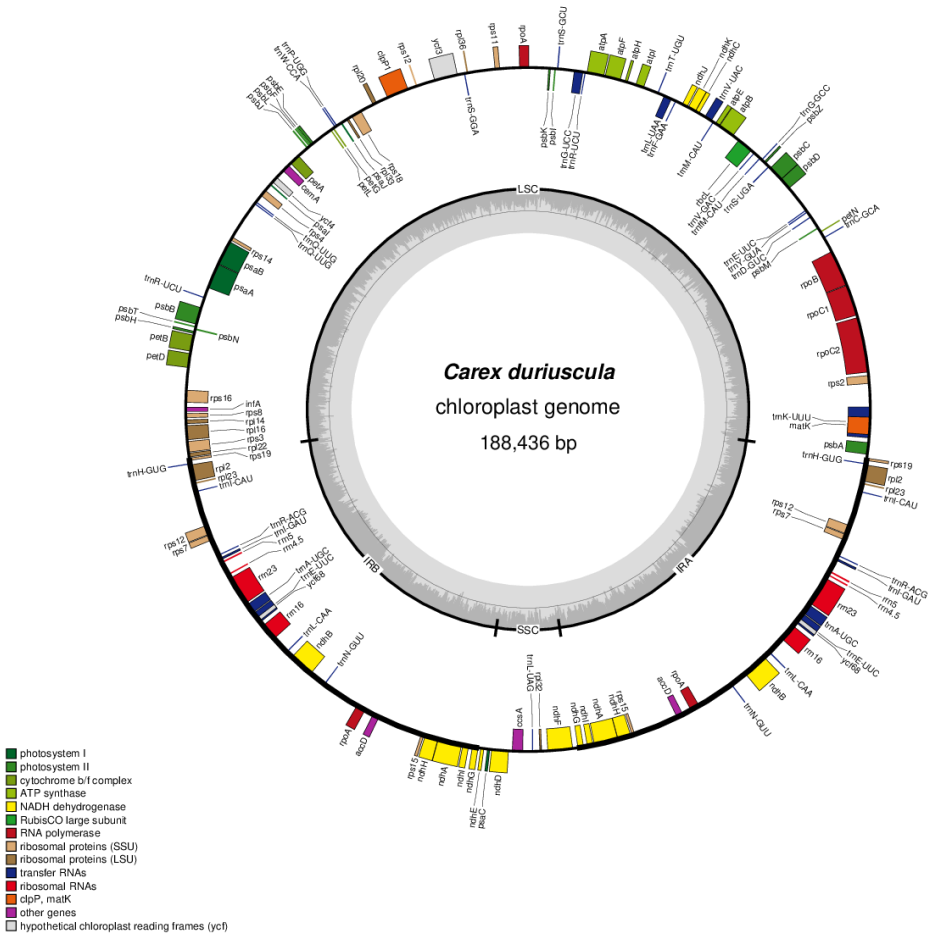


Fig. 2.13. Gene map of the chloroplast genome in *C. duriuscula* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

Fig. 2.14. Gene map of the chloroplast genome in *C. bostrychostigma* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

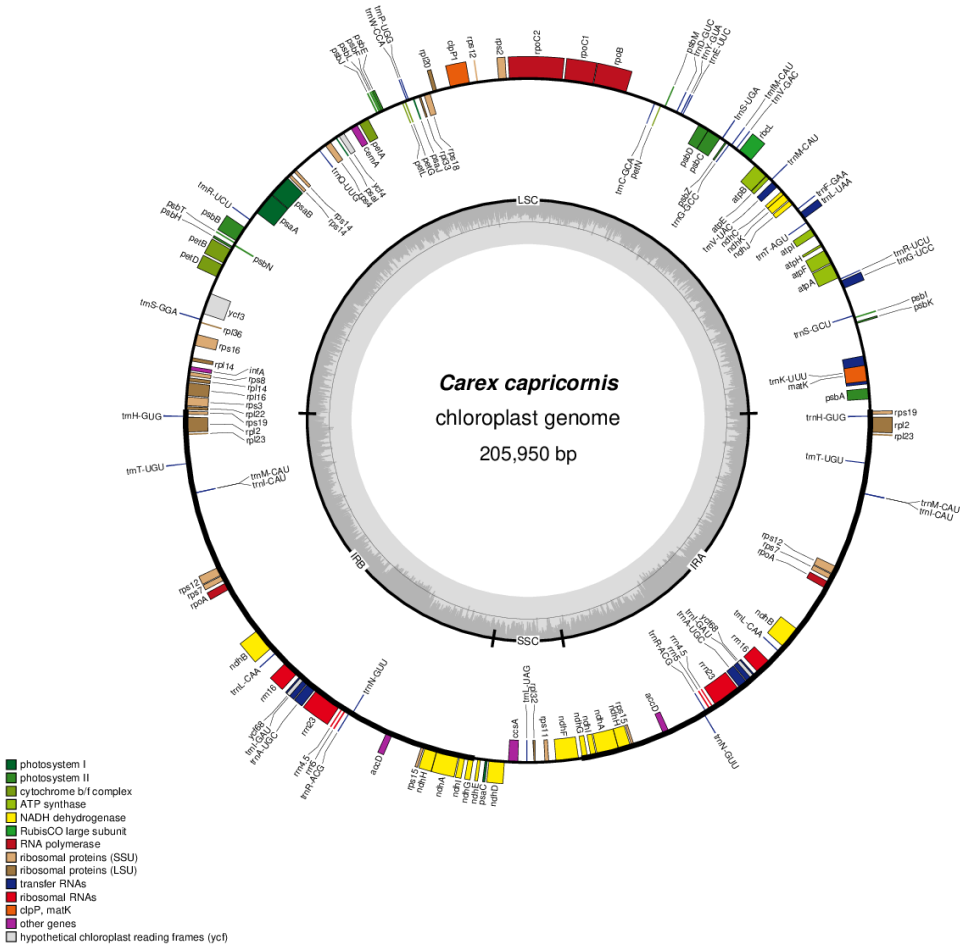


Fig. 2.15. Gene map of the chloroplast genome in *C. capricornis* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

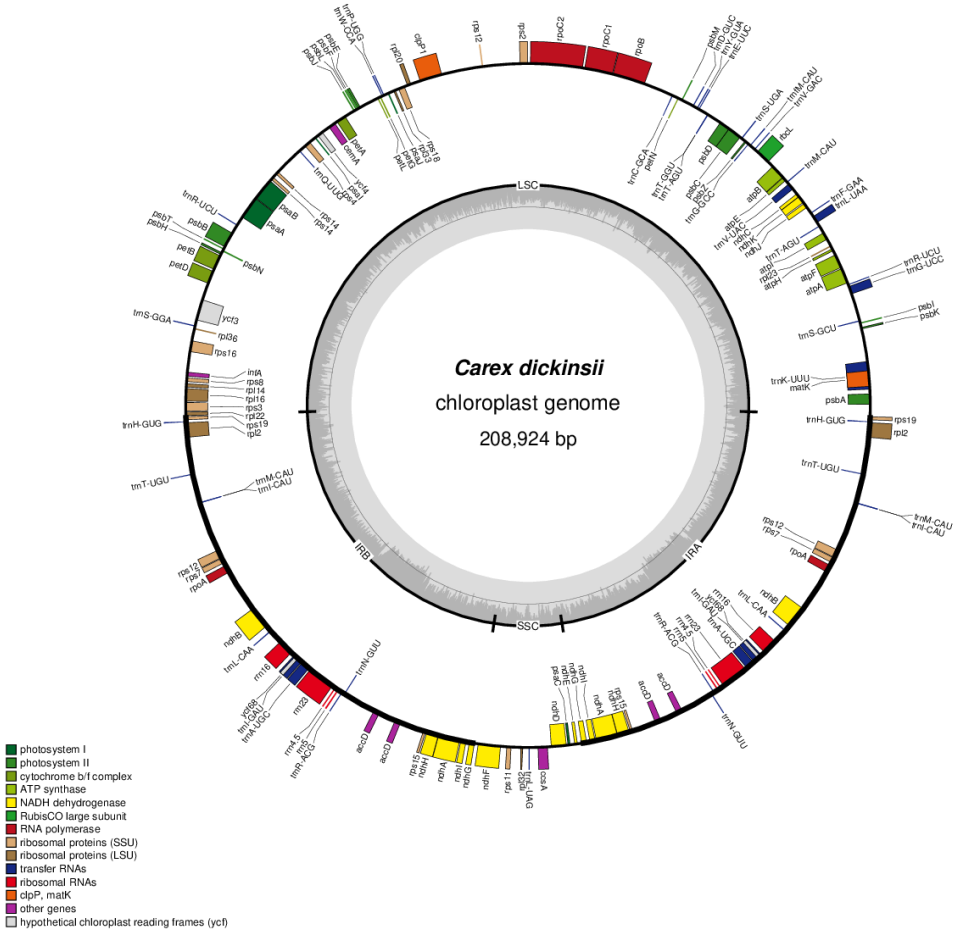


Fig. 2.16. Gene map of the chloroplast genome in *C. dickinsii* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

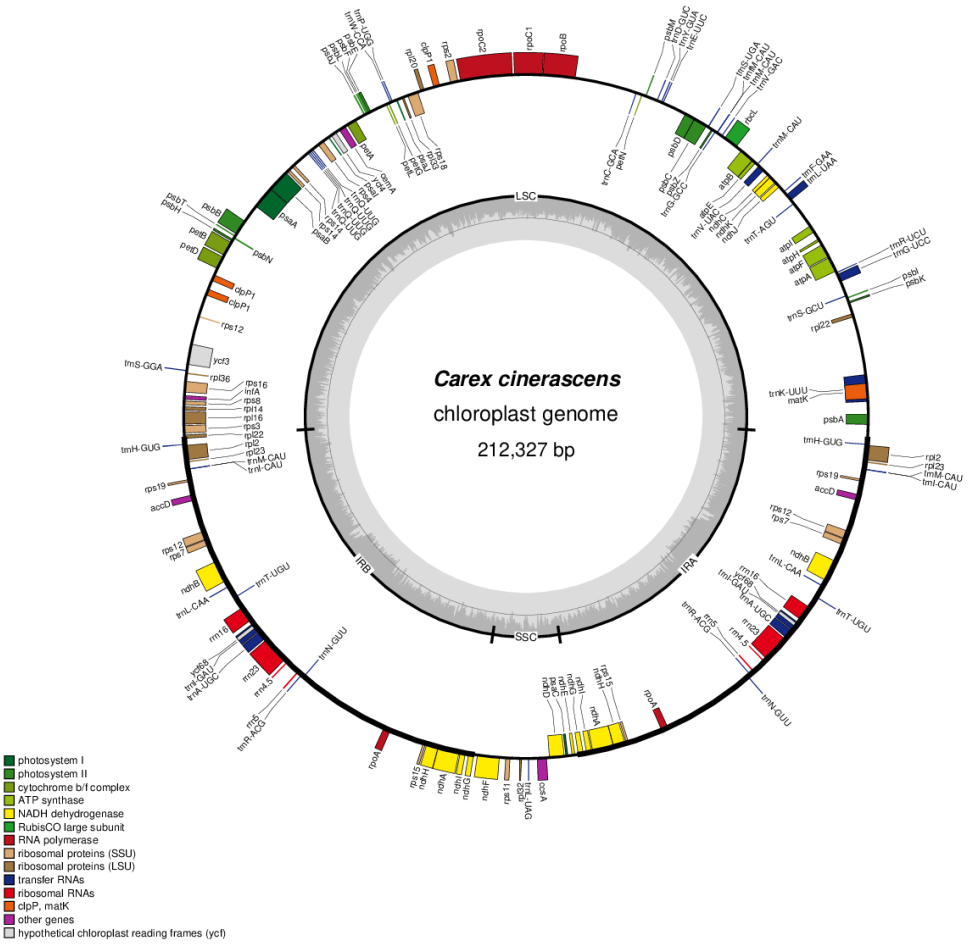


Fig. 2.17. Gene map of the chloroplast genome in *C. cinerascens* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

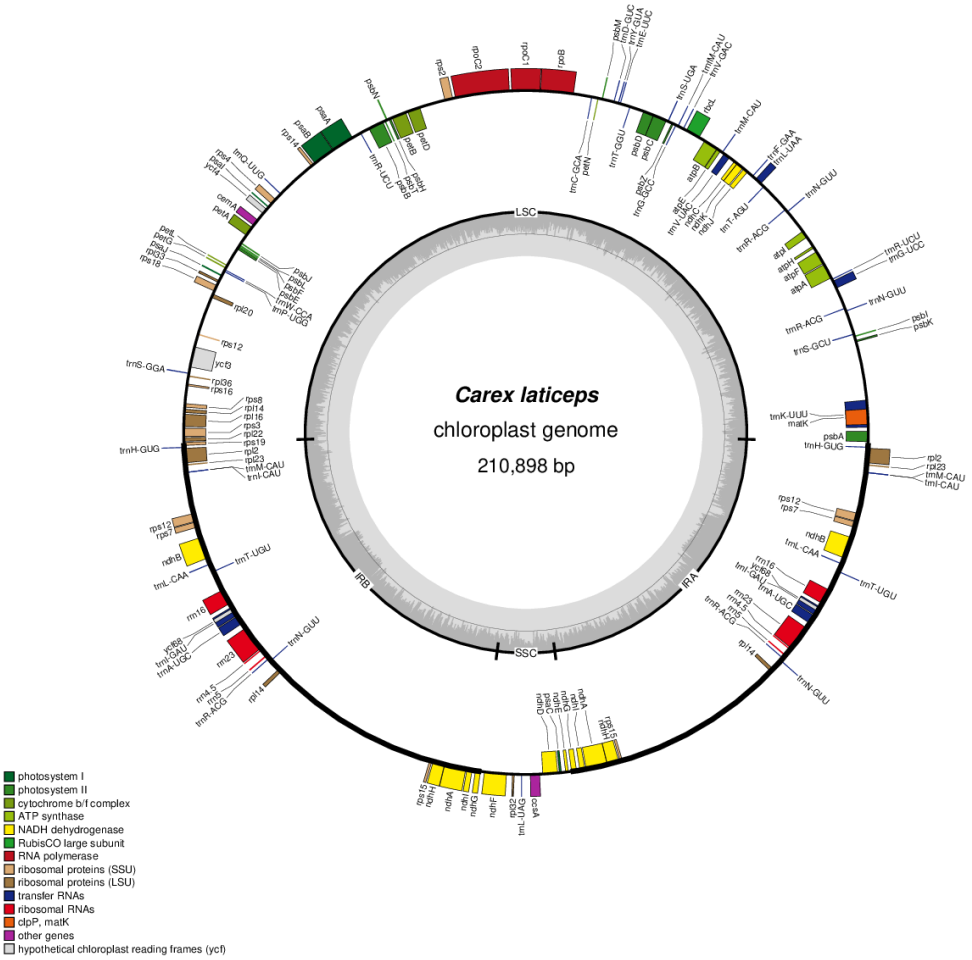


Fig. 2.18. Gene map of the chloroplast genome in *C. laticeps* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

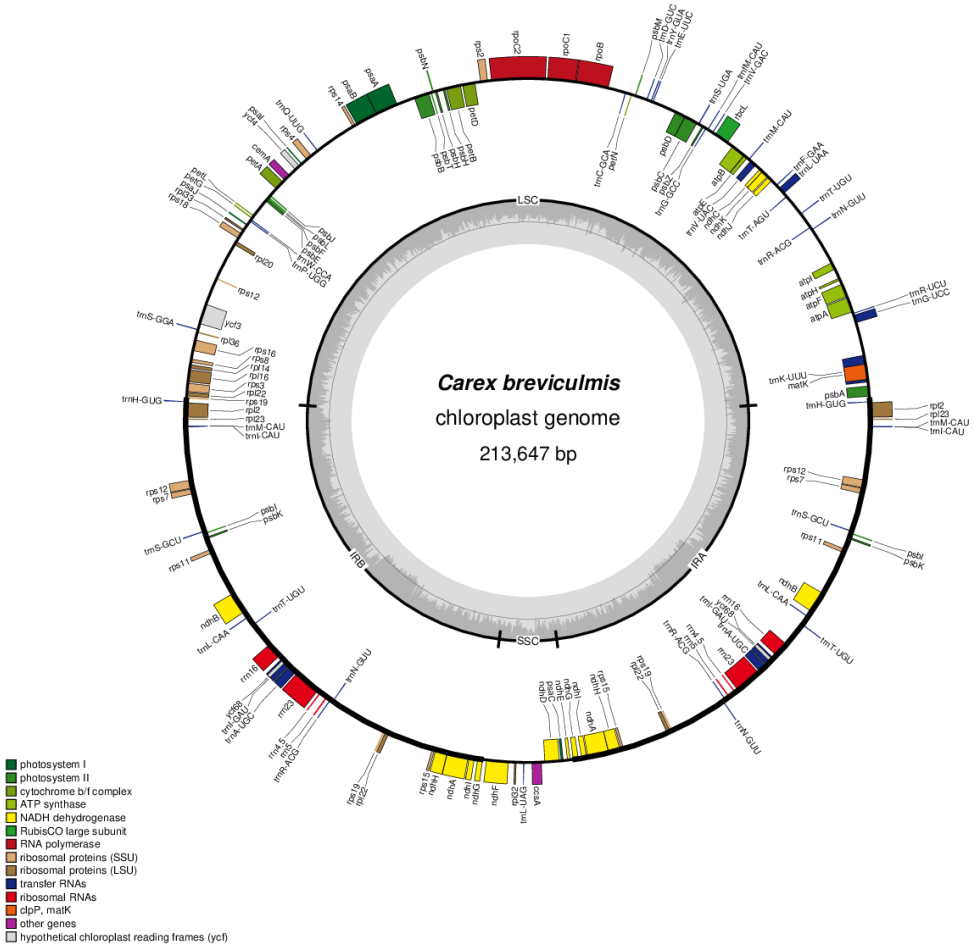


Fig. 2.19. Gene map of the chloroplast genome in *C. breviculmis* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

Fig. 2.20. Gene map of the chloroplast genome in *C. peiktusani* using the OGDRAW software (v1.3.1; Greiner et al., 2019)

2.3.3. Repetitive sequence patterns in chloroplast genomes within Poales: type, length, and distribution

The *Carex* species exhibit a notably higher number of repeats compared to other species (Fig. 2.21). Specifically, *C. breviculmis* and *C. peiktusani* contain over 3,000 repeats, predominantly of the forward and palindromic types, with fewer complement and reverse repeats. Notably, *C. peiktusani* reached REPuter's maximum detection threshold of 5,000 repeats, indicating that the actual repeat count in this species may exceed this limit.

Regarding repeat length distribution, over 10 repeats exceeding 1,000 bp were found in *C. cinerascens*, *C. dickinsii*, *C. laticeps*, and *C. peiktusani*, with additional long repeats observed in *J. fauriei* (Fig. 2.22). Conversely, *T. latifolia*, *E. buergerianum*, and *O. sativa*—belonging to Typhaceae, Eriocaulaceae, and Poaceae, respectively—showed no repeats over 1,000 bp. Similarly, *C. capricornis* and *C. paxii* (Cyperaceae) lacked repeats of this length but contained approximately 1,000–1,500 shorter repeats (Fig. 2.21), contrasting with *T. latifolia*, *E. buergerianum*, and *O. sativa*, which displayed very few repeats overall.

The distribution of repetitive sequences across chloroplast

genome regions varied by species. Most repeats were concentrated in the LSC region, followed by the IR region, with some species exhibiting repeats within the SSC region (Fig. 2.23). In *T. latifolia*, repeats appeared to a lesser extent in the SSC region, though numbers remained lower than in other regions. Other species also displayed low repeat levels in the SSC region.

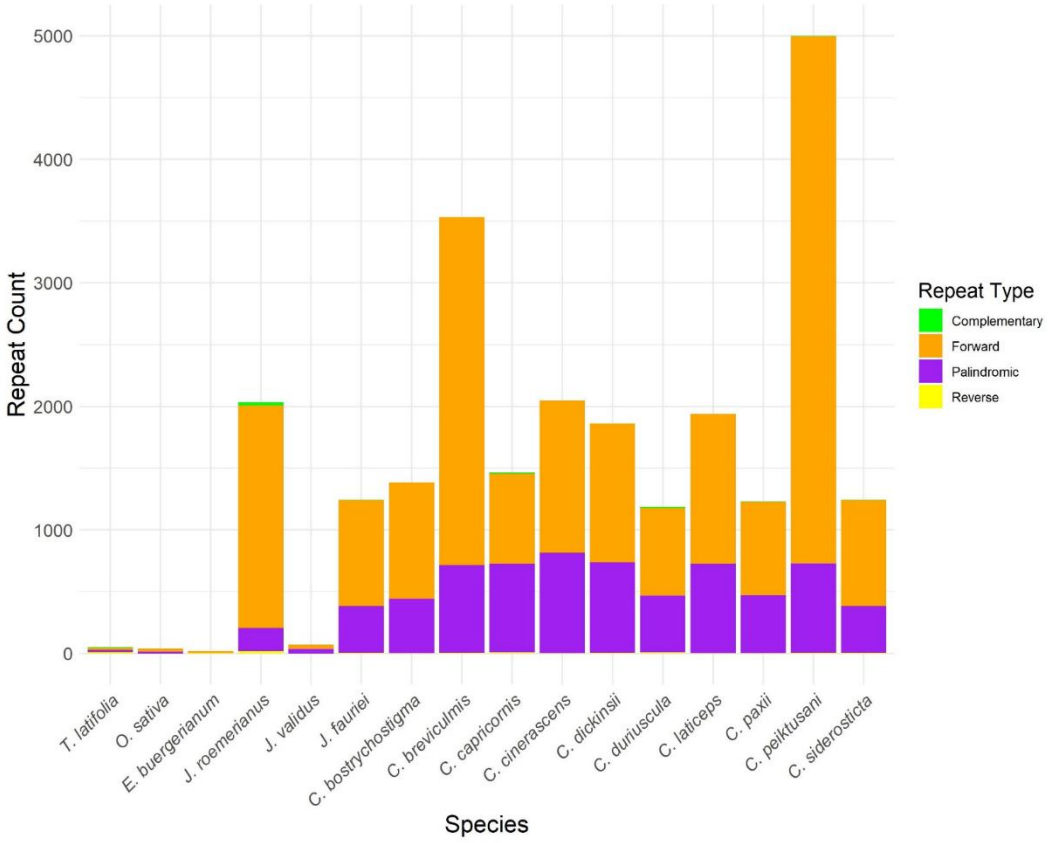


Fig. 2.21. Presents the results of a repeat sequence analysis conducted across 16 species in the Poales. The bar plot illustrates the counts of various repeat types (complementary, forward, palindromic, and reverse) within each species. This analysis was conducted using REPuter (Kurtz et al., 2001).

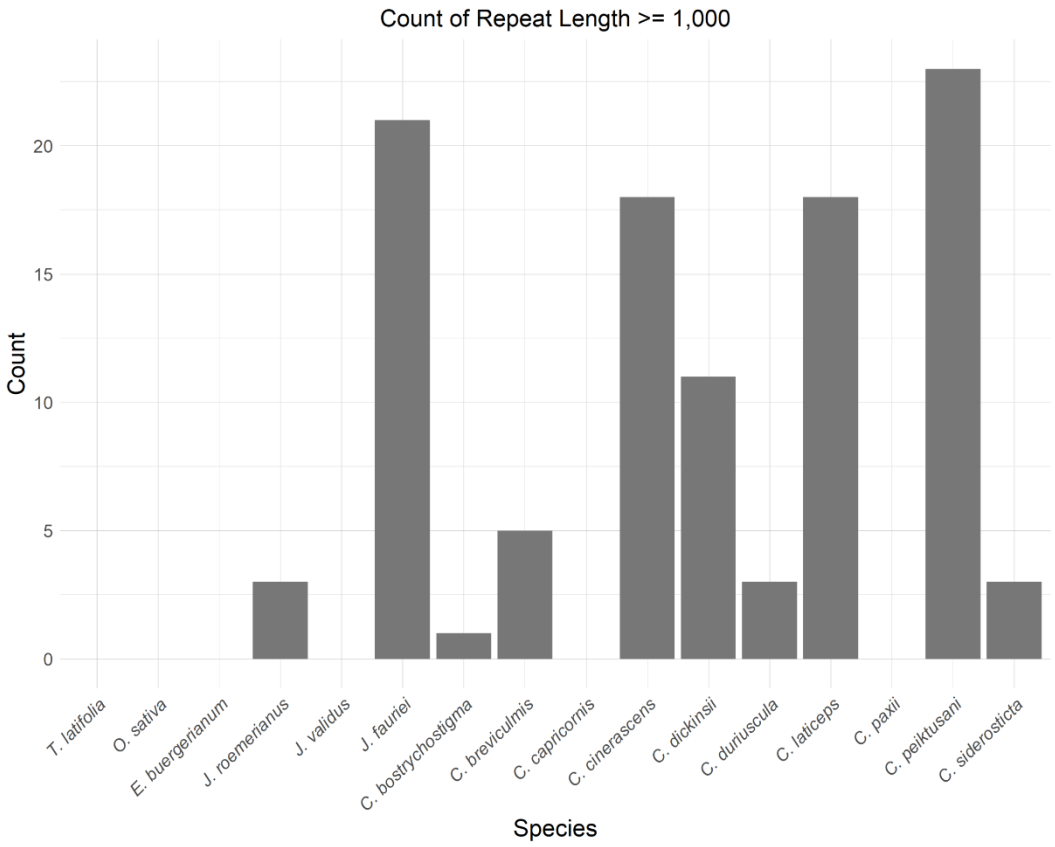


Fig. 2.22. Counts of repeat sequences with a length of 1 Kbp or greater across 16 species in Poales. The bar plot displays the number of long repeats (≥1,000 bp) identified within each species. This analysis was conducted using REPuter (Kurtz et al., 2001).

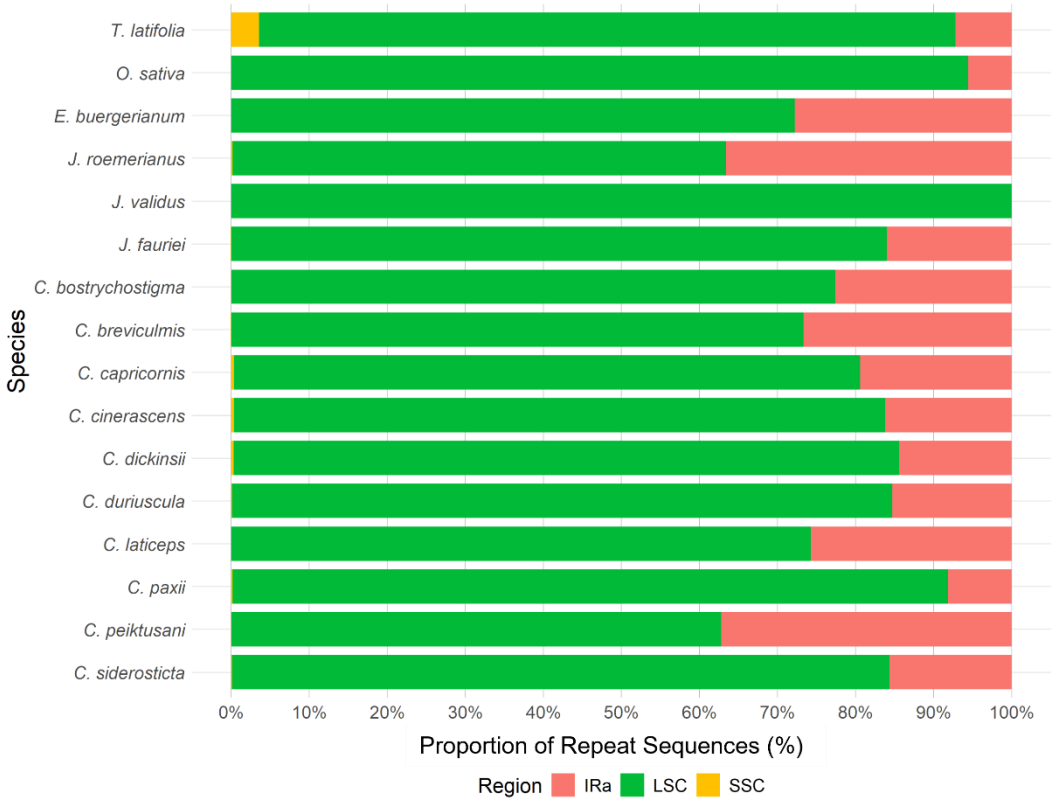


Fig. 2.23. Distribution of repeat sequences by genomic region across multiple Poales species. The horizontal stacked bar plot displays the proportion of repeat sequences located within the SSC, LSC, and IR regions for each species. This analysis was conducted using REPuter (Kurtz et al., 2001).

2.3.4. Structural complexity and rearrangements in chloroplast genomes of Poales

The chloroplast genome structures in *Carex* were compared with related taxa using pairwise dot plots and progressiveMAUVE (Fig. 2.24; Fig. 2.25). Typhaceae, representing one of the earliest-diverging lineages within Poales, served as a reference to investigate structural complexity and rearrangements across the order. Only one structural change was observed between *T. latifolia* (Typhaceae) and *E. buergerianum* (Eriocaulaceae) (Fig. 2.24A.a). In contrast, structural variations increased markedly as the comparison progressed from Typhaceae to Cyperaceae (Fig. 2.24B-D). This analysis revealed a gradual increase in structural changes and complexity from Typhaceae to Cyperaceae.

The comparison between *T. latifolia* (Typhaceae) and *O. sativa* (Poaceae) revealed relatively few structural variations, including a 3.9 kb translocation, a 23 kb inversion, and a 2 kb indel. Additionally, a small inverted translocation was detected (Fig. 2.24B). When comparing *T. latifolia* (Typhaceae) and *J. fauriei* (Juncaceae), the frequency and scale of structural variations increased. These variations included inversions ranging from 6 to 12 kb, translocations of 1.5 kb and 3 kb, and 6 kb

inversion + indel mutations. Additionally, indels up to 11 kb and short inverted translocations were identified. These results suggest that the chloroplast genome of Juncaceae exhibits greater structural dynamics compared to Poaceae (Fig. 2.24C). In the analysis of *T. latifolia* (Typhaceae) and Cyperaceae, represented by *C. siderosticta*, structural complexity was most pronounced. These included 16 and 20 kb inversion + indel mutations, a 3 kb translocation, and a 19 kb inverted translocation, along with additional indels. A short inverted duplication was also detected in the IR region. These structural rearrangements highlight the high genomic plasticity and structural diversity of the chloroplast genome in Cyperaceae, particularly in the genus *Carex* (Fig. 2.24D).

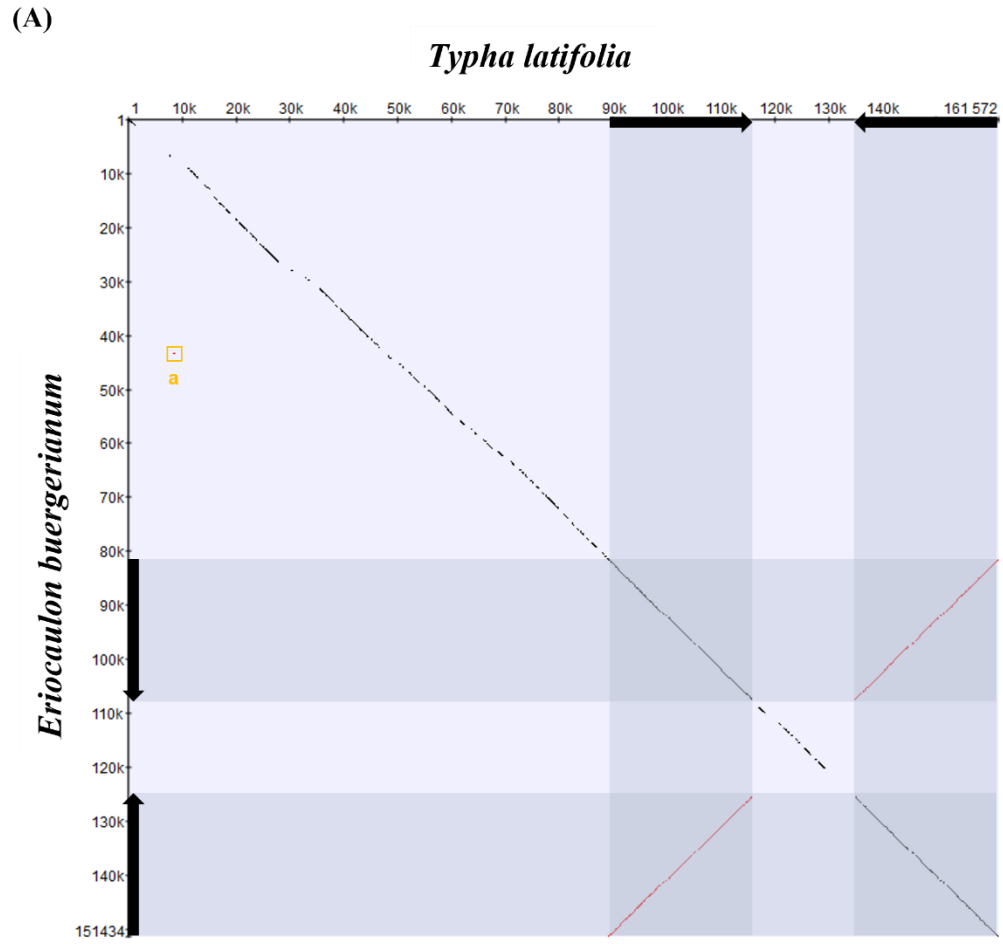


Fig. 2.24. Pairwise dot plots of Poales taxa (minimum repeat length of 100 bp and repeat identity threshold of 90%). Major structural changes are indicated in different colors (blue: inversion; green: indel; purple: translocation; orange: inverted translocation; pink: inversion + indel).

Fig. 2.24. (Continued) Dot plots include (A) between Typhaceae and Eriocaulaceae, (B) between Typhaceae and Poaceae, (C) between Typhaceae and Juncaceae, and (D) between Typhaceae and *C. siderosticta* (Cyperaceae). Among the 10 *Carex* species analyzed in this study, *C. siderosticta* represents the most basal lineage and serves as a representative of the Cyperaceae in this comparison.

(B)

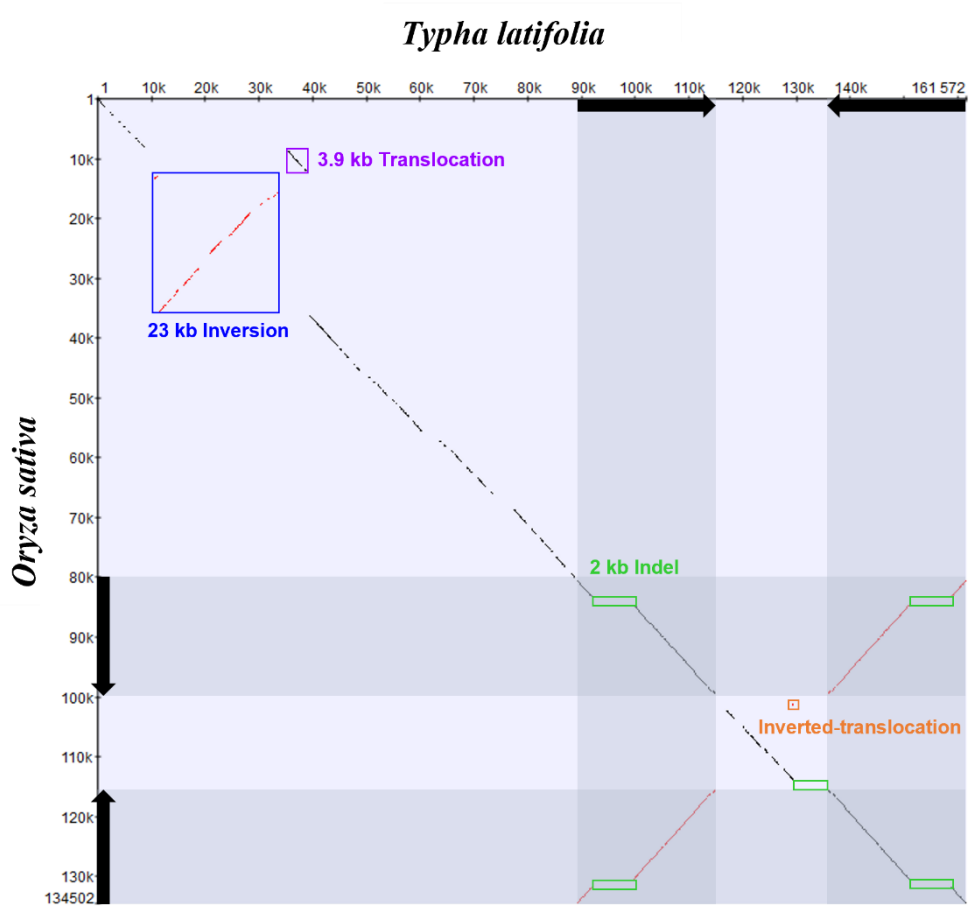


Fig. 2.24. (Continued)

(C)

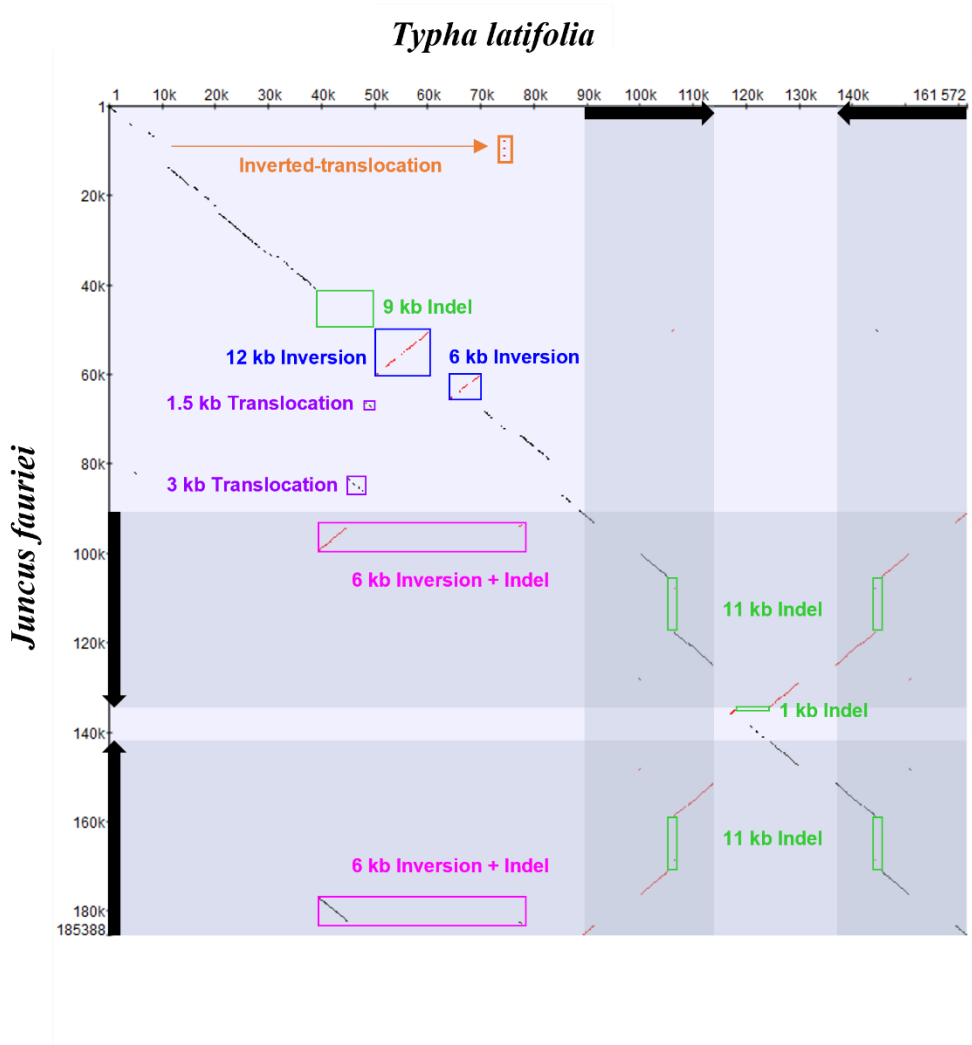


Fig. 2.24. (Continued)

(D)

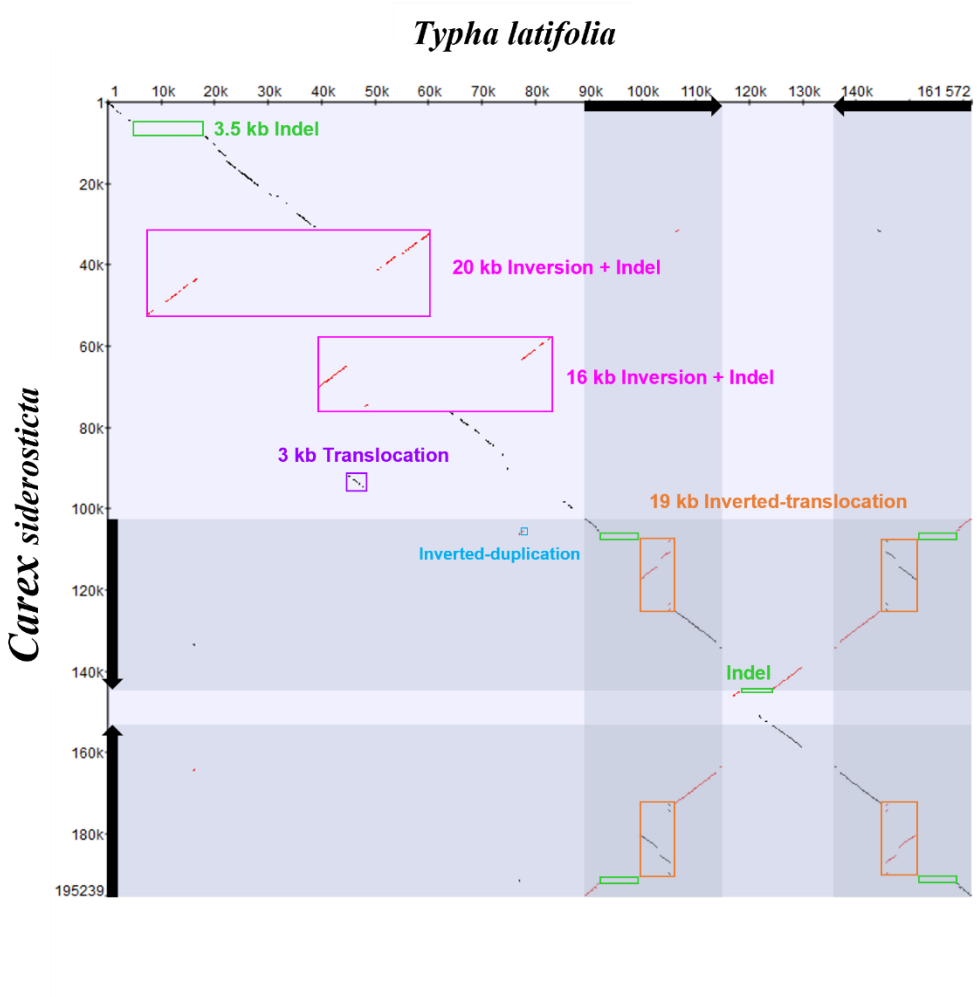


Fig. 2.24. (Continued)

Fig. 2.25. Whole genome alignment of Poales taxa. The Locally Collinear Blocks (LCBs) were identified by progressiveMAUVE alignment using the chloroplast genome of *C. siderosticta* as a reference, with one copy of the IR region removed prior to alignment to prevent redundancy.

Fig. 2.25. (Continued) Annotated features are displayed as follows: protein-coding regions (CDS) as white boxes, transfer RNAs (tRNAs) in green, ribosomal RNAs (rRNAs) in red, and miscellaneous RNA (misc_RNA) features in blue. Evolutionary insights into the chloroplast genomes of Cyperaceae: Phylogenetic relationships and gene structure divergence in Poales

To analyze the phylogenetic relationships and gene structure differences between Cyperaceae and other taxa within Poales, 79 homologous amino acid sequences were clustered and aligned using Orthofinder (v2.5.5) and the MUSCLE algorithm (v3.8.1551). For each gene, the optimal model was selected (Table 2.5), and a phylogenetic tree was constructed with IQ-TREE (v2.3.6) using the ultrafast bootstrap method.

Table 2.5. The best amino-acid substitution models for 79 coding genes used in the phylogenetic analyses of IQ-TREE (Minh et al., 2020) with the partition model (Chernomor et al., 2016)

Number	Gene	Best model	Number	Gene	Best model
1	*accD*	VT+F	41	*psbF*	mtMAM
2	*atpA*	Q.plant+F+G4	42	*psbH*	FLAVI+G4
3	*atpB*	Q.mammal+F+G4	43	*psbI*	mtMet
4	*atpE*	Q.mammal+F	44	*psbJ*	mtMAM
5	*atpF*	mtVer+F+G4	45	*psbK*	mtMet+G4
6	*atpH*	mtMAM+I	46	*psbL*	mtMet
7	*atpI*	Q.bird+F+G4	47	*psbM*	mtMet
8	*ccsA*	Q.bird+F+G4	48	*psbN*	mtREV+I
9	*cemA*	Q.mammal+F+I	49	*psbT*	mtMet
10	*clpP1*	JTTDCMut+F+G4	50	*psbZ*	mtInv
11	*infA*	VT+F+G4	51	*rbcL*	mtZOA+F+R2
12	*matK*	Q.mammal+F+I	52	*rpl14*	Q.plant+F+I
13	*ndhA*	mtMet+F+G4	53	*rpl16*	mtREV+F+I+R2
14	*ndhB*	Q.bird+F	54	*rpl2*	Q.mammal+F+I
15	*ndhC*	Q.bird+F	55	*rpl20*	mtMet+F+G4
16	*ndhD*	Q.bird+F+G4	56	*rpl22*	JTT+F+G4
17	*ndhE*	mtMAM+F+G4	57	*rpl23*	PMB+F
18	*ndhF*	Q.mammal+F+G4	58	*rpl32*	mtVer+F+G4
19	*ndhG*	Q.mammal+F	59	*rpl33*	Q.bird+F+G4
20	*ndhH*	Q.mammal+F+G4	60	*rpl36*	HIVw+F+G4
21	*ndhI*	JTT+F	61	*rpoA*	VT+F+G4
22	*ndhJ*	JTT+F+G4	62	*rpoB*	Q.mammal+F+G4
23	*ndhK*	HIVw+F+G4	63	*rpoC1*	Q.bird+F+G4
24	*pbf1*	mtMAM	64	*rpoC2*	Q.mammal+F+G4
25	*petA*	Q.mammal+F+G4	65	*rps11*	Q.mammal+F+G4
26	*petB*	FLU+F+G4	66	*rps12*	PMB+F
27	*petD*	mtMAM+F+I	67	*rps14*	mtVer+F+G4
28	*petG*	mtZOA+I	68	*rps15*	mtMet+F+I
29	*petL*	mtMet	69	*rps16*	JTT+F+I
30	*petN*	mtMAM	70	*rps18*	JTT+F+G4
31	*psaA*	mtMAM+F+G4	71	*rps19*	WAG+F+G4
32	*psaB*	FLU+F+G4	72	*rps2*	HIVb+F+G4
33	*psaC*	Q.bird+G4	73	*rps3*	JTT+F+G4
34	*psaI*	mtMet	74	*rps4*	mtMet+F+I+R2
35	*psaJ*	mtMAM	75	*rps7*	JTT+F+I+R2
36	*psbA*	FLU+F+G4	76	*rps8*	JTT+F+G4
37	*psbB*	mtMAM+F+I+R2	77	*ycf3*	HIVw+F+G4
38	*psbC*	mtMAM+F+G4	78	*ycf4*	Q.plant+F+G4
39	*psbD*	FLU+F+G4	79	*ycf68*	Q.mammal
40	*psbE*	mtMAM+G4			

The results indicated that the *clpP* was only present in Typhaceae, Eriocaulaceae, and Poaceae, while it was absent in Juncaceae and Cyperaceae (Fig. 2.26). This suggests that Cyperaceae possesses a gene structure that differentiates it from other Poales lineages such as Poaceae and Eriocaulaceae. Furthermore, analysis of the *accD* revealed that it was absent in Juncaceae but present as a single copy in Typhaceae, Eriocaulaceae, and Poaceae. Most species within Cyperaceae contained multiple copies of *accD*, except for *C. paxii*, *C. laticeps*, and *C. breviculmis*, which lacked *accD* altogether. These three species also lacked *clpP1* and *infA*. This pattern of multiple gene duplications and specific gene losses highlights the structural complexity and gene variability within the chloroplast genome of Cyperaceae (Fig. 2.26).

According to Zhou et al. (2023), the chloroplast genome size of *J. validus* is smaller than that of other *Juncus* species, likely due to the expansion of the IR region, which reduced the SSC region to approximately 2 kbp and resulted in the loss of 11 *ndh* genes as well as *rps15* and *ycf4*. These characteristics indicate that Juncaceae has structural features distinguishing it from other Poales taxa (Fig. 2.26).

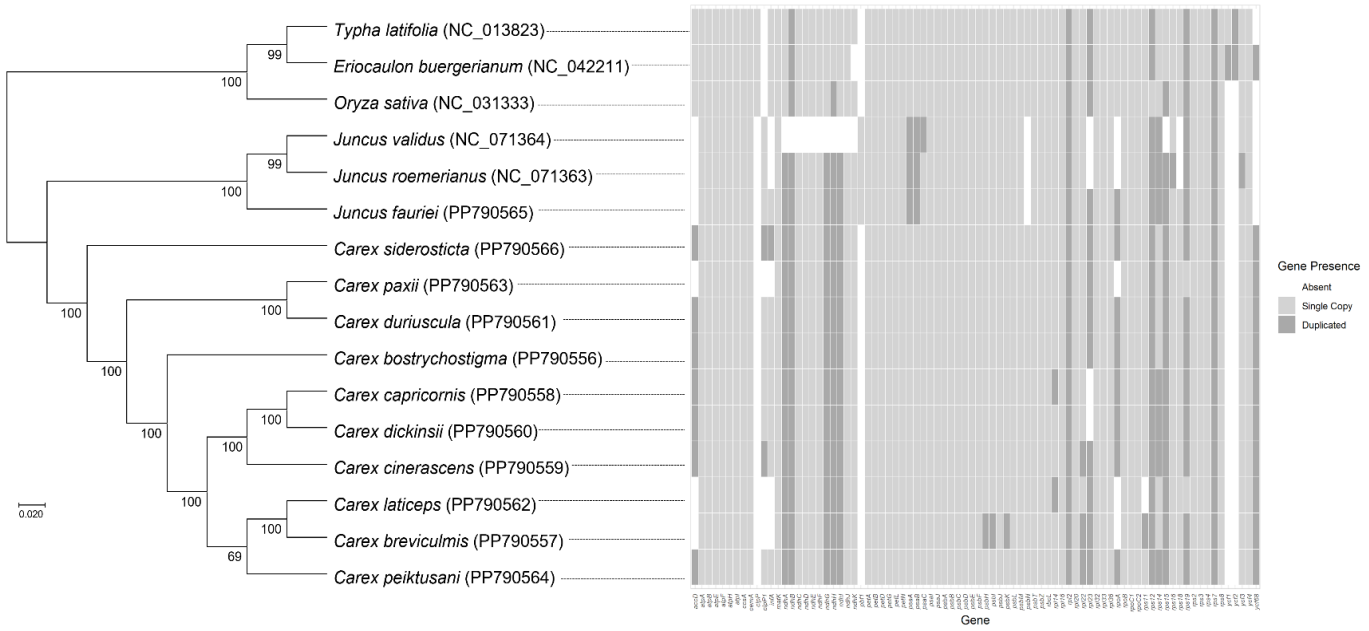


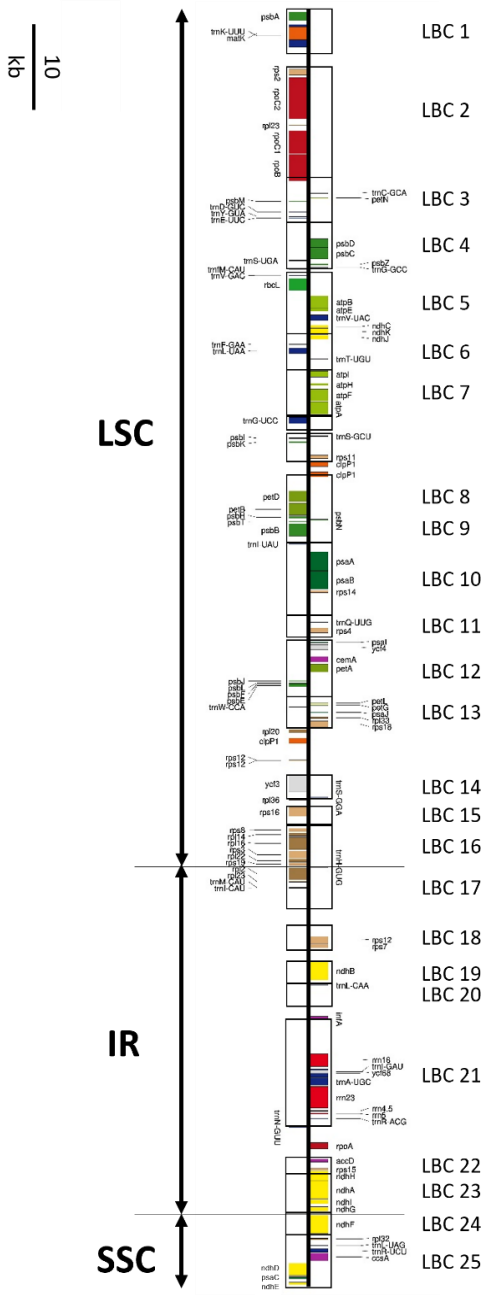
Fig. 2.26. Maximum-likelihood tree based on amino acid sequences of 79 coding genes with presence (gray) / absence (white) matrix. Dark gray indicates two or more duplicated paralogous genes in each taxon, and light gray indicates a single gene. One of the duplicated genes was randomly selected for analysis.

Using *C. siderosticta*, the earliest-diverging species in the genus *Carex* and a member of the subgenus *Siderosticta*, as a reference, 25 locally collinear blocks (LCBs) were identified to analyze chloroplast genome structural variations (Fig. 2.27; Table 2.6). By assigning LCB order based on *C. siderosticta*, structural rearrangements across *Carex* species were revealed, representing the simplest configuration involving the fewest rearrangements and serving as the most parsimonious arrangement across the genus.

Distinct structural rearrangements were observed within and across subgenera and clades. For instance, in the subgenus *Vignea*, *C. paxii* and *C. duriuscula* displayed inversions in LCB 10–15 and 19–21, suggesting unique gene arrangements that distinguish *Vignea* from other *Carex* subgenera (Fig. 2.28). In the transition to subgenus *Carex*, an inversion spanning approximately 50 kb across LCB 2–9 was identified, indicating a characteristic structural modification within this lineage (Fig. 2.28).

Additional structural differences were also found among specific sections and species. During the divergence of sections VI.H.1.Hirta (including *C. capricornis* and *C. dickinsii*) and VI.K.Phacocystis,

inversions in LCB 10–14 were observed, distinguishing them from other clade in *Carex* subgenus. Furthermore, *C. breviculmis* exhibited a deletion in LCB 9. *C. peiktusani* displayed another inversion in LCB 9–2, marking significant structural divergence within the genus *Carex*. These structural rearrangements, including inversions, translocations, and deletions annotated across the phylogenetic tree (Fig. 30), highlight the structural evolution underlying the diversity within *Carex*.



ID	Start site	End site	Length	Region
LBC 1	1	5103	5103	LSC
LCB 2	6897	20026	13130	LSC
LCB 3	20088	25625	5538	LSC
LCB 4	25626	31009	5384	LSC
LCB 5	31422	38764	7343	LSC
LCB 6	38765	42993	4229	LSC
LCB 7	42994	48593	5600	LSC
LCB 8	48880	50310	1431	LSC
LCB 9	50458	54242	3785	LSC
LCB 10	55850	64119	8270	LSC
LCB 11	64120	72686	8567	LSC
LCB 12	72687	75262	2576	LSC
LCB 13	75499	82328	6830	LSC
LCB 14	82329	86073	3745	LSC
LCB 15	91787	94757	2971	LSC
LCB 16	95657	97509	1853	LSC
LCB 17	97787	107778	9992	LSC, IR
LCB 18	109948	112837	2890	IR
LCB 19	113951	116856	2906	IR
LCB 20	116857	119463	2607	IR
LCB 21	121146	134095	12950	IR
LCB 22	137676	139647	1972	IR
LCB 23	139648	144364	4717	IR, SSC
LCB 24	144365	147050	2686	SSC
LCB 25	147051	153351	6301	SSC

Fig. 2.27. Linear map of the chloroplast genome of *C. siderosticta* (PP790566), with 25 identified locally collinear blocks (LCBs) indicated along the LSC, IR, and SSC regions.

Table 2.6. LCB order of *Carex* species using *C. siderosticta* as a reference

Region	Large Single Copy (LSC)															Inverted Repeat (IR)									Small single Copy (SSC)	
Spcies		LCB 2-9, 9-2 inversion and 9 indel								LCB 10-15 and 10-14 inversion						LCB 21 translocation				LCB 19-20 inversion			Conserved region			
C. siderosticta	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18	19	20	21	22	23	24	25
C. paxii	1	2	3	4	5	6	7	8	9	-15	-14	-13	-12	-11	-10	16	17		18	-21	-20	-19	22	23	24	25
C. duriuscula	1	2	3	4	5	6	7	8	9	-15	-14	-13	-12	-11	-10	16	17		18	-21	-20	-19	22	23	24	25
C. bostrychostigma	1	-9	-8	-7	-6	-5	-4	-3	-2	10	11	12	13	14	15	16	17	21	18	19	20		22	23	24	25
C. capricornis	1	-9	-8	-7	-6	-5	-4	-3	-2	-14	-13	-12	-11	-10	15	16	17		19	19	20	21	22	23	24	25
C. dickinsii	1	-9	-8	-7	-6	-5	-4	-3	-2	-14	-13	-12	-11	-10	15	16	17		20	19	20	21	22	23	24	25
C. cinerascens	1	-9	-8	-7	-6	-5	-4	-3	-2	-14	-13	-12	-11	-10	15	16	17		21	19	20	21	22	23	24	25
C. laticeps	1	-9	-8	-7	-6	-5	-4	-3	-2	10	11	12	13	14	15	16	17		22	19	20	21	22	23	24	25
C. breviculmis	1	LCB 9 indel	-8	-7	-6	-5	-4	-3	-2	10	11	12	13	14	15	16	17		23	19	20	21	22	23	24	25
C. peiktusani	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		24	19	20	21	22	23	24	25

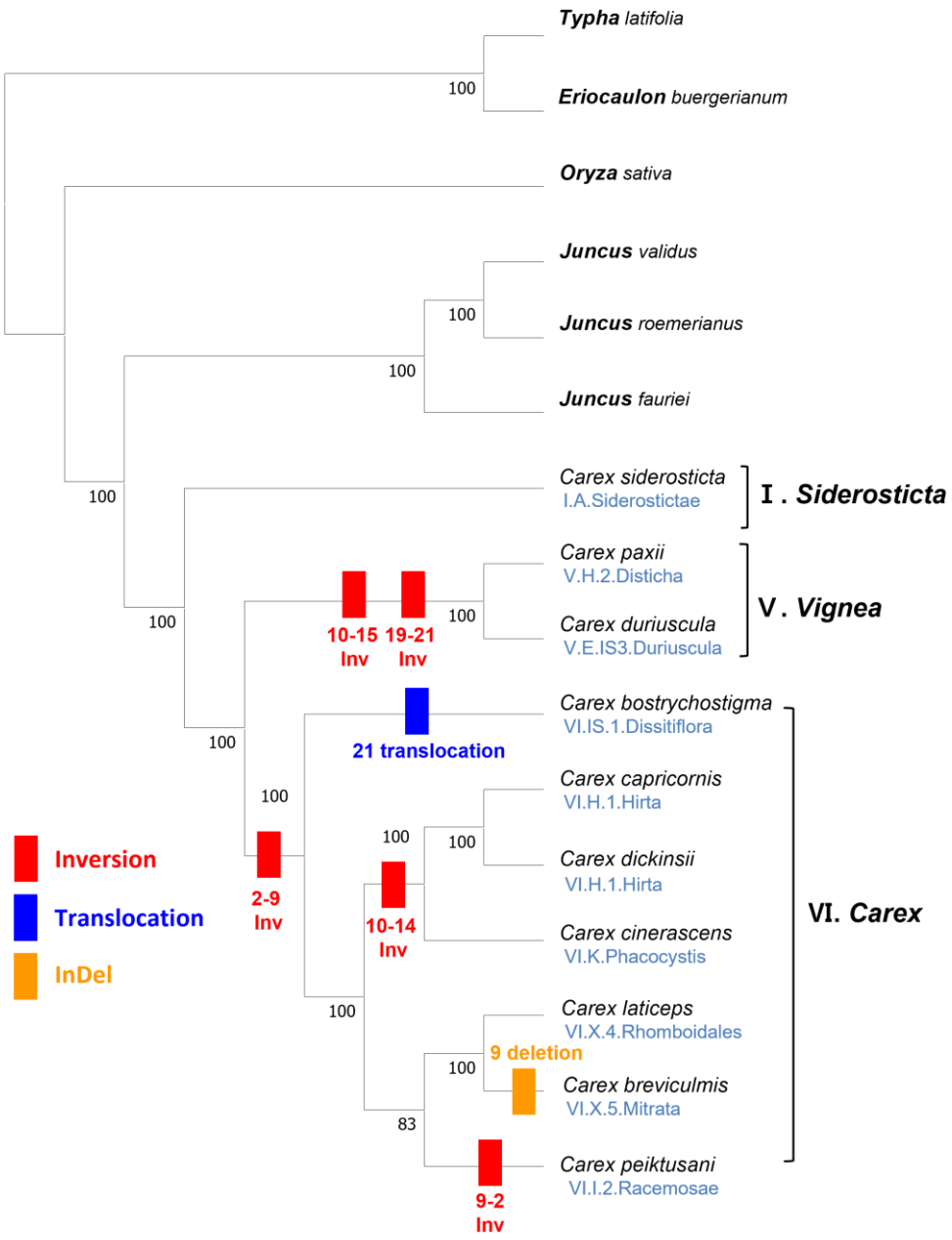


Fig. 2.28. Phylogenetic tree and distribution of structural variation in chloroplast genomes constructed from 25 synteny block DNA sequences.

Fig. 2.28. (Continued) This phylogenetic tree visualizes structural variation across species in chloroplast genomes, with major structural mutations indicated by color-coded synteny block numbers. Red indicates inversion mutations, blue indicates translocation mutations, and orange indicates insertion/deletion mutations; each number represents one synteny block. Bootstrap values reflect the confidence in each node. On the right, species are grouped by subgenus (*Siderosticta*, *Vignea*, *Carex*), with specific structural variations annotated within each clade. The blue text below each *Carex* species name represents the section to which that species belongs within the classification hierarchy.

Finally, the phylogenetic trees generated from both the 79 CDS sequences and the 25 LCBs sequences exhibited consistent topologies. Notably, the bootstrap values for the clade containing *C. laticeps*, *C. breviculmis*, and *C. peiktusani* increased when LCB sequences were used, suggesting that LCBs may better reflect structural variation compared to CDS sequences (Fig. 2.26; Fig. 2.28). Overall, Cyperaceae, particularly the genus *Carex*, exhibits considerable structural complexity and genomic rearrangements relative to other Poales taxa, providing

valuable insights into the evolutionary trajectory and ecological specialization within *Carex*.

2.4. Discussion

2.4.1. Chloroplast genome structure and stability

The chloroplast genome of *Carex* contains abundant short dispersed repeats (SDRs) and long repeat sequences, which contribute to its structural instability. As shown in Table 2.4, *Carex* species generally exhibit lower GC content than related taxa, potentially contributing to decreased structural stability (Ruhlman et al., 2017; Papanikolaou et al., 2009). Lower GC content reduces the binding energy between base pairs, making the genome more flexible and, consequently, more susceptible to abnormal recombination events. Therefore, the combination of low GC content and repetitive sequences in *Carex* may increase genome instability, potentially driving various structural rearrangements. These findings align with studies on *Eleocharis* (Cyperaceae), which also reported chloroplast genome heteroplasmy associated with structural polymorphism (Lee et al., 2020).

2.4.2. Increasing structural complexity in Poales chloroplast genomes

The dot-plot analysis of Poales chloroplast genomes (Fig. 2.24) demonstrates a clear trend of increasing structural complexity and rearrangements from Typhaceae to Cyperaceae. The comparison between *T. latifolia* (Typhaceae) and *E. buergerianum* (Eriocaulaceae) showed relatively stable chloroplast genome structures. In contrast, comparisons with Poaceae, Juncaceae, and Cyperaceae revealed a gradual increase in complexity, with the most pronounced structural changes observed in Cyperaceae. In particular, the large-scale inversions and complex rearrangements observed in *C. siderosticta* (Cyperaceae) suggest that structural changes in Poales chloroplast genomes extend beyond simple inversions, translocations, and indels to exhibit more complex patterns. These findings are consistent with the results of Lee (2021) and Xu et al. (2023a), supporting the high structural diversity and plasticity of the chloroplast genome in *Carex*. Overall, these results highlight a progressive increase in structural complexity across Poales chloroplast genomes, with Cyperaceae exhibiting structural changes that may have played a significant role in their evolutionary adaptation and diversity. The large-scale rearrangements identified in the chloroplast genome of *Carex* provide

valuable insights into the phylogenetic relationships and structural evolution within Poales.

2.4.3. Phylogenetic insights using chloroplast genome structural variations in *Carex*

The classification of the genus *Carex* has traditionally been based on morphological characteristics. However, the issue of homology in morphological traits has often led to inaccuracies in reflecting phylogenetic relationships, as convergent evolution can result in morphologically similar species belonging to entirely different phylogenetic groups. To address this limitation, nuclear ribosomal DNA markers such as ETS and ITS were introduced. While these molecular approaches have been useful in exploring phylogenetic relationships among closely related species, they have inherent challenges. Low gene concordance often makes it difficult to resolve specific phylogenetic relationships, and the high rate of incongruence, where multiple ITS copies or variants exist within a species, can distort phylogenetic analyses. For instance, some non-functional ITS variants may skew results or lead to an underestimation of genetic diversity when only the dominant copy is analyzed (The Global *Carex* Group, 2016). Recently, NGS-based HybSeq has improved phylogenetic resolution by targeting multiple exon regions. However, it is largely focused on coding regions and fails to adequately capture non-coding regions or large-scale

structural variations of entire genomes. Additionally, as HybSeq relies on Illumina short-read data, it is insufficient for exploring the complex genome structures typical of *Carex* (Villaverde et al., 2020). To overcome these limitations, this study utilized nanopore sequencing with long-read data to successfully assemble the complete chloroplast genomes of 10 *Carex* species. This approach allowed for simultaneous analysis of entire genome sequences, including both coding and non-coding regions, as well as structural variations. Given the intricate genome structure and abundant repetitive sequences characteristic of *Carex*, long-read data proved indispensable for detecting large-scale structural variations and repetitive sequence patterns that were previously undetectable using short-read methodologies. Furthermore, this study provided new insights by analyzing the relationship between genome structural variations and morphological traits. For example, Fig. 2.25 and Table 2.6 illustrate the LCB 2-10 order shared by *C. siderosticta*, *C. paxii*, *C. duriuscula*, and *C. peiktusani*. These four species share a common morphological characteristic of bisexual spikes, where staminate (male) flowers are positioned above pistillate (female) flowers within the same spike. This suggests that specific genome structural variations may be associated with bisexual spike morphology,

highlighting the importance of integrating molecular and morphological approaches for a more comprehensive understanding of phylogenetic relationships. Notably, this study demonstrates a novel and distinct approach by utilizing genome structural variation patterns for phylogenetic classification. To date, no such classification has been performed for *Carex*, making this research unique in leveraging the advantages of complete chloroplast genome data to address limitations in traditional methods. In conclusion, this study integrates chloroplast genome analysis using nanopore-based long-read data with morphological traits to clarify phylogenetic relationships in *Carex*. By presenting previously unrecognized patterns, this approach not only addresses gaps in earlier studies but also offers broad applicability to other large and complex taxonomic groups. Future studies with additional data will further elucidate the connections between molecular classifications and morphological characteristics, contributing to a deeper understanding of evolutionary patterns.

3. *Carex* mitochondrial genome

3.1. Introduction

3.1.1. Characteristics of plant mitochondrial genome

Mitochondria are organelles within eukaryotic cells that play a crucial role in energy production through cellular respiration. In plant cells, mitochondria are essential for breaking down sugars produced by photosynthesis to generate energy in the form of ATP (Notsu et al., 2002; Chevigny et al., 2020; Mϕller et al., 2021).

The plant mitochondrial genome differs significantly from the animal mitochondrial genome in terms of size, structure, and gene content (Knoop et al., 2010). Animal mitochondrial genomes are typically small, ranging from 15 to 20 kb, and contain 37 genes (Boore, 1999). In contrast, plant mitochondrial genomes are much larger and more variable, ranging from 191 kb to 11.3 Mb, reflecting their more complex structure compared to animals. Among angiosperms (flowering plants), *Cucumis melo* var. *momordica* is reported to have the largest mitochondrial genome, measuring approximately 2.71 Mb and containing 83 genes (as of November 12, 2024, INSDC; GenBank accession: MG947207).

A distinctive feature of plant mitochondrial genomes is their

abundance of non-coding and repetitive sequences. These repetitive sequences drive homologous recombination, resulting in mitochondrial genomes that exist not only as a single master circle but also as multipartite genomes comprising multiple circular or linear molecules. A master circle represents the entire genome as a single large circular molecule, while multipartite genomes consist of several independent molecules (Fischer et al., 2022). Such mechanisms lead to heteroplasmy, where different mitochondrial DNA copies coexist within the same species. Heteroplasmy plays a significant role in plant adaptation and evolution (Mower et al., 2012; Richardson et al., 2013; Kozik et al., 2019; Yang et al., 2022). Despite advancements in next-generation sequencing (NGS) technologies and improved assembly strategies, large repetitive sequences and multipartite genome structures continue to pose challenges for complete genome assembly. This structural and genetic diversity represents a critical research challenge for understanding the dynamics of plant evolution and the relationship between genome structure and function.

3.1.2. Limited mitochondrial research within the *Carex*

Currently, the only reported mitochondrial genome within the *Carex* is that of *C. breviculmis* (GenBank accession: NC_068626; Xu et al., 2023b). The Korean endemic species *Carex pseudochinensis* H. Lév. & Vaniot (1902) plays an important role in maintaining biodiversity, yet research on this species remains limited (Kim et al., 2009). Studies have documented its distribution in Heoninleung, Seoul (Kim et al., 2010) and its ecological significance within herbaceous plant communities on Mt. Cheongoksan, Korea (Son et al., 2014). However, genetic and genomic data are still scarce, limiting our understanding of its evolutionary history.

Given the high levels of polymorphism and structural variation often observed in plant mitochondrial genomes, investigating the mitochondrial genome of *C. pseudochinensis* presents a valuable opportunity to understand the evolutionary complexity of this genus and to explore the genomic plasticity that may underlie its adaptability and diversity.

3.2. Matrerials and methods

3.2.1. Plant materials

C. pseudochinensis was collected from Wolchon-ri, Gunbuk-myeon, Haman-gun, Korea (N35.306179°, E128.320333°), and the specimen was deposited in the herbarium at Sungshin Women's University (*Y. Cho s. n.,* SWU0036913) (Fig. 3.1; Table 1.1; Index 11). Species identification was based on the recently published Korean Cyperaceae manuals (Cho et al., 2016).

Fig. 3.1. The voucher specimen of *C. pseudochinensis* used in this study (SWU0036913)

3.2.2. DNA extraction and sequencing

A high-molecular-weight (HMW) DNA extraction method optimized for effective long-read sequencing was employed for high-quality third-generation genome sequencing (Kang et al., 2023). The concentration and quality of the extracted DNA were measured using Nanodrop (Thermo Fisher Scientific, Massachusetts) and Qubit (Introgen, Massachusetts) systems, and the DNA fragment length was verified with the FEMTO Pulse system (Agilent Technologies).

Long-read sequencing data were generated on the MinION platform using R9 flow cells and the SQK-LSK109 library preparation kit (Oxford Nanopore Technologies, Oxford). Short-read sequencing was performed on the Illumina NovaSeq 6000 S4 platform using the TruSeq Nano DNA Kit (Illumina, San Diego).

3.2.3. Mitochondrial genome assembly and polishing

To assemble the mitochondrial genome of *C. pseudochinensis*, the Plant Organelle-Genome Long-Read Assembly Pipeline (POLAP), a seven-step pipeline specialized for plant organelle genome assembly, was utilized. POLAP enables high-accuracy genome assembly even

without a reference genome, making it suitable for this study. The assembly process consists of the following seven steps:

Step 1. Long-Read Data Filtering

To estimate the genome size of *C. pseudochinensis*, K-mer analysis was performed on short-read data using Jellyfish (Marçais & Kingsford, 2011). Following this estimation, sequences longer than 3 kb were extracted from the long-read dataset using Minimap2 (Li, 2018), and these filtered long reads were prepared for whole-genome assembly.

Step 2. Whole-Genome Assembly

Whole-genome assembly was conducted using Flye software (v2.9.3-b1797; Kolmogorov et al., 2019) with the filtered long-read sequences. Coverage option was set to 30×, and the previously estimated genome size was specified as an assembly parameter.

Step 3. Organelle Gene Annotation

A BLAST homology search was conducted to annotate organelle genes within the assembled contigs (Altschul et al., 1997). Plant organelle amino acid sequences from the NCBI Reference Sequence Database (RefSeq, Release 218) were used to build a local database for annotation

purposes.

Step 4. Mitochondrial DNA Seed Contig Selection

To identify seed contigs potentially containing mitochondrial DNA, the Bandage software (Wick et al., 2015) was used to visualize the whole-genome assembly graph. The criteria for selecting seed contigs included the presence of organelle genes, the copy number of the contig, and its connectivity within the assembly graph. The "copy number" refers to the extent to which sequencing data for a specific contig is detected repeatedly (sequencing depth). This measure is provided by the Flye as multiplicity values (v2.9.3-b1797; Kolmogorov et al., 2019). In general, mitochondrial-derived contigs in plant leaves exhibit higher copy numbers than nuclear DNA but lower than chloroplast-derived contigs (Henry et al., 2022). Based on these copy number criteria, mitochondrial-derived contigs were selected according to the expected coverage levels, improving the accuracy of mitochondrial genome assembly.

Step 5. Mitochondrial Long-Read Data Collection

Long-read data mapped to the selected seed contigs were isolated using Minimap2 (Li, 2018). These filtered long-read sequences were then

prepared for subsequent rounds of mitochondrial genome assembly.

Step 6. Mitochondrial Genome Assembly

The isolated long-read data were reassembled specifically for the mitochondrial genome using Flye software (v2.9.3-b1797; Kolmogorov et al., 2019). Assembly coverage option was set to 30×, with the genome size parameter defined by the total base count of the selected seed contigs.

Step 7. Polishing and Assembly Validation

The assembled mitochondrial draft genome from the long-read data was refined using FMLRC (v1.0.0; Mak et al., 2023; Zhou et al., 2023) with short-read data to enhance accuracy. Subsequently, the accuracy of the assembled genome was further validated by confirming uniform sequencing coverage of short- and long-read sequences using Minimap2 (v2.24; Li, 2018) and Geneious Prime (v2024.0.5; Kearse et al., 2012).

For more detailed information on the tools used in each step, refer to Table 3.1.

Table 3.1. Seven steps of a plant mitochondrial genome assembly and main tools

Step	Description	Main tool	References
1	Genome size estimation	Jellyfish	Marçais & Kingsford (2011)
2	Flye whole-genome assembly	Flye	Kolmogorov et al. (2019)
3	Organelle gene annotation	BLAST	Altschul et al. (1997)
4	Mitochondrion-origin contig selection	Bandage	Wick et al. (2015)
5	Mitochondrion-origin reads selection	Minimap2	Li (2018)
6	Flye organelle-genome assembly	Flye	Kolmogorov et al. (2019)
7	Polishing	FMLRC	Mak et al. (2023)

3.2.4. Mitochondrial genome annotation

The mitochondrial genome was annotated using the GeSeq tool (v2.03; Tillich et al., 2017) with reference sequences from 15 mitochondrial genomes of Poaceae species and the mitochondrial genome of *C. breviculmis* (Table 3.2). Protein-coding genes were annotated using translated BLAT (Kent, 2002), while rRNA and tRNA genes were annotated using nucleotide BLAT (Kent, 2002). In addition, tRNA genes were specifically annotated using the tRNAscan-SE software (v2.0.7; Chan, 2019).

All annotated genes were manually checked and curated in Geneious Prime (v2024.0.5; Kearse et al., 2012). The final circular genome map of the mitochondrial genome was generated using OGDRAW software (v1.3.1; Greiner et al., 2019). All options during these processes were used with default settings.

Table 3.2. Reference genomes used for annotation, including mitochondrial genomes from 15 Poaceae species and *C. breviculmis* (Xu et al., 2023b), downloaded from NCBI on February 7, 2024.

Family	GenBank no.	Species
Cyperaceae	NC_068626.1	*Carex breviculmis* R.Br.
Poaceae	NC_040989.1	*Eleusine indica* (L.) Gaertn.
Poaceae	NC_036024.1	*Triticum aestivum* L.
Poaceae	NC_022714.1	*Triticum timopheevii* (Zhuk.) Zhuk.
Poaceae	NC_022666.1	*Aegilops speltoides* Tausch
Poaceae	NC_013816.1	*Oryza rufipogon* Griff.
Poaceae	NC_011033.1	*Oryza sativa* L. (Japonica Group)
Poaceae	NC_029816.1	*Oryza minuta* J.Presl
Poaceae	NC_007886.1	*Oryza sativa* L. (Indica Group)
Poaceae	NC_008333.1	*Zea luxurians* (Durieu & Asch.) R.M.Bird
Poaceae	NC_008331.1	*Zea perennis* (Hitchc.) Reeves & Mangelsd.
Poaceae	NC_007982.1	*Zea mays subsp. mays*
Poaceae	NC_008332.1	*Zea mays subsp. parviglumis* Iltis & Doebley
Poaceae	NC_008360.1	*Sorghum bicolor* (L.) Moench
Poaceae	NC_008362.1	*Tripsacum dactyloides* (L.) L.
Poaceae	NC_031164.1	*Saccharum officinarum* L.

3.2.5. Structural comparison of mitochondrial genomes in *C. breviculmis* and *C. pseudochinensis*

Pairwise sequence alignment of the mitochondrial genomes of *C. breviculmis* and *C. pseudochinensis* was performed using progressiveMauve software (v2.4.0; Darling et al., 2010) to examine structural variations and sequence similarity.

3.2.6. Maximum likelihood phylogeny tree of *Carex* and related taxa

Phylogenetic analysis was conducted using a mitochondrial proteome dataset comprising six Cyperaceae taxa, including *C. pseudochinensis*, along with representatives from other Poales and eudicots as outgroups (Angiosperm Phylogeny Website v14; http://www.mobot.org/MOBOT/research/APweb/) (Table 3.3). Amino acid sequences were clustered into 33 homologous groups using Orthofinder (v2.5.5; Emms & Kelly, 2019) and aligned with the MUSCLE algorithm (v3.8.1551; Edgar, 2004). The alignment matrices were used to infer the phylogeny of 10 species using IQ-TREE (v2.3.6; Minh et al., 2020), applying the partition model (Chernomor et al., 2016) and the ultrafast bootstrap option (Hoang et al., 2018). The final phylogenetic

tree was visualized using MEGA (v11.0.13; Kumar et al., 2018), and a gene presence/absence table was generated using the ggtree R package (v3.6.0; Yu et al., 2017; R Core Team 2023).

Table 3.3. Species and NCBI accessions of mitochondrial genomes included in the phylogenetic analysis

Family	GenBank no.	Species
Cyperaceae	NC_087862	*Carex pseudochinensis* H. Lév. & Vaniot
Cyperaceae	NC_068626	*Carex breviculmis* R.Br.
Cyperaceae	NC_058697	*Cyperus esculentus* L.
Cyperaceae	NC_068217	*Rhynchospora tenuis* Link
Cyperaceae	NC_068216	*Rhynchospora pubera* (Vahl) Boeckeler
Cyperaceae	NC_068215	*Rhynchospora breviuscula* H.Pfeiff.
Juncaceae	NC_069588	*Juncus effusus* L.
Juncaceae	NC_069587	*Luzula sylvatica* (Huds.) Gaudin
Poaceae	NC_066488	*Oryza sativa* L.
Brassicaceae	NC_037304	*Arabidopsis thaliana* (L.) Heynh.

3.3. Results

3.3.1. DNA sequencing and K-mer analysis

C. pseudochinensis was sequenced using Oxford Nanopore Technologies (ONT) and Illumina platforms. ONT sequencing generated a total read base of 2.09 Gbp, consisting of 355,010 reads. The maximum read length reached 117,698 bp, with an N50 value of 18,855 bp (Table 3.4). Illumina sequencing produced high quality short read data, generating approximately 3.16 Gbp from 42 million reads, which contributed to the enhancement of assembly accuracy.

Genome size was estimated to be approximately 98 Mb based on K-mer analysis using Jellyfish. Using the ONT long read data, the calculated coverage was 21.29x, which is lower than the 30x recommended by POLAP. Therefore, all long read data were used without subsampling, and the whole-genome assembly was performed with Flye software (Fig. 3.2A).

Table 3.4. Summary of DNA sequencing data for *C. pseudochinensis*

Species	Sequencing methods	Total read bases (Gbp)	Reads	Min length (bp)	Mean length (bp)	Max length (bp)	N50 (bp)
C. pseudochinensis	ONT	2.09	355010	34	6,097.51	117,698	18,855
	Illumina	3.16	20978104/20978104	N/A	N/A	N/A	N/A

* N/A: Illumina sequencing with a 500 bp insert size and generated 150 bp paired-end reads.

3.3.2. Selection of seed contigs for mitochondrial genome assembly

Mitochondrial DNA seed contigs were selected based on connectivity within the genome assembly graph, copy number, and the presence of organelle genes. Tables 3.5−3.6 provide detailed characteristics of the contigs assembled using Flye. Each column represents the following: the unique identifier assigned to each contig during the Flye assembly process (a; Contig ID), the estimated copy number of the contig, reflecting its relative abundance (b; Copy numbers of the contig), the sequence ID representing the edges of the contig in the assembly graph (c; Edge sequence ID), the total number of mitochondrial genes identified on the contig (d; Number of mitochondrial genes on the contig sequence), and the total number of plastid genes identified on the contig (e; Number of plastid genes on the contig sequence). Copy numbers were evaluated using Flye (v2.9.2; Kolmogorov et al., 2019), which provides multiplicity values—an important metric indicating the replication level of specific contigs within the genome. Multiplicity values were a critical criterion for distinguishing organellar genome contigs, with contigs showing copy numbers greater than 1 being considered organellar in origin and selected for subsequent steps.

A) Contigs with mitochondrial gene annotations exceeding plastid gene annotations

Initially, 11 contigs with more mitochondrial gene annotations than plastid gene annotations were selected. Copy number analysis revealed that seven of these contigs had copy numbers between 2 and 3, suggesting a mitochondrial origin. In contrast, the remaining four contigs exhibited copy numbers of 1 or very high values, indicating possible origins from nuclear or plastid genomes; these were excluded from further analysis (Table 3.5). In Table 3.5, contigs that met the first selection criterion are marked with a black "X," while the final selected contigs are highlighted in red.

B) Contigs with similar ratios of mitochondrial and plastid genes or slightly plastid-biased

Five additional contigs were selected where the ratio of mitochondrial to plastid gene annotations was similar or slightly biased toward plastid annotations (Table 3.6). These contigs had copy numbers ranging from 2 to 3, indicating a mitochondrial genome origin. In Table 3.6, the selected contigs are marked with a red "X" for identification.

C) Contigs with no organelle gene annotations

Two contigs without organelle gene annotations but demonstrating high connectivity within the genome assembly graph were also selected (Fig. 3.2B; Table 3.7, edge_2813 and edge_2044). These contigs were deemed critical for maintaining structural continuity in mitochondrial genome assembly. However, edge_1489 (Fig. 3.2B), despite meeting the connectivity criterion, was excluded from selection as a seed contig due to its significantly lower coverage (3x) compared to other contigs.

In total, 14 mitochondrial genome seed contigs were selected (Table 3.7). Long-read raw data with sequences exceeding 3 kb in length were mapped onto these seed contigs, resulting in 1,667 intra-contig mapped reads and 42 inter-contig mapped reads. These data were used to assemble the mitochondrial genome.

Table 3.5. Summary of contigs and organelle annotation counts in Flye whole-genome assembly

Contig[a]	Length	Copy[b]	Edge[c]	MT[d]	PT[e]	Selected
contig_369	129614	3	369	11	1	X
contig_2045	385224	2	2045	11	2	X
contig_368	100315	2	368	7	0	X
contig_2865	11463	95	2865	6	4	X
contig_367	60207	3	367	4	1	X
contig_2811	79923	3	2811	4	0	X
contig_2046	129919	2	2046	3	2	X
contig_9	384584	1	9	2	0	X
contig_1346	3250	172	1346	2	1	X
contig_2812	47958	3	2812	2	1	X
contig_1883	231324	1	1883	1	0	X
contig_2864	101257	55	2864	35	48	
contig_1347	13554	107	1347	6	8	
contig_1348	10081	52	1348	4	5	
contig_2866	14458	82	2866	3	6	
contig_2479	193764	1	2479	2	2	
contig_479	1226336	1	479	1	1	
contig_648	310328	2	648	1	1	
contig_1148	353948	1	1148	1	1	
contig_1149	43202	1	1149	1	1	
contig_2003	47220	2	2003	1	1	
contig_2632	223639	1	2632	1	1	
contig_2649	597676	1	2649	1	1	
contig_1340	7685	103	1340	1	1	
contig_35	384319	1	35	0	3	
contig_86	234633	1	86	0	1	
contig_162	443777	1	162	0	1	
contig_217	154080	2	217	0	1	
contig_230	228764	1	230	0	1	
contig_262	354456	1	262	0	3	
contig_306	56005	1	306	0	1	
contig_321	217888	1	321	0	1	
contig_342	192186	1	342,-341	0	10	
contig_355	64330	1	355	0	1	
contig_413	368475	1	413	0	2	
contig_495	162188	1	495	0	5	
contig_543	128600	1	543	0	1	
contig_598	186435	1	598	0	8	
contig_616	268419	1	616	0	2	
contig_695	177380	1	695	0	1	
contig_779	238875	1	779	0	1	
contig_814	233824	1	814	0	4	
contig_851	209075	1	851	0	1	
contig_1109	297420	1	1109	0	3	
contig_1221	133096	2	1221	0	3	
contig_1339	28994	1	-28,651,339	0	2	
contig_1485	185817	1	1485	0	1	
contig_1576	259904	1	1576	0	10	
contig_1597	307220	1	1597	0	1	
contig_1608	782773	1	1608	0	1	
contig_1803	71622	1	1803	0	2	
contig_1816	46768	1	1816	0	1	
contig_2175	304516	1	2175	0	1	
contig_2187	342701	1	2187	0	2	
contig_2201	227883	1	2201	0	1	
contig_2379	251347	1	-23,782,379	0	1	
contig_2426	136198	1	2426	0	1	
contig_2626	143122	1	2626	0	1	
contig_2652	137942	1	2652	0	1	
contig_2785	102637	2	2785	0	1	
contig_2852	238213	1	2852	0	1	
contig_2859	596014	1	2859	0	1	
contig_2868	305717	1	2868	0	1	
contig_653	39820	1	653	0	1	
contig_1337	5850	41	1337	0	2	
contig_2878	8190	1	2878	0	3	

Table 3.6. Selected contigs with mitochondrial and plastid gene ratios similar or slightly plastid-biased

Contig[a]	Length	Copy[b]	Edge[c]	MT[d]	PT[e]	Selected
contig_369	129614	3	369	11	1	
contig_2045	385224	2	2045	11	2	
contig_368	100315	2	368	7	0	
contig_2865	11463	95	2865	6	4	
contig_367	60207	3	367	4	1	
contig_2811	79923	3	2811	4	0	
contig_2046	129919	2	2046	3	2	
contig_9	384584	1	9	2	0	
contig_1346	3250	172	1346	2	1	
contig_2812	47958	3	2812	2	1	
contig_1883	231324	1	1883	1	0	
contig_2864	101257	55	2864	35	48	
contig_1347	13554	107	1347	6	8	
contig_1348	10081	52	1348	4	5	
contig_2866	14458	82	2866	3	6	
contig_2479	193764	1	2479	2	2	
contig_479	1226336	1	479	1	1	
contig_648	310328	2	648	1	1	X
contig_1148	353948	1	1148	1	1	
contig_1149	43202	1	1149	1	1	
contig_2003	47220	2	2003	1	1	X
contig_2632	223639	1	2632	1	1	
contig_2649	597676	1	2649	1	1	
contig_1340	7685	103	1340	1	1	
contig_35	384319	1	35	0	3	
contig_86	234633	1	86	0	1	
contig_162	443777	1	162	0	1	
contig_217	154080	2	217	0	1	X
contig_230	228764	1	230	0	1	
contig_262	354456	1	262	0	3	
contig_306	56005	1	306	0	1	
contig_321	217888	1	321	0	1	
contig_342	192186	1	342,-341	0	10	
contig_355	64330	1	355	0	1	
contig_413	368475	1	413	0	2	
contig_495	162188	1	495	0	5	
contig_543	128600	1	543	0	1	
contig_598	186435	1	598	0	8	
contig_616	268419	1	616	0	2	
contig_695	177380	1	695	0	1	
contig_779	238875	1	779	0	1	
contig_814	233824	1	814	0	4	
contig_851	209075	1	851	0	1	
contig_1109	297420	1	1109	0	3	
contig_1221	133096	2	1221	0	3	X
contig_1339	28994	1	-28,651,339	0	2	
contig_1485	185817	1	1485	0	1	
contig_1576	259904	1	1576	0	10	
contig_1597	307220	1	1597	0	1	
contig_1608	782773	1	1608	0	1	
contig_1803	71622	1	1803	0	2	
contig_1816	46768	1	1816	0	1	
contig_2175	304516	1	2175	0	1	
contig_2187	342701	1	2187	0	2	
contig_2201	227883	1	2201	0	1	
contig_2379	251347	1	-23,782,379	0	1	
contig_2426	136198	1	2426	0	1	
contig_2626	143122	1	2626	0	1	
contig_2652	137942	1	2652	0	1	
contig_2785	102637	2	2785	0	1	X
contig_2852	238213	1	2852	0	1	
contig_2859	596014	1	2859	0	1	
contig_2868	305717	1	2868	0	1	
contig_653	39820	1	653	0	1	
contig_1337	5850	41	1337	0	2	
contig_2878	8190	1	2878	0	3	

Table 3.7. Mitochondrial candidate contigs

Contig	Length	Mitochondrial Annotation	Copy Number	Graph Connectivity
edge_2045	385224	X	X	X
edge_648	310328		X	
edge_217	154080		X	
edge_1221	133096		X	
edge_2046	129919	X	X	X
edge_369	129614	X	X	X
edge_2785	102637		X	
edge_368	100315	X	X	X
edge_2811	79923	X	X	X
edge_367	60207	X	X	X
edge_2812	47958	X	X	X
edge_2003	47220		X	
edge_2813	20176			X
edge_2044	8698			X

3.3.3. Mitochondrial genome assembly

The organelle genome assembly generated using the Flye (v2.9.2; Kolmogorov et al., 2019) yielded 13 contigs (Fig. 3.2C). Four of these contigs were found to contain mitochondrial genes based on annotation using GeSeq (v2.03; Tillich et al., 2017) with the *C. breviculmis* (Xu et al., 2023b) as a reference. Among these, three contigs (edges 7, 8, and 13) were linked and had high sequence coverage (>10x), while the remaining independent contig (edge 6) had low sequence coverage (4x). These three fragments were extracted as a circular DNA sequence using Bandage software (v0.8.1; Wick et al., 2015), and subsequently polished with short-read data using FMLRC (v1.0.0; Mak et al., 2023; Zhou et al., 2023).

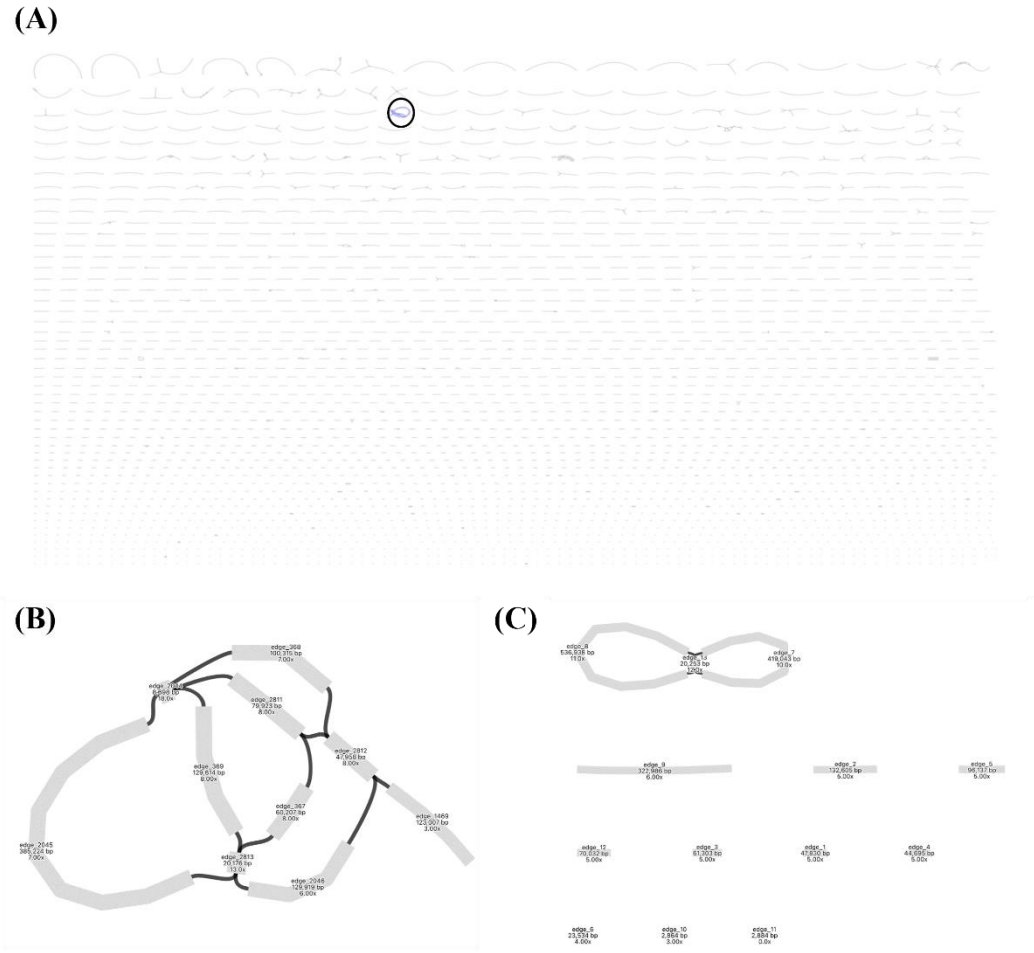


Fig. 3.2. Bandage graphs from each step of the mitochondrial genome assembly of *C. pseudochinensis* using POLAP. (A) Graphs of a long-read whole-genome assembly. (B) A graph of potential mitochondrial-derived contigs, which is indicated as a circle in the whole-genome assembly graphs (A). The contigs, except for edge 1469, were used as seed contigs for the organelle genome assembly.

Fig. 3.2. (Continued) Edge 1469 was excluded as it did not have any gene annotations. (C) Graphs of mitochondrial-genome assembly based on the seed contigs of the graph in (B). All figures were generated using Bandage (Wick et al., 2015).

3.3.4. Polishing and verification

The assembled mitochondrial genome was polished using short-read data to enhance its accuracy. Furthermore, the consistency of sequencing depth across the mitochondrial genome assembly was verified by mapping both short-read (Fig. 3.3A) and long-read data (Fig. 3.3B) using Minimap2 (v2.24; Li, 2018) and Geneious Prime (v2024.0.5; Kearse et al., 2012).

The final assembled mitochondrial genome is 997,628 base pairs long, with a GC content of 41%, comprising 28% A, 20% C, 20% G, and 32% T. The genome contains 57 genes in total, including 31 protein-coding genes, 6 ribosomal RNAs, and 20 tRNAs. Notably, the *nad1*, *nad2*, and *nad5* underwent trans-splicing, and the number of trans-spliced elements in each gene was consistent with that of *C. breviculmis* (*nad1*, 3; *nad2*, 4; *nad5*, 3). The final circular genome map for the *C. pseudochinensis* mitochondrial genome was visualized using OGDRAW software (v1.3.1; Greiner et al., 2019) (Fig 3.4).

The mitochondrial genome also includes two large regions of 536.94 kbp and 419.04 kbp, along with direct repeat regions. These

repeat regions consist of two pairs, measuring 20,276 bp and 8,742 bp, respectively, and include the *rrn18* and *rrn26*. Repeats were detected using the Repeat Finder tool in Geneious Prime (v2024.0.5; Biomatters Ltd., Auckland, New Zealand; Kearse et al., 2012) with the following settings: Minimum repeat length: 1,000 and Maximum repeats (approximate) to find: 10,000.

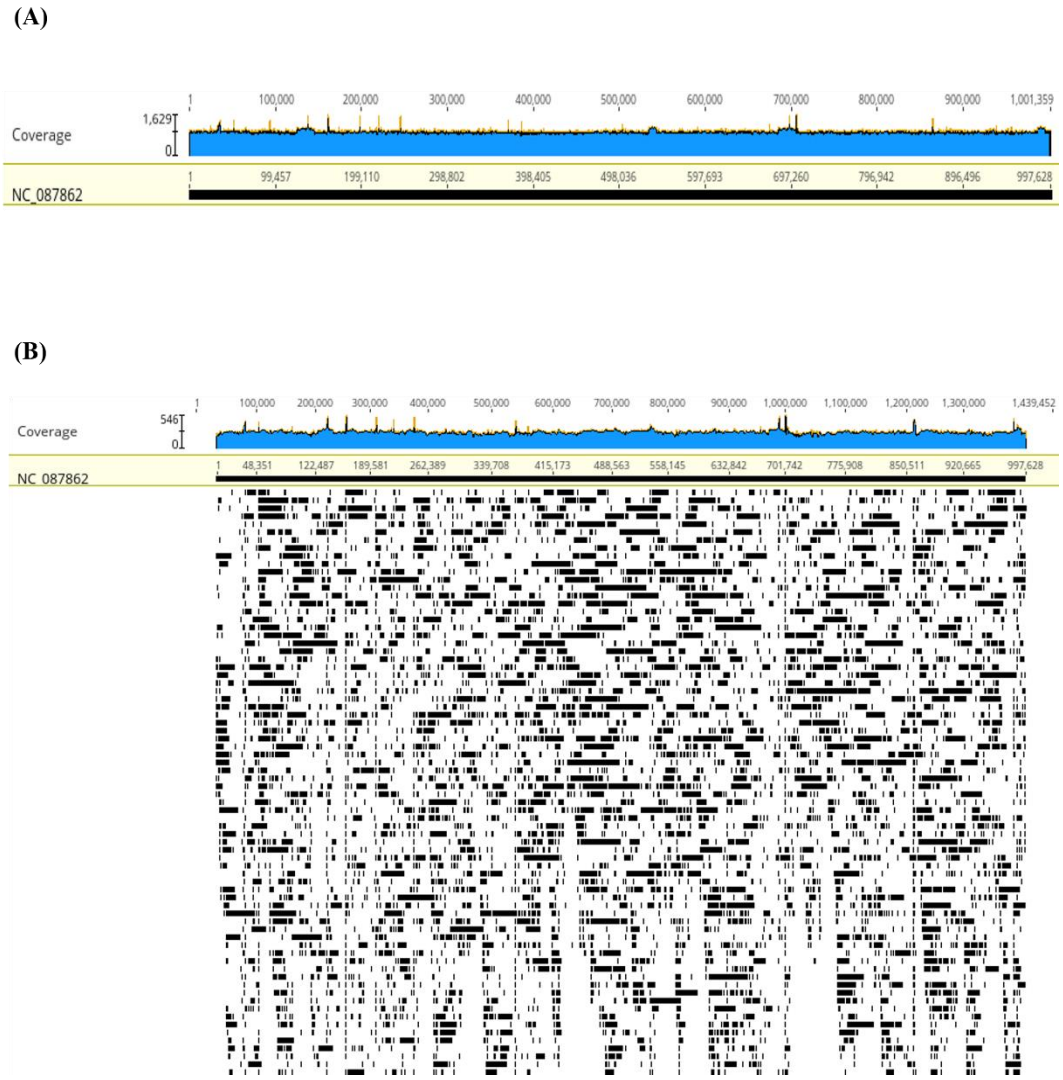


Fig. 3.3. Sequencing coverage of short-reads (A) and long-reads (B) against the final assembled mitochondrial genome was analyzed using Minimap2 (Li, 2018) through the Geneious Prime (v2024.0.5; Biomatters Ltd., Auckland, New Zealand; Kearse et al., 2012). Only the upper part of the mapped long reads was displayed.

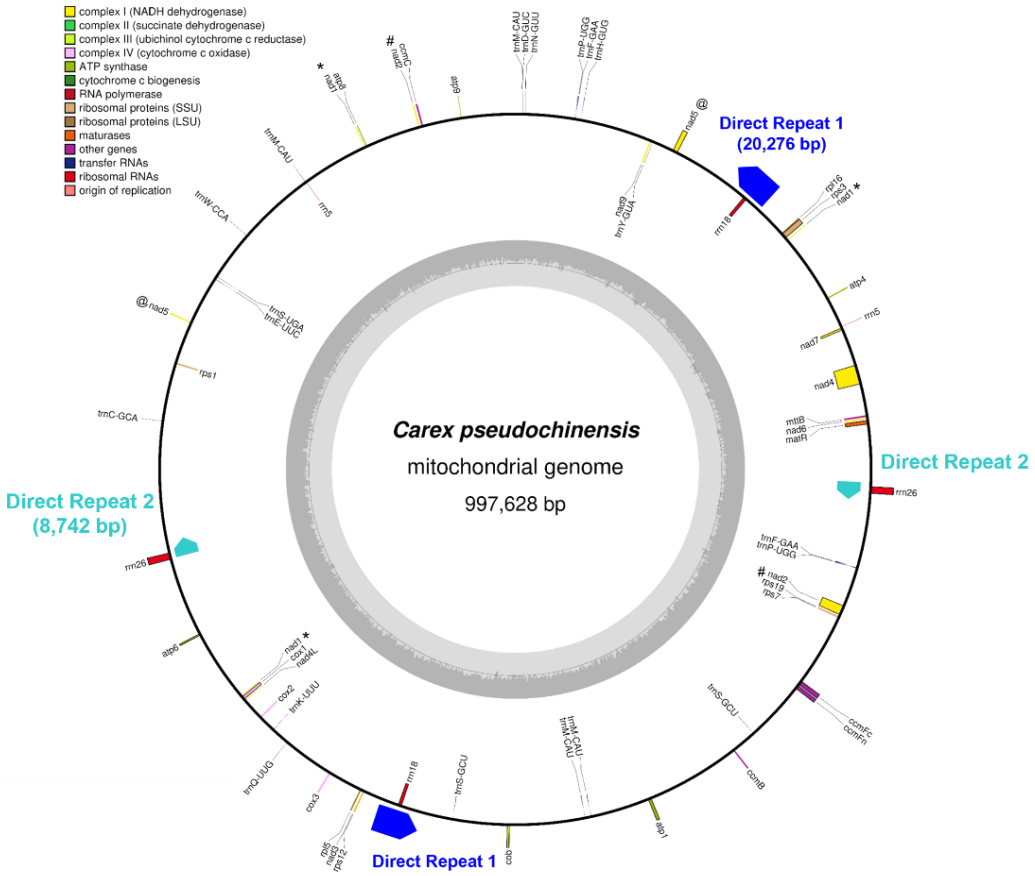


Fig. 3.4. Circular genome map of the mitochondrial DNA of *C. pseudochinensis*. Three trans-spliced genes, *nad1*, *nad2*, and *nad5*, are indicated by different special characters, *, #, and @, respectively. Blue arrows indicate direct repeats.

3.3.5. Phylogenetic analysis and structural divergence of *Carex* mitochondrial genomes

The phylogenetic tree of nine monocot mitochondrial genomes was reconstructed, with *Arabidopsis thaliana* used as an outgroup (Fig. 3.5; Table 3.3). Among the six Cyperaceae species, *C. pseudochinensis* and *C. breviculmis* clustered closely, indicating a close phylogenetic relationship based on 33 protein-coding genes, with strong support values across all nodes. For stability in the phylogenetic analysis, genes shared by five or fewer species (*rpl2*, *rpl10*, *rpl14*, and *rps2A*) were excluded, and one duplicate gene per group was randomly selected for inclusion. Despite this close phylogenetic relationship, the mitochondrial genome structures of *C. pseudochinensis* and *C. breviculmis* exhibited distinct differences in gene arrangement, composition, and the presence of duplicated homologous genes (Fig. 3.6).

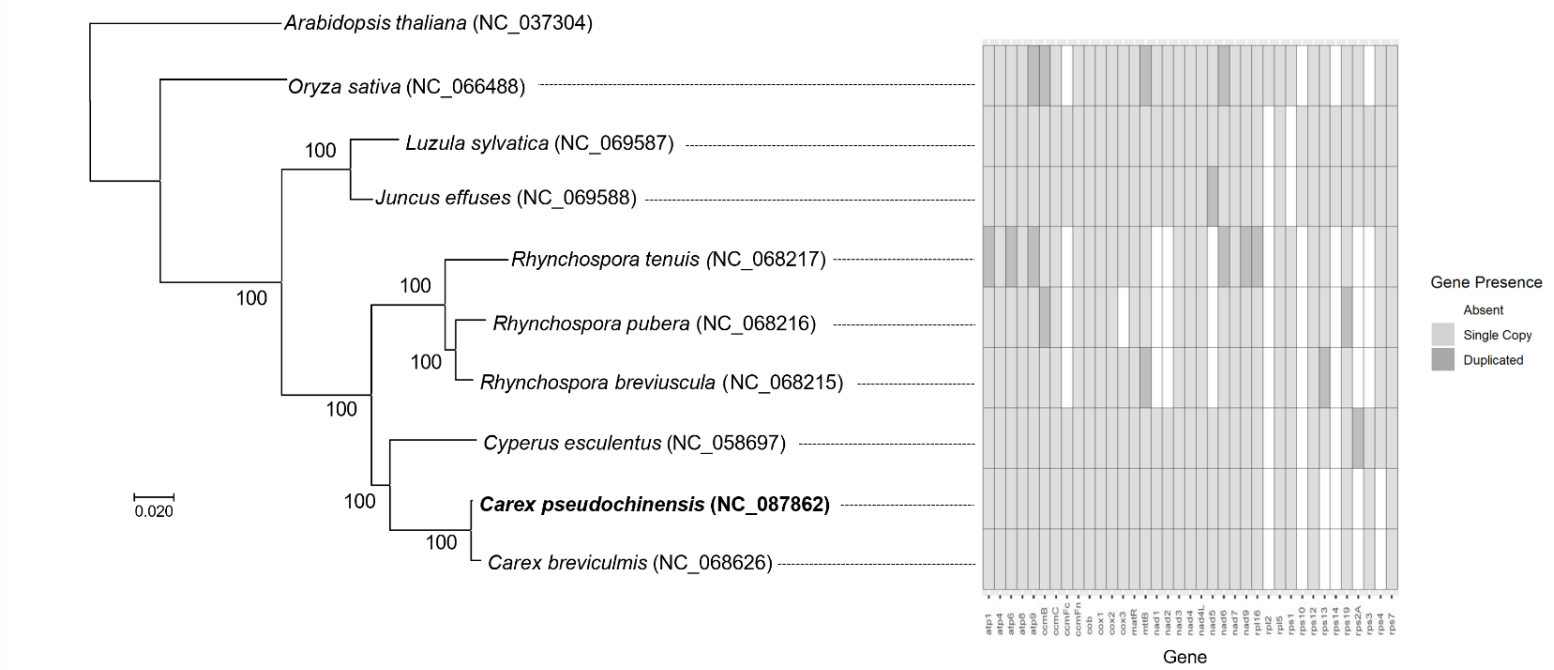


Fig. 3.5. Maximum-likelihood tree based on amino acid sequences of 33 coding genes with presence (gray) / absence (white)matrix. Numbers above nodes indicate bootstrap support (1,000 replicates).

Fig. 3.5. (Continued) Dark gray indicates the presence of two or more duplicated homologous genes in each taxon, while light gray represents single genes. Optimal models for each gene are listed in Table 3.8.

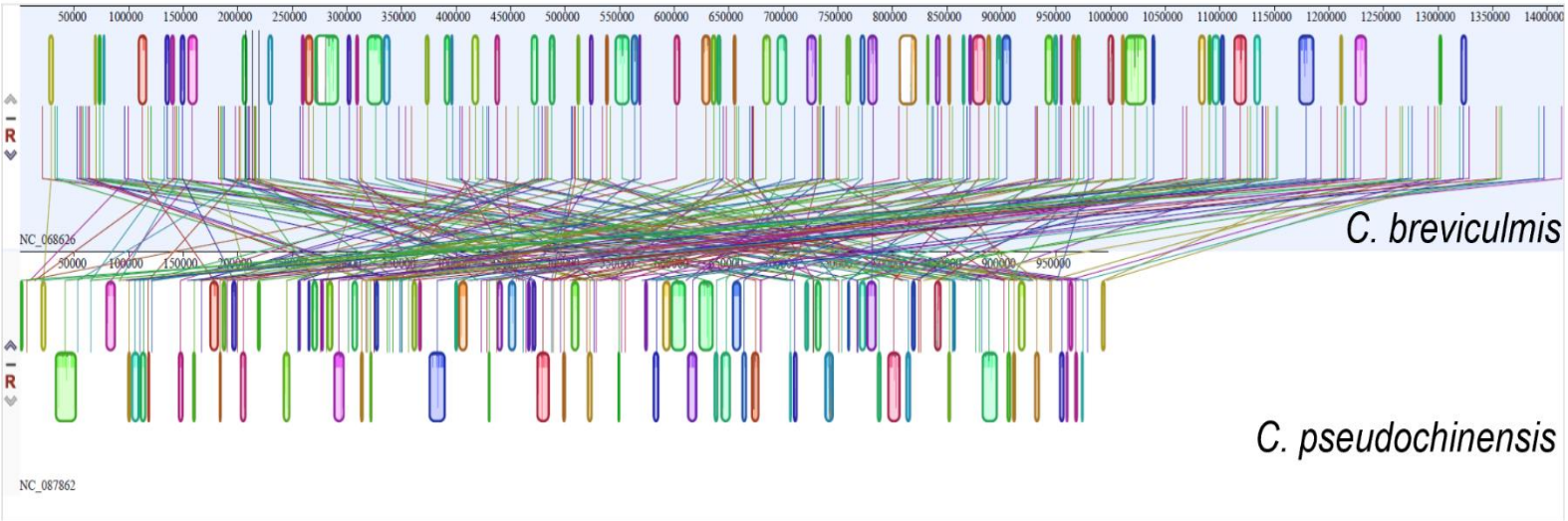


Fig. 3.6. Synteny block comparison between *C. pseudochinensis* (NC_087862) and *C. breviculmis* (NC_068626) using progressiveMauve (Darling et al., 2010)

Table 3.8. The best amino-acid substitution models for 33 coding genes used in the phylogenetic analyses of IQ-TREE (Minh et al., 2020) with the partition model (Chernomor et al., 2016)

Number	Gene	Best model	Number	Gene	Best model
1	*atp9*	cpREV	18	*nad3*	Q.bird+G4
2	*ccmB*	Q.bird+I	19	*rps12*	cpREV
3	*atp6*	VT+I	20	*ccmC*	FLAVI+R2
4	*nad6*	cpREV+R2	21	*ccmFn*	Q.mammal+G4
5	*mttB*	cpREV+G4	22	*nad4L*	cpREV
6	*rpl16*	JTT	23	*atp4*	Q.mammal+I
7	*nad9*	Q.bird+G4	24	*rps19*	JTTDCMut+G4
8	*atp1*	Q.plant+I+R2	25	*cox3*	cpREV
9	*cox2*	JTT+R2	26	*nad5*	cpREV+F+G4
10	*rps7*	HIVb+G4	27	*rps4*	cpREV+F+R2
11	*cox1*	mtZOA+R2	28	*rps13*	cpREV+R2
12	*cob*	cpREV+I	29	*nad1*	cpREV
13	*rpl5*	cpREV+I	30	*nad2*	HIVw+F
14	*nad4*	HIVw+F+G4	31	*rps1*	cpREV+G4
15	*matR*	cpREV+G4	32	*rps3*	cpREV+G4
16	*nad7*	Q.bird	33	*ccmFc*	Q.mammal
17	*atp8*	cpREV			

3.4. Discussion

The mitochondrial genome of *C. pseudochinensis* was successfully assembled using the POLAP. The complex structure and polymorphism of plant mitochondrial genomes, especially those of the Cyperaceae, present a challenge for constructing "master" genome sequences. Our initial attempts to assemble the *C. pseudochinensis* mitochondrial genome using the ptGAUL (v1.0.5; Zhou et al., 2023), with *C. breviculmis* (Xu et al., 2023b) as a reference, were unsuccessful, likely due to structural and sequence differences between the two species (Fig. 3.6; Fig. 3.7). This outcome aligns with previous observations that assembling Cyperaceae mitochondrial genomes is particularly challenging. The assembled mitochondrial genome for *C. pseudochinensis* was validated by ensuring uniform sequencing coverage across the genome, confirmed through the mapping of short- and long-read sequences using Minimap2 (v2.24; Li, 2018) and Geneious Prime (v2024.0.5; Kearse et al., 2012) (Fig. 3.4A). The mitochondrial genome of *C. pseudochinensis* generated in this study provides a foundation for understanding structural evolution in the genus *Carex* and offers valuable insights for delineating species boundaries.

In conclusion, POLAP presents a valuable alternative for assembling plant organelle genomes with high structural complexity and numerous repeat sequences. While ptGAUL is generally effective for chloroplast genome assembly, it relies on the presence of a reference genome to guide its assembly. This dependency becomes a significant limitation in the case of *Carex*, where substantial structural variation exists even within the same genus. Such extensive variations make it challenging to assemble chloroplast genomes accurately using ptGAUL, as any reference genome struggles to accommodate the unique structural rearrangements found in *Carex*. By contrast, POLAP circumvents this issue through a reference-free approach, constructing its assembly solely from the organism's own sequence data. This capability makes POLAP particularly advantageous for *Carex* chloroplast genomes, where high structural variability would otherwise hinder assembly. Overall, POLAP stands out as a robust tool for advancing future studies on mitochondrial and chloroplast genomes in this genus.

Fig. 3.7. Assembly graph of the mitochondrial genome of *C. pseudochinensis* generated using the ptGAUL (v1.0.5; Zhou et al., 2023) approach, with the *C. breviculmis* mitochondrial genome as a reference. This method did not yield a single master genome sequence for *C. pseudochinensis*, likely due to structural and sequence dissimilarities. The figure was produced with Bandage (Wick et al., 2015).

4. Conclusion

This research contributes foundational knowledge regarding the structural evolution and instability of chloroplast genomes in *Carex*. The findings illustrate the high levels of repetitive sequences, particularly short dispersed repeats (SDRs) and long repeat sequences, which are major drivers of genomic rearrangements and instability. These structural complexities have significant implications for understanding the adaptive evolution of *Carex* and its ecological diversification within Cyperaceae. Moreover, the use of synteny analysis to incorporate larger genomic structures into the phylogenetic framework has proven effective in elucidating evolutionary relationships that are not easily captured by analyses limited to individual genes or smaller genomic regions.

The mitochondrial genome analysis of *C. pseudochinensis* further underscores the high structural plasticity within the genus, reflecting the challenges of assembling organelle genomes with substantial repetitive content. The successful application of the POLAP for mitochondrial genome assembly demonstrates the necessity of using reference-free methods for complex taxa like *Carex*, where high structural variability can complicate reference-based assembly approaches. These

advancements pave the way for more accurate genomic studies and improve our understanding of evolutionary adaptation and speciation processes within Cyperaceae.

In conclusion, this study highlights the need for further research into the diverse subgenera of *Carex*, especially beyond subgenera *Siderosticta*, *Vignea*, and *Carex*. The current research was limited to these subgenera due to their prevalence in Korea, as depicted in Figure 1.2. However, to gain a more comprehensive understanding of the evolutionary dynamics and structural variations within *Carex*, it is essential to include species from the remaining subgenera and a wider range of clades. Expanding the study to other subgenera, such as *Psyllophorae*, *Euthyceras*, and *Uncinia*, will provide valuable insights into the full spectrum of genetic diversity and evolutionary adaptations within this complex genus.

Overall, this study provides a crucial stepping stone for future research into the molecular evolution of *Carex*. By incorporating additional subgenera and employing both molecular and morphological approaches, future studies can refine the phylogenetic framework of *Carex* and contribute to a more nuanced understanding of its structural

diversity, evolutionary trajectories, and ecological significance.

LITERATURE CITED

Akiyama, S. (1955). *Carices of the Far Eastern region of Asia*. Sapporo: The Faculty of Science, Hokkaido University.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research, 25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N., & Christin, P. A. (2009). Phylogenomics of C(4) photosynthesis in sedges (Cyperaceae): Multiple appearances and genetic convergence. *Molecular Biology and Evolution, 26*(8), 1909–1919. https://doi.org/10.1093/molbev/msp103

Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M. L., & Ruhlman, T. A. (2016). Variable presence of the inverted repeat and plastome stability in *Erodium*. *Annals of Botany, 117*(7), 1209–1220. https://doi.org/10.1093/aob/mcw065

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research, 27*(8), 1767–1780. https://doi.org/10.1093/nar/27.8.1767

Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2019). tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *bioRxiv*. https://doi.org/10.1101/614032

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace-aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, *65*(6), 997–1008. https://doi.org/10.1093/sysbio/syw037

Chevigny, N., Schatz-Daas, D., Lotfi, F., & Gualberto, J. M. (2020). DNA repair and the stability of the plant mitochondrial genome. *International Journal of Molecular Sciences*, *21*(1), Article 328. https://doi.org/10.3390/ijms21010328

Cho, Y. H., Kim, J. H., & Park, S. H. (2016). *Grasses and sedges in South Korea*. Seoul: Geobook. (in Korean)

Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., & Jansen, R. K. (2006). The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution*, *23*(11), 2175-2190. https://doi.org/10.1093/molbev/msl089

Dai, L. K., Liang, S. Y., Zhang, S. R., Tang, Y. C., Koyama, T., Tucker, G. C., Simpson, D. A., Noltie, H. J., Strong, M. T., Bruhl, J. J., Wilson, K. L., & Muasya, A. M. (2010). Cyperaceae. In Z. Y. Wu, P. H. Raven, & D. Y. Hong (Eds.), *Flora of China* (Vol. 23). Science Press; Missouri Botanical Garden Press.

Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology*, *17*(1), 1-29. https://doi.org/10.1186/s13059-016-1004-2

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*(6), e11147. https://doi.org/10.1371/journal.pone.0011147

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340

Egorova, T. V. (1999). *The sedges (Carex L.) of Russia and adjacent states (with the limits of the former USSR)*. St.-Petersburg: St.-

Petersburg State Chemical-Pharmaceutical Academy; St. Louis: Missouri Botanical Garden Press.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, *39*(Web Server issue), W29-W37. https://doi.org/10.1093/nar/gkr367

Fischer, A., Dotzek, J., Walther, D., & Greiner, S. (2022). Graph-based models of the Oenothera mitochondrial genome capture the enormous complexity of higher plant mitochondrial DNA organization. *NAR Genomics and Bioinformatics, 4*(2), lqac027. https://doi.org/10.1093/nargab/lqac027

Flora of China Editorial Committee. (n.d.). *Carex duriuscula*. In Z. Y. Wu & P. H. Raven (Eds.), *Flora of China* (Vol. 23, p. 453). Retrieved November 11, 2024, from http://www.efloras.org

Flora of China Editorial Committee. (n.d.). *Carex paxii*. In Z. Y. Wu & P. H. Raven (Eds.), *Flora of China* (Vol. 23, p. 446). Retrieved November 11, 2024, from http://www.efloras.org

Flora of China Editorial Committee. (n.d.). *Carex peiktusani*. In Z. Y. Wu & P. H. Raven (Eds.), *Flora of China* (Vol. 23, p. 308). Retrieved November 11, 2024, from http://www.efloras.org

Flora of China Editorial Committee. (n.d.). *Carex siderosticta*. In Z. Y. Wu & P. H. Raven (Eds.), *Flora of China* (Vol. 23, p. 347). Retrieved November 11, 2024, from http://www.efloras.org

Flora of North America Editorial Committee. (n.d.). *Carex duriuscula*. In Flora of North America Editorial Committee (Eds.), *Flora of North*

America (Vol. 23, pp. 303–305). Retrieved November 11, 2024, from http://www.efloras.org

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The world checklist of vascular plants, a continuously updated resource for exploring global plant diversity. *Scientific Data, 8*(1), 215. https://doi.org/10.1038/s41597-021-00997-6

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

Henry, R. J. (2022). Progress in plant genome sequencing. *Applied Biosciences*, *1*(2), 113–128. https://doi.org/10.3390/applbiosci1020008

Hipp, A. L. (2007). Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). *Evolution, 61*(9), 2175–2194. https://doi.org/10.1111/j.1558-5646.2007.00183.x

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). Ufboot2: Improving the ultrafast bootstrap approximation. *Mol Biol Evol*, *35*(2), 518–522. doi:10.1093/molbev/msx281

Hoshino, T., Masaki, T., & Nishimoto, M. (2011). *Illustrated sedges of Japan*. Tokyo: Heibonsha Ltd.

Kang, M., Chanderbali, A., Lee, S., Soltis, D. E., Soltis, P. S., & Kim, S. (2023). High-molecular-weight DNA extraction for long-read sequencing of plant genomes: An optimization of standard methods. *Applications in Plant Sciences*, *11*(3), e11528. https://doi.org/10.1002/aps3.11528

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and

analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keeling, P. J. (2010). The endosymbiotic origin, diversification, and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1541), 729–748. https://doi.org/10.1098/rstb.2009.0103

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664. https://doi.org/10.1101/gr.229202

Kim, K. O., Hong, S. H., Lee, Y. H., Na, C. S., Kang, B. H., & Son, Y. W. (2009). Taxonomic status of endemic plants in Korea. *Journal of Ecology and Field Biology*. Open Access Library.

Kim, K. O., Hong, S. H., Lee, Y. H., Na, C. S., Kang, B. H., & Son, Y. W. (2010). Distribution of vascular plants at the ecological landscape conservation area Heoninlleung in Seoul. *Korean Journal of Plant Resources*, *23*(1), 60–78.

Knoop, V., Volkmar, U., Hecht, J., & Grewe, F. (2010). Mitochondrial genome evolution in the plant lineage. In F. Kempken (Ed.), *Plant mitochondria* (pp. 3–29). https://doi.org/10.1007/978-0-387-89781-3_1

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Koyama, T. (1961). Classification of the family Cyperaceae (1). *Journal of the Faculty of Science, University of Tokyo, Section III, Botany, 8*, 37–148.

Koyama, T. (1962). Classification of the family Cyperaceae (2). *Journal of the Faculty of Science, University of Tokyo, Section III, Botany, 8*, 149–278.

Kozik, A., Rowan, B. A., Lavelle, D., Berke, L., Schranz, M. E., Michelmore, R. W., & Christensen, A. C. (2019). The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genetics*, *15*(8), e1008373. https://doi.org/10.1371/journal.pgen.1008373

Kreczetovicz, V. (1935). *Carex*. In V. L. Komarov (Ed.), *Flora URSS* (Vol. 3). Acad. Sci. URSS, Leningrad.

Kükenthal, G. (1909). Cyperaceae-Caricoidae. In A. Engler (Ed.), *Das Pflanzenreich* (IV, 20, Heft 38). Leipzig: W. Engelmann.

Kükenthal, G. (1935). Cyperaceae-Scripoideae-Cypereae. In A. Engler (Ed.), *Das Pflanzenreich* (IV, 20, Heft 101), pp. 1-160. Berlin: Engelmann.

Kükenthal, G. (1936). Cyperaceae-Scripoideae-Cypereae. In A. Engler (Ed.), *Das Pflanzenreich* (IV, 20, Heft 101), pp. 161-671. Berlin: Engelmann.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research, 29*(22), 4633-4642. https://doi.org/10.1093/nar/29.22.4633

Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*(1), 11-16. https://doi.org/10.1093/nar/gkh152

Lee, B. (2021). An integrative study of Cyperaceae: Focused on expressions of floral MADS-box genes and the structural

evolution of chloroplast genomes (Unpublished doctoral dissertation). Sungshin Women's University, Seoul, Korea.

Lee, C., Ruhlman, T. A., & Jansen, R. K. (2020). Unprecedented intraindividual structural heteroplasmy in *Eleocharis* (Cyperaceae, Poales) plastomes. *Genome Biology and Evolution*, *12*(5), 641-655. https://doi.org/10.1093/gbe/evaa076

Lee, H. L., Jansen, R. K., Chumley, T. W., & Kim, K. J. (2007). Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Molecular Biology and Evolution*, *24*(5), 1161-1180. https://doi.org/10.1093/molbev/msm036

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics, 34*(18), 3094-3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

López-Juez, E., & Pyke, K. A. (2005). Plastids unleashed: Their development and their integration in plant development. *The Plant Cell*, *17*(7), 1860-1878. https://doi.org/10.1387/ijdb.051997el

Mak, Q. X. C., Wick, R. R., Holt, J. M., & Wang, J. R. (2023). Polishing *de novo* Nanopore assemblies of bacteria and eukaryotes with FMLRC2. *Molecular Biology and Evolution, 40*(3). https://doi.org/10.1093/molbev/msad048

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764-770. https://doi.org/10.1093/bioinformatics/btr011

Márquez-Corro, J. I., Martín-Bravo, S., Jiménez-Mejías, P., Hipp, A. L., Spalink, D., Naczi, R. F. C., Roalson, E. H., Luceño, M., & Escudero, M. (2021). Macroevolutionary insights into sedges (*Carex*: Cyperaceae): The effects of rapid chromosome number evolution on lineage diversification. *Journal of Systematics and Evolution, 59*(4), 776-790. https://doi.org/10.1111/jse.12730

Márquez-Corro, J. I., Martin-Bravo, S., Spalink, D., Luceno, M., & Escudero, M. (2019). Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). *Molecular Phylogenetics and Evolution, 135*, 203-209. https://doi.org/10.1016/j.ympev.2019.03.006

Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. (2012). Holocentric chromosomes: Convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Research*, *20*, 579-593.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530-1534. https://doi.org/10.1093/molbev/msaa015

Mohlenbrock, R. H., & Nelson, P. W. (1999). *Sedges: Carex (The illustrated flora of Illinois)*. Carbondale: Southern Illinois University Press.

Møller, I. M., Rasmusson, A. G., & Van Aken, O. (2021). Plant mitochondria: Past, present and future. *The Plant Journal*, *108*(4), 912-959. https://doi.org/10.1111/tpj.15495

Mower, J. P., & Vickrey, T. L. (2018). Structural diversity among plastid genomes of land plants. In S.-M. Chaw & R. K. Jansen (Eds.), *Advances in botanical research* (Vol. 85, pp. 263-292). Academic Press. https://doi.org/10.1016/bs.abr.2017.11.013

Mower, J. P., Sloan, D. B., & Alverson, A. J. (2012). Plant mitochondrial genome diversity: The genomics revolution. In *Plant genome diversity* (Vol. 1, pp. 123–144). https://doi.org/10.1007/978-3-7091-1130-7_9

Naczi, R. F. C. (2009). Insights on using morphologic data for phylogenetic analysis in sedges (Cyperaceae). *The Botanical Review*, *75*(1), 67–95. https://doi.org/10.1007/s12229-008-9017-5

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., & Kadowaki, K. (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics*, *268*(4), 434–445. https://doi.org/10.1007/s00438-002-0767-1

Ohwi, J. (1936). Cyperaceae Japonicae I: A synopsis of the Caricoidea of Japan, including the Kuriles, Saghalin, Korea, and Formosa. *Memoirs of the College of Science, Kyoto Imperial University. Series B, Biology, 11*, 229–530.

Ohwi, J. (1944). Cyperaceae Japonicae II: A synopsis of the Rhynchosporoideae and Scirpoideae of Japan, including the Kuriles, Saghalin, Korea, and Formosa. *Memoirs of the College of Science, Kyoto Imperial University. Series B, Biology*, *18*, 1–182.

Okonechnikov, K., Golosova, O., Fursov, M., & the UGENE team. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Palmer, J. D., & Thompson, W. F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell, 29*(2), 537–550.

Papanikolaou, N., Trachana, K., Theodosiou, T., Promponas, V. J., & Iliopoulos, I. (2009). Gene socialization: Gene order, GC content

and gene silencing in Salmonella. *BMC Genomics*, *10*, 597. https://doi.org/10.1186/1471-2164-10-597

Richardson, A. O., Rice, D. W., Young, G. J., Alverson, A. J., & Palmer, J. D. (2013). The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology*, *11*(1), 29. https://doi.org/10.1186/1741-7007-11-29

Roalson, E. H., Jiménez-Mejías, P., Hipp, A. L., Benítez-Benítez, C., Bruederle, L. P., Chung, K. S., Escudero, M., Ford, B. A., Ford, K., Gebauer, S., Gehrke, B., Hahn, M., Hayat, M. Q., Hoffmann, M. H., Jin, X. F., Kim, S., Larridon, I., Léveillé-Bourret, É., Lu, Y. F., … Zhang, S. R. (2021). A framework infrageneric classification of *Carex* (Cyperaceae) and its organizing principles. *Journal of Systematics and Evolution*, *59*(4), 726-762. https://doi.org/10.1111/jse.12722

Ruhlman, T. A., Zhang, J., Blazier, J. C., Sabir, J. S. M., & Jansen, R. K. (2017). Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *American Journal of Botany*, *104*(4), 559-572. https://doi.org/10.3732/ajb.1600453

Saxena, R. K., Edwards, D., & Varshney, R. K. (2014). Structural variations in plant genomes. *Briefings in Functional Genomics*, *13*(4), 296-307. https://doi.org/10.1093/bfgp/elu016

Simpson, D. A., & Inglis, C. A. (2001). Cyperaceae of economic, ethnobotanical, and horticultural importance: A checklist. *Kew Bulletin, 56*(2). https://doi.org/10.2307/4110962

Sokoloff, D. D., Remizowa, M. V., Bateman, R. M., & Rudall, P. J. (2018). Was the ancestral angiosperm flower whorled throughout? *American Journal of Botany, 105*(1), 5-15. https://doi.org/10.1002/ajb2.1003

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). *Journal of Korean Forest Society*, *103*(3), 313-320. https://doi.org/10.14578/jkfs.2014.103.3.313

Spalink, D., Drew, B. T., Pace, M. C., Zaborsky, J. G., Li, P., Cameron, K. M., Givnish, T. J., & Sytsma, K. J. (2016). Evolution of geographical place and niche space: Patterns of diversification in the North American sedge (Cyperaceae) flora. *Molecular Phylogenetics and Evolution, 95*, 183-195. https://doi.org/10.1016/j.ympev.2015.09.028

Spalink, D., Pender, J., Escudero, M., Hipp, A. L., Roalson, E. H., Starr, J. R., Waterway, M. J., Bohs, L., & Sytsma, K. J. (2018). The spatial structure of phylogenetic and functional diversity in the United States and Canada: An example using the sedge family (Cyperaceae). *Journal of Systematics and Evolution*, *56*(5), 449-465. https://doi.org/10.1111/jse.12423

Sugiura, M. (2003). History of chloroplast genomics. *Photosynthesis Research*, *76*(1-3), 371-377. https://doi.org/10.1023/A:1024913304263

The Angiosperm Phylogeny Group. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society, 181*(1), 1-20. https://doi.org/10.1111/boj.12385

The Global *Carex* Group, Jiménez-Mejías, P., Hahn, M., Lueders, K., Starr, J. R., Brown, B. H., Chouinard, B. N., Chung, K.-S., Escudero, M., Ford, B. A., Ford, K. A., Gebauer, S., Gehrke, B., Hoffmann, M. H., Jin, X.-F., Jung, J., Kim, S., Luceño, M., Maguilla, E., Martín-Bravo, S., … Roalson, E. H. (2016). Megaphylogenetic specimen-level approaches to the *Carex* (Cyperaceae) phylogeny using ITS, ETS, and *matK* sequences: Implications for

classification. *Systematic Botany, 41*(3), 500–518. https://doi.org/10.1600/036364416X692497

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq: Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, *45*(W1), W6–W11. https://doi.org/10.1093/nar/gkx391

Villaverde, T., Jiménez-Mejías, P., Luceño, M., Waterway, M. J., Kim, S., Lee, B., Rincón-Barrado, M., Hahn, M., Maguilla, E., Roalson, E. H., Hipp, A. L., Wilson, K. L., Larridon, I., Gebauer, S., Hoffmann, M. H., Simpson, D. A., Naczi, R. F. C., Reznicek, A. A., Ford, B. A., … Martín-Bravo, S. (2020). A new classification of *Carex* (Cyperaceae) subgenera supported by a HybSeq backbone phylogenetic tree. *Botanical Journal of the Linnean Society*, *194*(2), 141–163. https://doi.org/10.1093/botlinnean/boaa042

Weng, M. L., Blazier, J. C., Govindu, M., & Jansen, R. K. (2014). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution*, *31*(3), 645–659. https://doi.org/10.1093/molbev/mst257

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of *de novo* genome assemblies. *Bioinformatics*, *31*(20), 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Wicke, S., Schneeweiss, G. M., de Pamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, *76*(3–5), 273–297. https://doi.org/10.1007/s11103-011-9762-4

Wu, H., Li, D. Z., & Ma, P. F. (2024). Unprecedented variation pattern of plastid genomes and the potential role in adaptive evolution in

Poales. *BMC Biology*, *22*(1), 97. https://doi.org/10.1186/s12915-024-01890-5

Xu, S., Teng, K., Zhang, H., Gao, K., Wu, J., Duan, L., Yue, Y., & Fan, X. (2023a). Chloroplast genomes of four *Carex* species: Long repetitive sequences trigger dramatic changes in chloroplast genome structure. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1100876

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

Yang, J. C., Shin, C. H., Choi, K., Jang, K. S., & Ji, S. J. (2014). *Field guide to the sedges 100 of Korea*. Pochon: Korea National Arboretum.

Yang, J., Ling, C., Zhang, H., Hussain, Q., Lyu, S., Zheng, G., & Liu, Y. (2022). A comparative genomics approach for analysis of complete mitogenomes of five Actinidiaceae plants. *Genes*, *13*(10), Article 1827. https://doi.org/10.3390/genes13101827

Yu, G., Smith, D. K., Zhu, H., Guan, Y., Lam, T. T. Y., & McInerny, G. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zhou, W., Armijos, C. E., Lee, C., Lu, R., Wang, J., Ruhlman, T. A., Jansen, R. K., Jones, A. M., & Jones, C. D. (2023). Plastid genome assembly using long-read data. *Molecular Ecology Resources*, *23*(6), 1442–1457. https://doi.org/10.1111/1755-0998.13787

Zhu, A., Guo, W., Gupta, S., Fan, W., & Mower, J. P. (2016). Evolutionary dynamics of the plastid inverted repeat: The effects of expansion,

contraction, and loss on substitution rates. *New Phytologist*, *209*(4), 1747-1756. https://doi.org/10.1111/nph.13743

ABSTRACT IN KOREAN

국 문 초 록

이 지 은

생물학과

성신여자대학교 대학원

본 연구는 벼목에 속하는 사초과(莎草科, Cyperaceae) 중 가장 큰 속인 사초속(莎草屬, *Carex*)의 소기관 유전체를 대상으로, 진화 양상과 계통학적 관계를 규명하고자 하였다. 이를 위해 10종의 사초(대사초, 대구사초, 들사초, 길뚝사초, 양뿔사초, 도깨비사초, 회색사초, 갯보리사초, 청사초, 백두사초)의 엽록체 유전체와 한국 고유종인 햇사초의 미토콘드리아 유전체 서열을 밝히고 이를 벼목 내의 다른 분류군들과 비교 · 분석하였다. 사초속은 약 2,000종으로 구성되어 있으며, 종 간 형태적 유사성이 높아 전통적인 형태학적 분류 방법만으로는 복잡한 분류학적 문제를 해결하기 어렵다. 제1장에서는 이러한 한계를 극복하기 위해 형태학적 분류와 분자 계통학적 분류를 결합한 통합 접근법의 필요성을 논의하였다. 특히 사초 유전체는 다수의 반복 서열로 인해 빈번한 구조적 재배열이 발생하

며, 이는 유전체 조립과 분석을 복잡하게 만든다고 알려져 있다. 본 연구에서는 차세대 유전체 시퀀싱(NGS) 기술을 활용하여 사초속 식물의 소기관 유전체를 조립하고, 벼목 계통 내 다른 분류군과의 구조적 변이를 비교하였다.

제2장에서는 사초 엽록체 유전체의 낮은 GC 함량과 빈번한 구조적 재배열이 짧고 긴 반복 서열의 존재와 밀접하게 연관되어 있음을 밝혔다. 또한 synteny 분석을 통해 사초속 종 간 구조적 변이 패턴을 확인하였으며, 이러한 변이가 양성화 구조와 같은 특정 형태적 특징과 연관될 가능성을 제시하였다. 제3장에서는 POLAP을 활용하여 한국 고유종 햇사초의 미토콘드리아 유전체를 조립하고, 청사초의 미토콘드리아 유전체와 비교 분석을 통해 사초속 미토콘드리아 유전체의 높은 구조적 가변성을 확인하였다. 이러한 결과는 사초속 미토콘드리아 유전체 연구의 기초 자료를 제공할 뿐만 아니라, 사초과 식물의 진화적 적응과 구조적 다양성을 이해하는 데 기여할 것이다. 본 연구는 사초속 유전체의 구조적 특성과 진화적 양상을 다각적으로 조명하였으며, 형태학적 특징과 분자적 변이를 연계함으로써 분류학적 문제 해결을 위한 새로운 관점을 제시하였다. 이를 통해 사초과 식물의 진화와 다양성에 대한 심층적 이해를 돕는 중요한 기초 자료를 제공할 것으로 기대된다.

AKNOWLEDGEMENTS

감사의 글

이 논문을 완성하기까지 많은 분들의 도움이 있었습니다. 부족하고 지친 순간에도 따뜻한 격려와 도움을 아낌없이 보내주신 모든 분께 깊은 감사를 드립니다.

가장 먼저, 김상태 지도 교수님께 깊은 감사의 말씀을 드립니다. 학부생 시절 연구실에 첫발을 내디뎠던 순간부터 오늘에 이르기까지, 교수님께서는 저에게 도전과 성장을 경험할 수 있는 기회를 아낌없이 주셨습니다. 학부생 신분으로 학회 구두 발표라는 뜻깊은 경험을 할 수 있도록 해주셨으며, 연구의 어려움에 부딪힐 때마다 해결책을 함께 고민해 주시고 늘 용기를 북돋아 주셨습니다. 교수님의 지도 덕분에 연구자로서의 길을 시작할 수 있었습니다.

제 연구에 지속적인 관심과 실질적인 도움을 주신 최상철 교수님께도 깊이 감사드립니다. 생물정보학이라는 낯선 분야를 이해하고 탐구할 수 있도록 이끌어 주셨으며, 논문의 방향과 완성도를 높이는 데 큰 역할을 해 주셨습니다. 교수님께서 주신 세심한 조언 덕분에 이 논문은 끝까지 완성될 수 있었습니다.

따뜻한 배려와 응원으로 함께해 주신 이재원 교수님께도 진심으로 감사드립니다. 실험 수업 조교로 함께했던 시간 동안 학생과 조교들에게 진심으로 대하시는 교수님의 모습을 보며 많은 것을 배웠습니다. 대학원 생활의 어려움을 함께 이해해 주시고, 힘든 순간마다 용기를 북돋아 주셔서 감사합니다. 또한, 교수님께서 이번 논문 심사를 맡아 주신 점에도 진심으로 감사드립니다.

귀한 사초속 식물들을 제공해 주신 조양훈 선생님께도 감사의 말씀을 드립니다. 선생님 덕분에 다양한 사초속 연구를 수행할 수 있었습니다. 또한, 학부 시절 검정납작골풀 연구의 기초 데이터를 제공해 주신 김성민 교수님께도 감사드립니다. 이 연구들이 지금의 논문을 완성하는 데 중요한 발판이 되었습니다.

연구실에서 함께했던 선배들에게도 감사드립니다. 이보라 언니, 랩장으로서 연구실의 중심을 잡아 주시고 궁금한 것들을 친절하게 알려주셨던 모습이 기억에 남습니다. 연구 이야기가 아니더라도 즐겁게 웃고 떠들었던 시간들이 소중했습니다. 박수현 언니, 비록 연구실 생활을 함께 하진 못했지만, 졸업 후에도 시간을 내주셔서 연구 초기의 시행착오를 극복할 수 있도록 조언과 도움을 아끼지 않으셔서 진심으로 감사드립니다. 선배들의 날카롭고 정확한 조언은 저와 후배들의 연구 방향을 설정하는 데 큰 도움이 되었습니다. 현지 언니, 식물에 대한 많은 열정과 지식을 나누며 따뜻한 응원으로 연구실의 분위기를 이끌어 주셔서 감사합니다. 현지 언니의 성실함과 지구력은 매우 큰 강점이며 본받고 싶은 점이었습니다. 같이 있으면 마음이 편해지는 세은 언니와, 그리고 먼저 졸업한 승연이는 사초속

연구의 든든한 동료로서 고민을 나누고 서로를 응원하며 큰 힘이 되었습니다. 연주 언니와 졸업 준비를 함께하며 힘든 순간을 나눴던 시간은 큰 위로가 되었습니다. 밤 늦게까지 연구실에서 함께 고민하고 웃었던 기억은 지금도 소중하게 남아 있습니다. 대학 입학 시절부터 대학원 졸업까지 매 순간을 함께한 나의 일기장, 데이터베이스 같은 정우 언니는 저에게 가장 소중한 동료이자 친구입니다. 항상 제 편이 되어 주고, 힘든 순간에는 진심으로 공감해 주며 기쁜 일에는 함께 웃어준 시간들 덕분에 행복한 대학과 대학원 생활을 만들어 주었습니다. 정우 언니의 빈자리가 정말 클 것 같아 벌써부터 아쉽습니다. 유쾌하고 솔직하며 추진력까지 겸비한 바질 엄마 지예가 연구실에 들어온 첫날부터 보여준 열정과 성실함, 그리고 총명함은 깊은 인상을 남겼습니다. 정우 언니와 셋이 함께 쌓은 추억들도 정말 소중합니다. 저의 졸업에 이어 지예의 대학원 입학을 진심으로 축하하며, 앞으로도 연구실에서 좋은 추억을 쌓아 가길 진심으로 바랍니다.

밝고 열정적인 후배들에게도 감사의 마음을 전합니다. 민경이는 낯선 학문적 배경에도 불구하고 열정적으로 새로운 지식을 흡수하며 연구에 몰두하는 모습이 매우 인상 깊었습니다. 명보는 밝은 성격과 유머 감각으로 연구실 분위기를 따뜻하고 활기차게 만들어주었으며, 항상 주변 사람들에게 긍정적인 에너지를 전해주었습니다. 연구 과정에서 많은 도움을 준 민경이와 명보 덕분에 연구실에서의 시간이 한층 더 뜻깊고 즐거웠습니다. 조용하고 점잖지만 따뜻한 마음씨를 가지고 있는 성실한 지은(서), 그리고 은보라, 김예슬, 최소윤 학부 연구생으로 연구실에 밝은 에너지를 더해 준 후배들에게도 고마운 마음을 전합니다.

친구들과 소중한 인연들도 큰 힘이 되었습니다. 다른 연구실 소속이었지만 같은 시기에 대학원생으로서 고민을 나누며 서로 위로할 수 있었던 석수아 언니, 대학 시절부터 연구실 생활까지 함께하며 여름 락 페스티벌과 같은 즐거운 추억을 만들어준 승진이, 오랜만에 만나도 어색함 없이 늘 편안하고 즐거운 지연이와의 시간들이 많은 위안이 되었습니다. 고등학교 시절부터 지금까지 이어져 온 한나, 유경이, 예진이와의 우정은 지금도 저를 지탱해주는 힘이 되어 주었고 앞으로도 가끔씩 오래 볼 수 있었으면 좋겠습니다. 멀리 떨어져 있지만 예린이가 전해주는 따뜻한 안부는 언제나 큰 위로와 기쁨을 안겨주었습니다. 이어 몇 없는 이성 친구지만 허심탄회하게 이야기를 나눌 수 있는 나병현, 박승준에게 고맙습니다. 마영신 작가님, 바쁘신 와중에도 맛있는 음식을 사 주시며 함께 시간을 나누고 좋은 어른으로서 격려를 보내주셔서 감사합니다. 그리고 권영찬, 심적으로 지치고 힘들 때마다 항상 곁에서 이야기를 들어주고, 긍정적인 에너지로 응원해 주어 진심으로 고맙습니다. 덕분에 어떤 일이든 씩씩하게 다시 해낼 수 있는 용기를 얻었습니다. 마 작가님과 영찬, 셋이 함께 보낸 시간들은 좋은 추억으로 남았습니다. 그 기억들이 큰 힘이 되었고, 앞으로도 저의 삶을 지탱해 줄 것이라고 느껴집니다.

마지막으로, 제 가족들에게 가장 깊은 감사를 드립니다. 할머니와 할아버지, 언제나 따뜻한 사랑과 기도로 저를 응원해 주셔서 진심으로 감사드립니다. 나의 동생 이지현, 서울에서 둘이 함께 지내며 가끔은 다투기도 했지만, 늘 나의 옆에서 내가 아프거나 바쁠 때 집안일도 대신해 주고 맛있는 음식도 많이 해 주어서 고마워. 그리고 부모님, 정서적 · 경제적으로 언제나 저를 뒷받침해 주시고 믿어 주신

두 분 덕분에 오늘의 제가 있습니다. 부모님의 사랑과 헌신을 기억하며 앞으로도 보답하며 살아가겠습니다.

이 논문은 결코 혼자 이루어진 것이 아닙니다. 여러분 덕분에 오늘의 제가 있을 수 있었습니다. 저를 지지해 주시고 함께해 주신 모든 분들께 다시 한번 깊이 감사드리며 글을 마칩니다.

2024년 12월 26일

이 지 은