김 상 태 교수 지도

석사학위 청구논문

MADS-box gene evolution in *Carex* (Cyperaceae) based on genome data

2024

성신여자대학교 대학원

생물학과

이승연

MADS-box gene evolution in *Carex* (Cyperaceae) based on genome data

김상태 교수 지도

이 논문을 석사학위논문으로 제출함

2023년 11월

성신여자대학교 대학원
생물학과
이승연

인 준 서

이승연의 석사학위 논문으로 인준함

2023년 11월

심사위원장 최 상 철 (서명 또는 인)

심 사 위 원 이 재 원 (서명 또는 인) 이재원

심 사 위 원 김 상 태 (서명 또는 인)

성신여자대학교 대학원

ABSTACTS

MADS-box gene evolution in *Carex* (Cyperaceae) based on genome data

Seungyeon Lee
Department of Biology
Graduate School of
Sungshin University

MADS-box genes are transcription factors that control eukaryote development and growth, particularly in regulating flower development in plants through the ABC(D)E model. Revealing the functions of MADS-box genes and their regulatory network, together with other related genes, will play a key role in understanding the evolution of angiosperms. However, studies on floral MADS-box genes in many major lineages of angiosperms have not yet been conducted. *Carex*, with approximately 2,000 species, is the fifth largest genus in angiosperms and has unique floral morphology, including the perigynium, a sac-like structure covering gynoecium. Detection of MADS-box genes and their functional studies in Poales have primarily focused on Poaceae, which includes several model plants, rather than Cyperaceae. In this study, I detected MADS-box genes in *Carex* as a representative of Cyperaceae. I first determined draft genomes from five *Carex* species (*C. siderosticta*, *C. paxii*, *C. dickinsii*, *C. breviculmis*, and *C. capricornis*) based on both long-read (Nanopore) and short-read (Illumina and MGI) sequences. The genomes were of high quality, with 48 to 168 contigs, and had BUSCO values over 94.5%. I analyzed two previously reported *Carex* genomes and representatives of other Poales lineages in conjunction with five newly sequenced genomes from this study. The relationship of the family level of the constructed species tree for twelve genomes was well-matched with known angiosperm phylogeny. In *Carex*, subg. *Carex* formed a sister to subg. Euthyceras. The shared orthogroups of six angiosperm species (*A. trichopoda, A. thaliana, O. sativa, J. effusus, C. siderosticta*, and *C. littledalei*)were 8,300 and 1,436 orthogroups were *Carex*-specific. The MADS-box genes from *Amborella trichopoda*, *Oryza sativa*, and *Arabidopsis thaliana* were used as references in the detection process. Approximately 900 MADS-box genes were detected from these twelve genomes. The maximum-likelihood tree using an amino-acid-aligned DNA matrix was constructed and a detailed analysis of each Type Ⅰ and Type Ⅱ MADS-box genes was conducted. The phylogenetic tree of whole MADDS-box genes showed a clear grouping of Type Ⅰ and Type Ⅱ genes with a few exceptions. In Type Ⅰ analysis, our results show that the four major subgroups of Type Ⅰ

MADS-box genes defined based on the *Arabidopsis* genes should be divided into an additional number of subgroups when considering the evolution of the entire angiosperm MADS-box genes. In the analysis of Type Ⅱ MADS-box gene, seventeen subfamilies were recognized. In *AGL12, StMADS11, AGL17, TM3, SEP*, and *SQUA* subfamilies, highly supported shared gene duplications across *Carex* species were detected. Especially, *Carex*-specific duplication events were found in *AGL17* and *StMADS11* subfamilies. I also compared various methods for detecting MADS-box genes and discussed the boundary of the MADS-box gene family. Our findings will shed light on the evolution of the floral structure and genome of *Carex*, which serves as a representative for the Cyperaceae.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Carex L., included in Cyperaceae Juss., contains over 2,000 species distributed worldwide (Hip, 2007), and it is the fifth largest genus in angiosperms (Fordin, 2004). Cyperaceae (ca. 5500 species) is placed in one of two major lineages in Poales Small with Poaceae Barnhart (Cole et al., 2019; Fig. 1). In Poales, Juncaeae L. is a sister to Cyperaceae, and it consists of eight genera and about 464 known species (Christenhusz et al., 2016; Fig. 1). Morphologically, Juncaceae differs from Cyperaceae in that Juncaceae flowers singly or with multiple flowers forming a cyme or glomerule, and the fruit is a capsule containing multiple seeds (Yang et al., 2014). Cyperaceae, Juncaceae, and Poaceae are collectively called 'graminoid' plants, characterized by their grass-like appearance (Park and Allaby, 2017). To gain a comprehensive understanding of the plants within the Poales, a thorough study of graminoid species is essential. However, most monocot researchers have focused on the Poaceae group in Poales but have not paid much attention to the Cyperaceae because Poaceae contains many food-resource plants. Although the species of Cyperaceae don't have a noticeable economic value, they are important species in grassland ecosystems.

Carex is the largest genus in Cyperaceae (Smith & Faulkner, 1976). Recent morphological and phylogenetic studies have shown that most genera in Cariceae have merged into *Carex*, and sister to

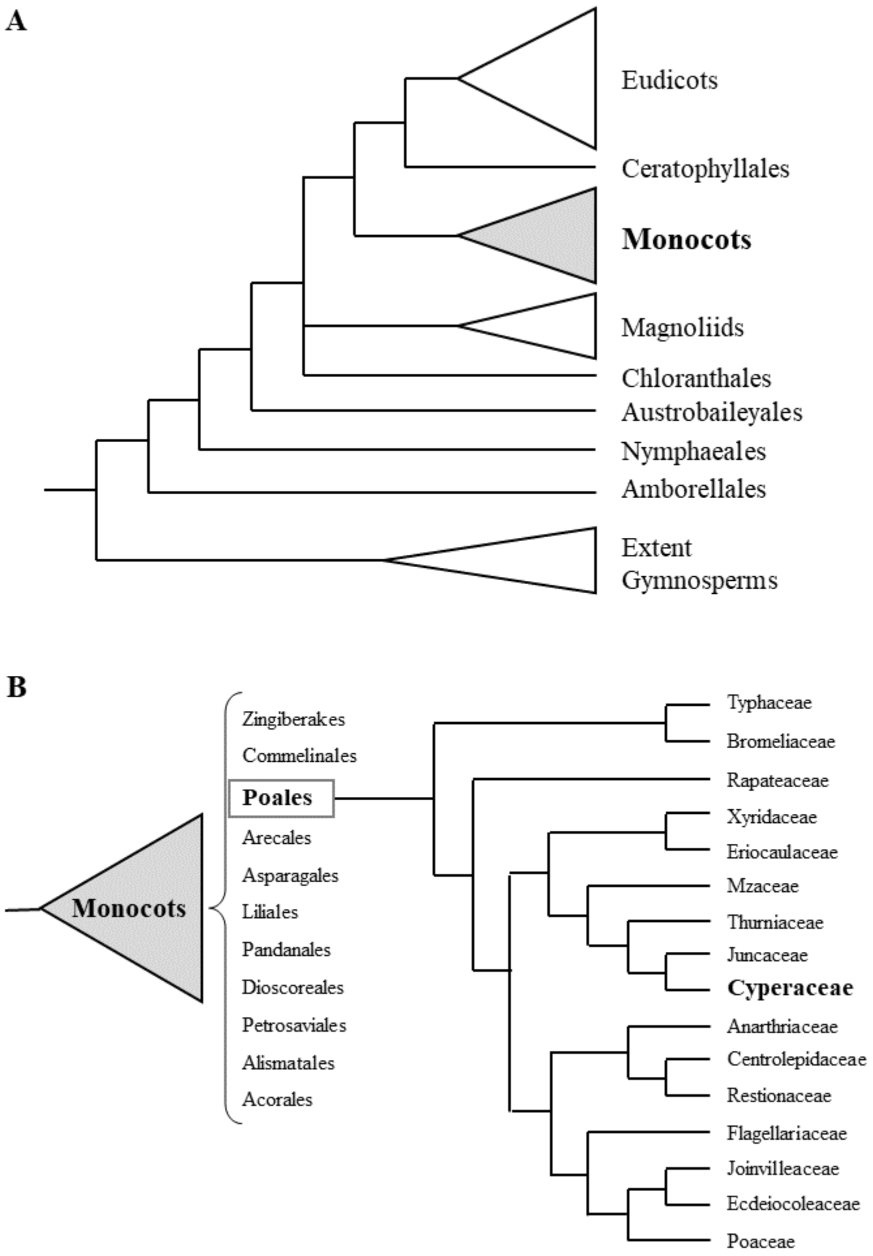


Fig. 1. Position of monocots in the angiosperm phylogenetic tree (A), and position of Cyperaceae in Poales phylogenetic tree (B). Topologies are summarized by Sokoloff et al. (2018) and the angiosperm phylogeny website (ver. 14).

Carex is the genus *Sumatroscripus* Oteng−Yeb. (Larridon et al., 2021).

Almost all species of *Carex* are perennial, and they typically have rhizomes, stolons or short rootstocks (Mohlenbrock and Nelson, 1999). The culm, the flower−bearing stalk, is unbranched and usually erect and is usually distinctly triangular in section (Mohlenbrock and Nelson, 1999). The leaves of *Carex* consist of a blade that extends outward from the stalk and a sheath that encloses a portion of the stalk (Mohlenbrock and Nelson, 1999). At the point where the blade meets the culm, there is a structure known as the ligule (Mohlenbrock and Nelson, 1999).

The flowers of *Carex* are combined into spikes, which are further grouped into a larger inflorescence (Mohlenbrock and Nelson, 1999). Almost all *Carex* species are monoecious, so each flower is either male (Staminate) or female (pistillate) (Mohlenbrock and Nelson, 1999). The floral structure of *Carex* is somewhat different from typical angiosperm flowers. Instead of a typical four−whorled structure, *Carex* has a unique floral structure. This perigynium is a sac−like structure covering gynoecium (Fig. 2). Each perigynium has one scale. The morphology of the scale is similar in female and male flowers (Fig. 2). The evolutionary origin of these floral organs of *Carex* is unclear.

The evolution and diversification process of *Carex* has attracted significant interest from many botanists. After Linne first described this genus (Linnaeus, 1753), many regional studies have been performed based on morphology, e.g., Kukenthal (1909), Mackenzie (1931−1935), Kreczetowicz (1935), Egorova (1999), and Liang and Koyama (2010.) In particular, Kukenthal (1909) suggested a relatively comprehensive global classification system of Caricoideae, including *Carex*, focused on the morphology of flowers and inflorescences. However, recent molecular phylogenetic studies using DNA (Starr et al., 1999; Waterway et al., 2009) showed very different results from the traditional classification system. In particular, it has been suggested that *Cymophyllus* (monotypic), *Kobresia* (ca. 60 species), *Schoenoxiphium* (ca. 15 species), and *Uncinia* (ca. 70 species), which have been considered to be closely related genera of *Carex*, were included in the clade of *Carex*.

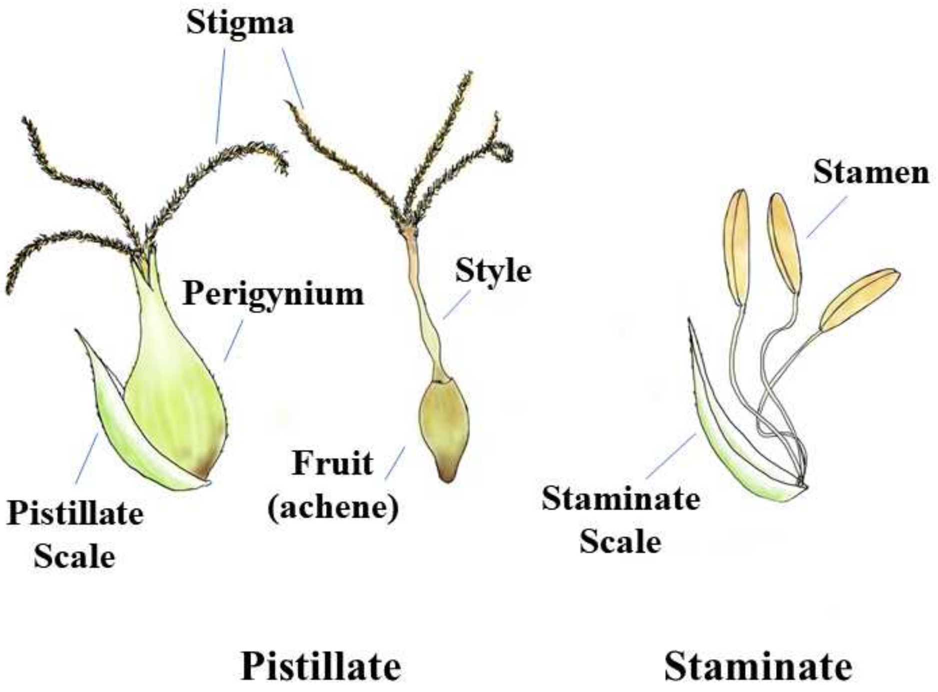


Fig. 2. The floral structure of *Carex*.

This considerable result has been confirmed again in the study of Global *Carex* Group (2016) based on the ITS and ETS of the nuclear genome and the *matK* sequence of the chloroplast genome using a comprehensive sampling of the genus (996 species; 50.23%). Recently, HybSeq (hybridization sequencing or captured enrichment sequencing), which is a combined method of target-specific capturing and large-capacity sequencing based on next-generation sequencing (NGS), has been applied in the study of *Carex* (Villaverde et al., 2020). The phylogenetic tree clearly showed six major lineages in the *Carex* (*Siderostictae, Schoenoxiphium, Unispicate, Uncinia, Vignea*, and Core *Carex*) with high supporting values (Villaverde et al., 2020). Recognition of genus boundaries and major lineages made it possible to suggest a new classification system for *Carex* (Global *Carex* Group, 2021). The new system organizes *Carex* into six subgenera: subg. *Siderosticta,* subg. *Psyllophorae,* subg. *Euthyceras,* subg. *Uncinia,* subg. *Vignea,* and subg. *Carex* (Fig. 3).

To understand the evolutionary origin of floral organs in *Carex*, we need to understand the study of the MADS-box gene family, which contains most floral organ identity genes in model plants. Genes of the MADS-box gene family have a MADS domain that contacts DNA directly and acts as a transcription factor. Members of this gene family have diverged during the long evolutionary history based on gene duplications (Nam et al., 2004). There are two different types of MADS-box genes in plants: Type I and

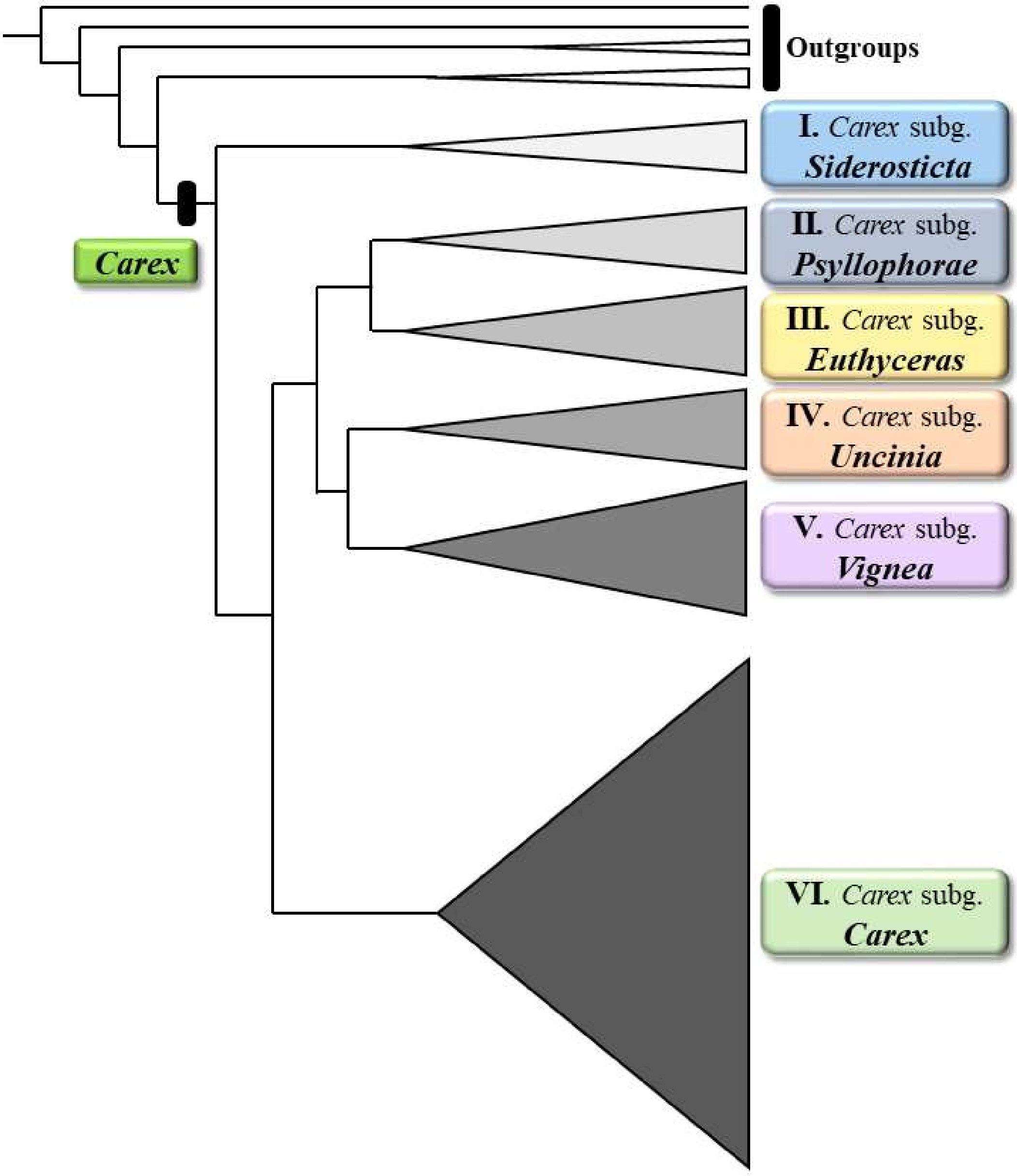


Fig. 3. New classification system of *Carex* based on the phylogenetic tree (summarized from Global *Carex* Group, 2021).

Type Ⅱ (Fig. 4). Type Ⅰ gene has only the MADS domain, and Type Ⅱ gene in plants contains M, I, K, and C regions (MIKC-type; Nam et al., 2004; Fig. 4). Some MIKC-type genes contain long I (intervening sequence) region (MIKC*-type) and others are not (MIKC^{c}-type) (Fig. 4). Some members of the MIKC^{c}-type MADS-box genes are highlighted because they are involved in the identification of floral organs in angiosperms. Parenicova et al. (2003) showed there are six groups of MADS-box genes in *Arabidopsis*: MIKC type, Mα, Mβ, Mγ, Mδ groups, and AGL33 (Fig. 5). However, the supporting values for these clades were very low.

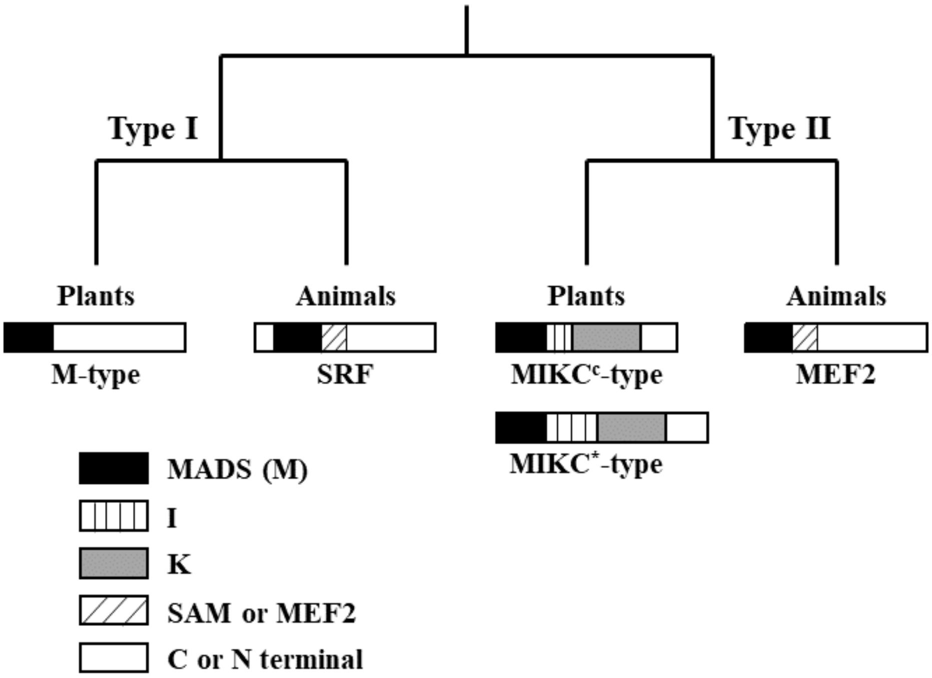


Fig. 4. Domain structure of types Ⅰ and Ⅱ MADS-box genes in plants and animals. Redrawn from Nam et al. (2004).

In *Arabidopsis*, the basic ABC model, extended ABC(D)E model, and quartet model explain floral organ identities in a combinatorial manner (Fig. 6). The classical ABC model posits that organ identity within each floral whorl is regulated by distinct combinations of three organ identity genes, called A, B, and C (Theissen and Melzer, 2007; Fig. 6). Expression of class A genes alone specifies the formation of sepals; the combination of A and B expressions specifies the development of petals; the combination of B and C expressions specifies stamen formation; expression of C alone determines the development of carpels (Theissen and Melzer, 2007; Fig. 6).

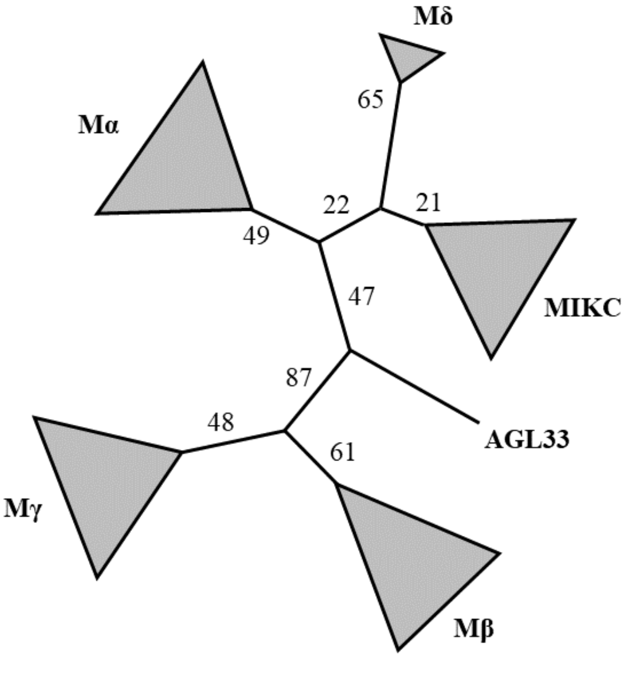


Fig. 5. Phylogeny of MADS-box genes in *Arabidopsis*. A result from Parenicová et al. (2003).

Later, D and E class genes were added to explain ovule identity (D−function; Colombo et al., 1995) and those that contribute to sepal, petal, stamen, and carpel identity (E−function; Pelaz et al., 2000). In many studies, D−function is not considered independently (indicated as "(D)" in this study; Soltis et al., 2007; Soltis et al., 2008) because ovules are not separate floral organs.

In *Arabidopsis*, class A genes are represented by *APETALA1* (*AP1*) and *APETALA2* (*AP2*), class B genes by *APETALA3* (*AP3*) and *PISTILLATA* (*P.I.*), and class C genes by *AGAMOUS* (*A.G.*) (Theissen and Melzer, 2007). Molecular cloning of these genes revealed that, except *AP2*, they all represent MIKC−type MADS−box genes (Theissen and Melzer, 2007). The floral quartet model of floral organ specification proposes that different tetramers of MIKC−type MADS−domain transcription factors regulate gene expression, thereby determining the identity of floral organs during development (Theißen et al., 2016; Fig. 6). These models are also globally well−conserved in Poaceae (Ciaffi et al., 2011). Poaceae has a peculiar floral structure, the floret, which contains carpels and stamens and reproductive organs are surrounded by two lodicules, which correspond to eudicot petals, and by a palea and lemma, whose correspondence to eudicot organs remains controversial (Ciaffi et al., 2011). In Poaceae, a significant number of genes with unique and diversified functions were identified, and genes for floral development corresponding to class A, B, C, D, and E genes in eudicots were also identified (Ciaffi et al., 2011).

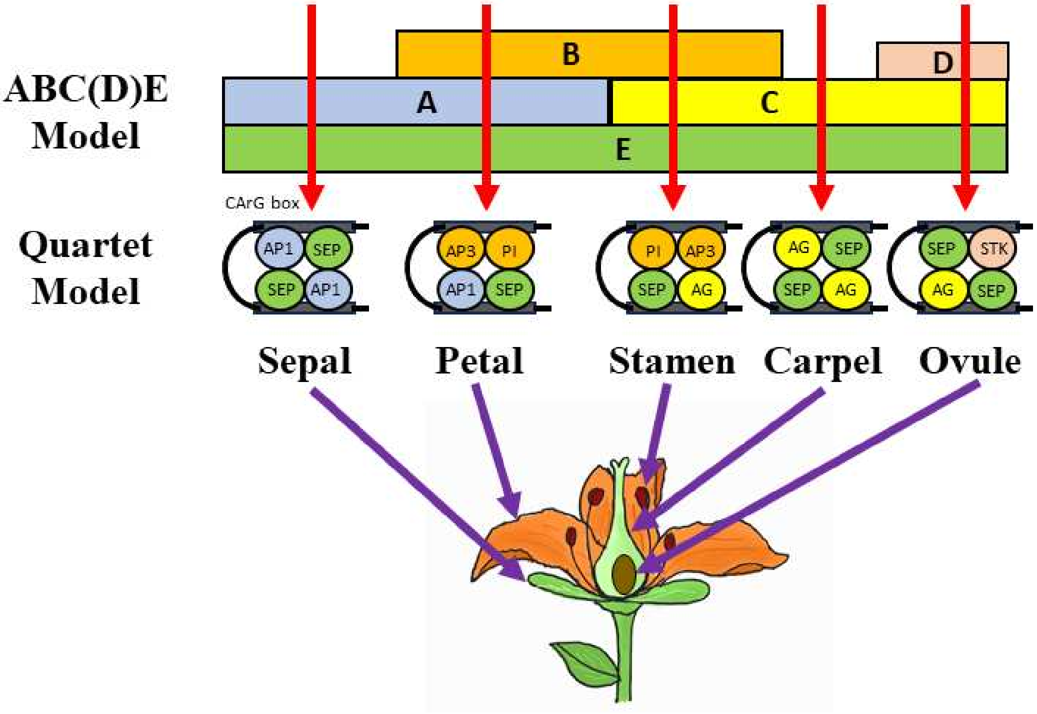


Fig. 6. Schematic drawing showing ABC(D)E model and quartet model. The ABC(D)E protein complex specifies floral organ identity by binding CArG boxes in the promoter region of the target gene (Coen and Meyerowitz, 1991; Pelaz et al., 2000; Pinyopich et al., 2003; Theissen, 2001).

MADS-box genes have not been studied in *Carex*. In this study, I aim to determine the MADS-box genes of *Carex*, identify their orthology with known MADS-box genes through phylogenetic analyses, and understand their evolution. Eventually, by examining the expression of these genes in different floral parts of *Carex* and comparing them to the classical ABC(D)E model, I expect to infer the evolutionary origins of specialized organs in *Carex*, such as the perigynium and female/male scale.

Complete genome study in the MADS-box studies has a significant meaning. For example, in the first MADS-box gene study in *Amborella*, which is a critical taxon in angiosperm evolution, only B- and C-class genes were identified (Kim et al., 2005). The sequence determination of these genes has been done by the PCR-based cloning method with MADS-box-specific degenerate primers. The expression pattern of these genes suggested the ancestral expression status of all angiosperms. However, after the determination of *Amborella* genome, the ancestral status of the expression has been dramatically changed because of the detection of two additional genes (*AP1* and *PI*-2; *Amborella* genome project, 2013). Additional MADS-box genes from a completed genome changed our understanding of the ancestral expression pattern of angiosperms when I reconstructed the ancestral status of expressions (unpublished data). Therefore, identifying the complete set of a particular gene group in a genome is critical for interpreting their expression and function.

Although the speed of plant genome publication has dramatically increased, published *Carex* genomes are few now. Even though there are more than 2,000 known species of *Carex* worldwide, only six *Carex* genomes have been published to date (plaBiPD, https://www.plabipd.de/plant_genomes_pa.ep). In this study, for five representative species of *Carex*, I generated 1) both long-read and short-read NGS sequences, 2) assembled draft genomes, 3) compared their genomes with other angiosperm plants, and 4) addressed the evolution and lineage-specific gene duplications of MADS-box genes. The five *Carex* genomes determined in this study will contribute to the research of underrepresented sedges. Furthermore, these studies will provide insights into the evolutionary patterns of floral structure and genome evolution in *Carex*, which represents Cyperaceae.

2. Materials and Methods

2.1. Plant materials

Five representative *Carex* species for the genome study are *Carex siderosticta* Hance, *Carex paxii* Kük., *Carex breviculmis* R.Br., *Carex capricornis* Meinsh. ex Maxim, *and Carex dickinsii* Franch. & Sav. (Fig. 7, 8, 9, 10, 11, and Table 1). These plants were collected from the wild in Korea, transplanted, and cultivated at the cultivation field in the Sungshin Women's University (Fig. 7, 8, 9, 10, 11, and Table 1). For comprehensive analyses, genome of two published *Carex* species, *Carex littledalei* (C.B.Clarke) S.R.Zhang and *Carex scoparia* Schkuhr ex Willd., are included in this study (Table 1). Two *Juncus* species, *Juncus effusus* L. and *Juncus inflexus* L., and three references representing other angiosperms, *Oryza sativa* L. (other monocots), *Arabidopsis thaliana* (L.) Heynh. (eudicots), and *Amborella trichopoda* Baill. (basal angiosperms), were also included (Table 1).

Three reference species, *A. trichopoda*, *A. thaliana*, and *O. sativa*, were used to detect MADS-box genes from *Carex* and *Juncus* species. MADS-box gene sequences of *A. trichopoda* were based on *Amborella* Genome Project (2013), those of *A. thaliana* were from 'The *Arabidopsis* Information Resource (TAIR)', and those of *O. sativa* were from 'Rice Genome Annotation Project (RGAP 7)'.

2.2. DNA extraction

High-molecular-weight (HMW) DNAs were extracted from fresh leaves following the recently published protocol by Kang et al. (2023), which combines (1) a nuclei extraction method followed by (2) a traditional CTAB DNA extraction method for plants with optimized extraction conditions. Extracted DNAs were used for both long-read and short-read sequencing.

The concentration, purity, and length of all DNA samples were measured by NanoDrop (Thermo Fisher Scientific, Massachusetts), Qubit (Invitrogen, Massachusetts), and FEMTO Pulse (Agilent Technologies, Santa Clara).

Table 1. Plant materials for comparative genomic analyses in this study. Bold characters indicate genomes determined in this study

Carex subgenus*	Species	Voucher / Reference
Ⅰ. *Carex* subg. *Siderosticta*	***Carex sidserosticta***	*Y. Cho s. n.*, SWU0036866
Ⅲ. *Carex* subg. *Euthyceras*	*Carex littledalei*	Can et al. (2020)
Ⅴ. *Carex* subg. *Vignea*	***Carex paxii***	*Y. Cho s. n.*, SWU0036903
Ⅴ. *Carex* subg. *Vignea*	*Carex scoparia*	Planta et al. (2022)
Ⅵ. *Carex* subg. *Carex*	***Carex breviculmis***	*S. Kim 2019-005*, SWU0029244
Ⅵ. *Carex* subg. *Carex*	***Carex capricornis***	*Y. Cho 2022-055*, SWU0054311
Ⅵ. *Carex* subg. *Carex*	***Carex dickinsii***	*Y. Cho 2022-03*, SWU0054291
Sisters to *Carex*	*Juncus effusus*	Planta et al. (2022)
	Juncus inflexus	Planta et al. (2022)
A sister to Cyperaceae	*Oryza sativa*	GCF_001433935.1; Kawakara et al. (2013)
A sister to Monocots	*Arabidopsis thaliana*	GCF_000001735.4; Tanya et al. (2015)
A sister to angiosperms	*Amborella trichopoda*	GCF_000471905.2; *Amborella* Genome Project (2013)

*: Subgeneric classification follows Global *Carex* Group (2021).

성신여자대학교 식물표본관
HERBARIUM, SUNGSHIN WOMEN'S UNIVERSITY

대사초 Cyperaceae
Carex siderosticta Hance

Near the summit of Mt Hambaek in Gohan-eup, Jeongseon-gun, Gangwon-

강원 정선군 고한읍 함백산 정상부근

2017.06.05

Coll.: Yanghoon Cho
Det.: Yanghoon Cho

Note: CA047

Fig. 7. The voucher specimen of *C. siderosticta* used in this study (*Y. Cho s. n.*, SWU0036866)

성신여자대학교 식물표본관
HERBARIUM, SUNGSHIN WOMEN'S UNIVERSITY

대구사초 Cyperaceae
Carex paxii Kük.

Gangju-ri, Beopsu-myeon, Haman-gun, Gyeongsangnam-do

경남 함안군 법수면 강주리

2019.05.16 35.310974, 128.321493

Coll.: Yanghoon Cho
Det.: Yanghoon Cho

Note: CA073

WR-190517-9

SWU0036903

Fig. 8. The voucher specimen of *C. paxii* used in this study (*Y. Cho s. n.*, SWU0036903)

성신여자대학교 식물표본관
HERBARIUM, SUNGSHIN WOMEN'S UNIVERSITY

청사초
Cyperaceae
Carex breviculmis R.Br.

Start point of the path from Opesan tunel to the summit,
Gangbuk-gu, Seoul
서울시 강북구 오패산터널 ->정상 입구

2019.05.01 N37°38'01.66 E127°02'42.47
Coll.: Sangtae Kim

Det.: N37°38'01.66 E127°02'42.47
Note: S.Kim 2019-005
엽초 2mm이하, 수과표면 매끈, 자인편 녹백색,
소수: 줄기윗부분 모여달림 과포 박막실, 수두 3개,

SWU0029244

Fig. 9. The voucher specimen of *C. breviculmis* used in this study (*S. Kim 2019-005*; SWU0029244)


성신여자대학교 식물표본관
HERBARIUM, SUNGSHIN WOMEN'S UNIVERSITY

양뿔사초 Cyperaceae
Carex capricornis Meinsh. ex Maxim.

692 Dangbuk-ri, Oksan-myeon, Gunsan-si, Jeollabuk-do

전북 군산시 옥산면 당북리 692

2022.06.02 N35.936016 E126.717543

Coll.: Yanghoon Cho, Suksoon Kim
Det.: Yanghoon Cho

Note: CA148
Y. Cho 2022-055
저수지나 습지 주변, 수온이 낮은 곳을 선호

SWU0054311

Fig. 10 The voucher specimen of *C. capricornis* used in this study (*Y. Cho 2022-055,* SWU0054311)

성신여자대학교 식물표본관
HERBARIUM, SUNGSHIN WOMEN'S UNIVERSITY

도깨비사초 Cyperaceae
Carex dickinsii Franch. & Sav.

Mountain 2-3 in Jungsan-dong, Buk-gu, Ulsan

울산 북구 중산동 산 2-3

2022.05.23 N35.669307 E129.371376

Coll.: Yanghoon Cho, Suksoon Kim
Det.: Yanghoon Cho

Note: CA133
Y. Cho 2022-037
산간 샘수가 나오는 햇볕이 좋은 곳

SWU0054291

Fig. 11 The voucher specimen of *C. dickinsii* used in this study (*Y. Cho s. n.,* SWU0054291)

2.3 Sequencing

Long-read sequencing data was generated usingthe ONT MinION platform (Oxford Nanopore Technologies, Oxford) with R9 version flowcells. The ONT libraries were constructed with the LSK109 kit (SQK-LSK109, Oxford Nanopore Technologies, Oxford). Short-read sequencing data was generated using HiSeq2000 (Illumina, San Diego) or MGISEQ-2000 (MGI-Tech, Wehrheim) platforms.

2.4 Assembly and annotation

Raw ONT long-reads were adapter-trimmed using Porechop (v0.2.4; Wick et al., 2017), and de novo assembly of the adapter-trimmed ONT long-reads was performed using NextDenovo assembler (v2.5.0; Hu et al., 2023). Long read polishing was performed using NextPolish-lgs (v1.4.1; Hu et al., 2019) and Medaka (v1.2.3; Oxford Nanopore Technologies Ltd., 2018), and short read polishing was performed using NextPolish-sgs (v1.4.1, Hu et al., 2019) and Polypolish (v0.5.0; Wick and Holt, 2022). The assembled genome was evaluated using BUSCO (v5.3.0; Manny et al., 2021) with ‘ebmbryophyta_odb10’ data.

The transcriptomes from six tissues (leaf, female carpel, female perigynium, female scale, male scale, and male stamen; unpublished) of *C. dickinsii* and three tissues (leaf, stem, and root; unpublished) of *C. paxii* were used for genome annotations except for *C. siderosticta.* In the case of *C. siderosticta,* a transcriptome from a leaf was added (unpublished).

Repeat library was generated by inputting the genome fasta file into RepeatModeler (ver. 2.0.1; Flynn et al., 2020) with a parameter 'engine = ncbi'. A combined repeat library was generated by integrating the generated repeat library and building a repeat library (Dfam 3.2, Repbase release 20181026). RepeatMasking result was generated using RepeatMasker (ver. 4.1.1; Smit et al., 2013–2015) with option '-e ncbi -nolow' and converted to gff file format with rmOutToGFF3.pl script. The final RepeatMasker result is in the gff3 format with the complex repeats separated.

Gene prediction was conducted using both BRAKER (Bruna et al.,2021; Hoff et al., 2016; Hoff et al., 2019; Stanke et al., 2006; Stanke et al., 2008) and MAKER (ver. 3.01.04; Cantarel et al., 2008). For protein evidence, I used Viridiplantae data in the case of BRAKER, and *C. littledalei* (the closest previously published genome to current working species) in the case of MAKER. To test the effectiveness of different prediction tools and protein evidence data, I also perform predictions on *C. dickinsii* using Viridiplantae data in MAKER. After gene prediction with MAKER, functional annotations were performed using BLAST 2.6.0+ (Camacho et al., 2008) with E-value of 1e-5 and GO Mapping/GO annotation protocol (Gotz et al., 2008) from OmicsBox (ver. 2.2.4; Bioinformatics and Valencia, 2019) with the default option. GO annotations were also performed with transcriptomes of the current study using the InterProScan protocol (ver. 5.59-91.0; Blum et al., 2020).

tRNA prediction was performed using tRNAscan-SE (ver. 1.3.1;

Chan et al., 2021) with option '-o, -f, -H', and ribosomal RNA prediction was conducted using Rnammer (ver. 1.2; Lagesen et al., 2007) with option '-S euk -m tsu, ssu, lsu'.

The assembly completeness of the assembled genome was evaluated using BUSCO (v5.3.0; Manni et al., 2021) with 'embryophyta_odb10' data, and K-mer analyses were performed using GenomeScope (Vurture et al., 2017).

2.5 Genome comparison

All analyses were performed with CDS sequences extracted from genome assembly and annotated gff files generated by the MAKER gene prediction tool with *C. littledalei* using the gffread (v0.12.7; Pertea and Pertea, 2020). A species tree of twelve angiosperms was constructed using an orthofinder (v2.5.4; Emms & Kelly, 2015, 2017, 2018, 2019). Genes Venn diagrams were generated based on OrthoVenn3 (Sun et al., 2023).

The genome, gene, and intron sizes were analysed using the Genious Prime (ver. 2023.2.1.; Kearse et al., 2012) with an annotation result from the BRAKER.

2.6 Identification of MADS-box genes and phylogenetic analyses

MADS-box genes from three well-studied species, *Amborella trichopoda* (41 sequences; *Amborella* genome project, 2013), *Arabidopsis thaliana* (107 sequences; Parenicova et al., 2003), and

Oryza sativa (76 sequences; Rice Full-Length cDNA Consortium et al., 2003), were used as references for detecting MADS-box genes from *Carex* and *Juncus* species included in this study. BlastP 2.5.0+ (Camacho et al., 2008) with an E-value of 1e-5 was conducted to detect MADS-box genes, and an InterProScan search (Jones et al., 2014; Blum et al., 2020) was performed to identify definitive MADS-box genes based on the presence of MADS domain.

MADS-box genes of 12 species, including *Carex* and *Juncus* species detected by BlastP and InterProScan search, were aligned with MUSCLE from MEGA7. A phylogenetic tree was generated using IQ-TREE (v. 2.0.3; Minh et al., 2020) with the best model searched by the model test option. Subsequent phylogenetic analyses were performed to give better resolution with 1) the realigned Type II genes matrix and 2) each realigned subset of subfamilies of the MIKC^c gene (see results). For all phylogenetic analyses, each node was evaluated with 1,000 bootstrapping replications. To obtain a better alignment, I rearranged the matrix of taxa according to their order in a phylogenetic tree because the sequence alignment result is input order-sensitive, and I used it to perform the phylogenetic analysis again. I applied three rounds of repeats of these processes for all of the phylogenetic analyses.

3. Results

3.1 Genome sequencing, assembly, and annotation.

The average genome size of five *Carex* species was 0.46 Gbp (0.28 Gbp ~ 0.75 Gbp) as a result of measurement by flow cytometry (Table 2). An average of 97x long-read data sequenced with ONT and an average of 83x short-read data sequenced with Illumina or MGI were generated (Table 2).

The sizes of assembled genomes ranged from 0.25 Gbp to 0.8 Gbp, and they were matched well compared to the genome size measured by flowcytometry (Table 3). The total number of contigs of five assembled genomes ranged from 48 to 168 (Table 3). All five assembled genomes showed high complete BUSCOs values of approximately 95 % (Table 3). From 19,923 to 36,105 genes were predicted as a result of gene prediction (Table 4).

Table 2. Statistics of sequencing data

Results	*Carex siderosticta*	*Carex paxii*	*Carex breviculmis*	*Carex capricornis*	*Carex dickinsii*
Genome size (flowcytometry)	0.75 Gbp	0.28 Gbp	0.49 Gbp	0.37 Gbp	0.45 Gbp
Long-read					
Total sequence	70.83 Gbp	51.6 Gbp	46.6 Gbp	23.0 Gbp	23.2 Gbp
Million reads	12	2.1	9.9	4.6	3.5
Coverage	94x	184x	95x	62x	52x
N50 read length	18 Kbp	46.4 Kbp	13 Kbp	13 Kbp	14 Kbp
Short-read					
Total sequence	36 Gbp	20 Gbp	36 Gbp	39 Gbp	36 Gbp
Million reads	237	134	242	129	241
Coverage	48x	71x	73x	105x	80x

Table 3. Assembly results of five *Carex* species

Results	*Carex siderosticta*	*Carex paxii*	*Carex breviculmis*	*Carex capricornis*	*Carex dickinsii*
Genome data					
Genome size (flowcytometry)	0.75 Gbp	0.28 Gbp	0.49 Gbp	0.37 Gbp	0.45 Gbp
Chromosome number	2n=24	2n=76	2n=68	(Measuring)	(Measuring)
Assembled sequence					
Total length (bp)	796,166,905	254,810,564	456,716,150	323,290,232	401,660,442
Total number of contigs	141	48	168	55	112
Longest contig (bp)	55,216,364	16,242,166	19,397,854	12,929,454	16,288,113
Contigs N50 length (bp)	14,891,410	8,666,829	7,798,935	8,998,381	9,594,122
Contigs average length (bp)	5,646,573.79	5,308,553	2,718,548.50	5,878,004.20	3,586,253.90
GC content (%)	36.2	33.46	34.57	33.47	33.82

Table 3.- (continued)

Results	*Carex siderosticta*	*Carex paxii*	*Carex breviculmis*	*Carex capricornis*	*Carex dickinsii*
BUSCO Result					
Complete BUSCOs	94.50 %	94.67 %	95.00 %	94.80 %	95.10 %
Complete and single-copy BUSCOs	83.40 %	92.38 %	90.80 %	90.80 %	91.20 %
Complete and duplicated BUSCOs	11.10 %	2.29 %	4.20 %	4.00 %	3.90 %
Fragmented BUSCOs	1.80 %	1.67 %	1.50 %	2.00 %	1.60 %
Missing BUSCOs	3.70 %	3.66 %	3.50 %	3.20 %	3.40 %
K-mer analysis					
Heterozygosity	2.38 %	0.0926 %	0.669 %	0.311 %	0.298 %
Haploid length (bp)	533,859,065	262,110,773	421,300,803	308,249,794	361,354,948
Error Rate	0.51 %	0.294 %	0.124 %	0.26 %	0.0606 %
K-mer cov.	19.5	30.6	16.3	29	20.2

Table 4. Annotation results of five *Carex* species based on the MAKER gene prediction tool with *C. littledalei* as protein evidence

Results	*Carex siderosticta*	*Carex paxii*	*Carex breviculmis*	*Carex capricornis*	*Carex dickinsii*
RepeatMasker result					
Total interspersed repeats (% of the genome)	486,765,964 bp (61.14 %)	94,507,129 bp (37.09 %)	248,086,137 bp (54.32 %)	142,054,508 bp (43.94 %)	211,534,698 bp (52.67%)
Gene prediction Result					
Number of predicted genes	36,015	19,923	21,852	20,685	26,120
Function annotation results (Number of annotated genes / %)					
NR	30,524 / 84.75 %	N/A	20,744 / 94.93 %	19,701 / 95.24 %	22,298 / 85.37 %
InterProScan	19,416 / 53.91 %	N/A	21,098 / 96.55 %	20,020 / 96.79 %	22,595 / 86.50 %
GO	23855 / 66.24 %	N/A	17,590 / 80.50 %	16,740 / 80.93 %	18,754 / 71.80%
tRNA & rRNA results					
Total tRNAs	817	530	194	647	653
rRNA sequences	667	540	620	316	227

N/A, not applicable

3.2 Comparison of gene and genome characteristics and phylogenetic analysis

The orthology relationships of six species of angiosperm, *Amborella trichopoda, Arabidopsis thaliana, Oryza sativa, Juncus effusus, Carex siderosticta*, and *Carex littledalei,* were investigated using OrthoVenn3 (Fig. 12). The shared orthogroups of all six species were 8,300 and 1,436 orthogroups were *Carex*-specific (Fig. 12).

The phylogenetic tree was constructed by the Orthofinder tool (Fig. 13). The relationship of the family level was well-matched with known angiosperm phylogeny (Fig. 13). In the genus *Carex*, a species of subg. *Siderosticta*, is placed at the base of the *Carex* tree (Fig. 13). Subg. *Carex*, which contains the largest number of species in *Carex*, forms a sister to subg. *Euthyceras* (Fig. 13).

The genome, gene, and intron sizes are shown by species (Fig. 14). Geneious prime was used to fit each gff file to the genome and examine the respective numbers. *C. siderosticta* had the largest genome size, gene size, and intron size within *Carex* (Fig. 14). The gene size and intron size of *C. littledalei* were unusually small compared to other *Carex* species, suggesting that there may have been an annotation error in the laboratory that published the genome (Fig. 14).

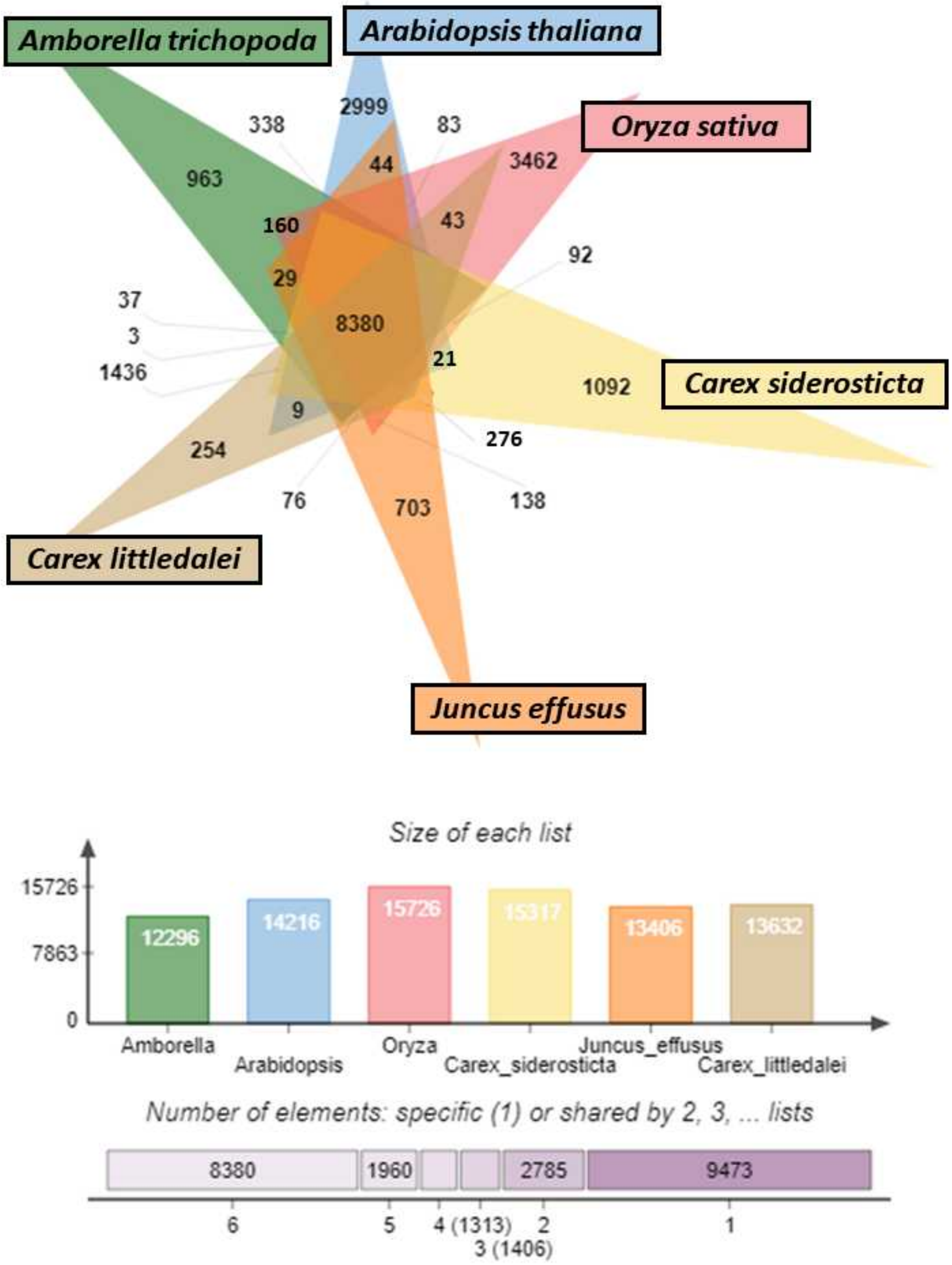


Fig. 12. Orthogroups relationship of six species of angiosperms. Data of *C. siderosticta* is based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.

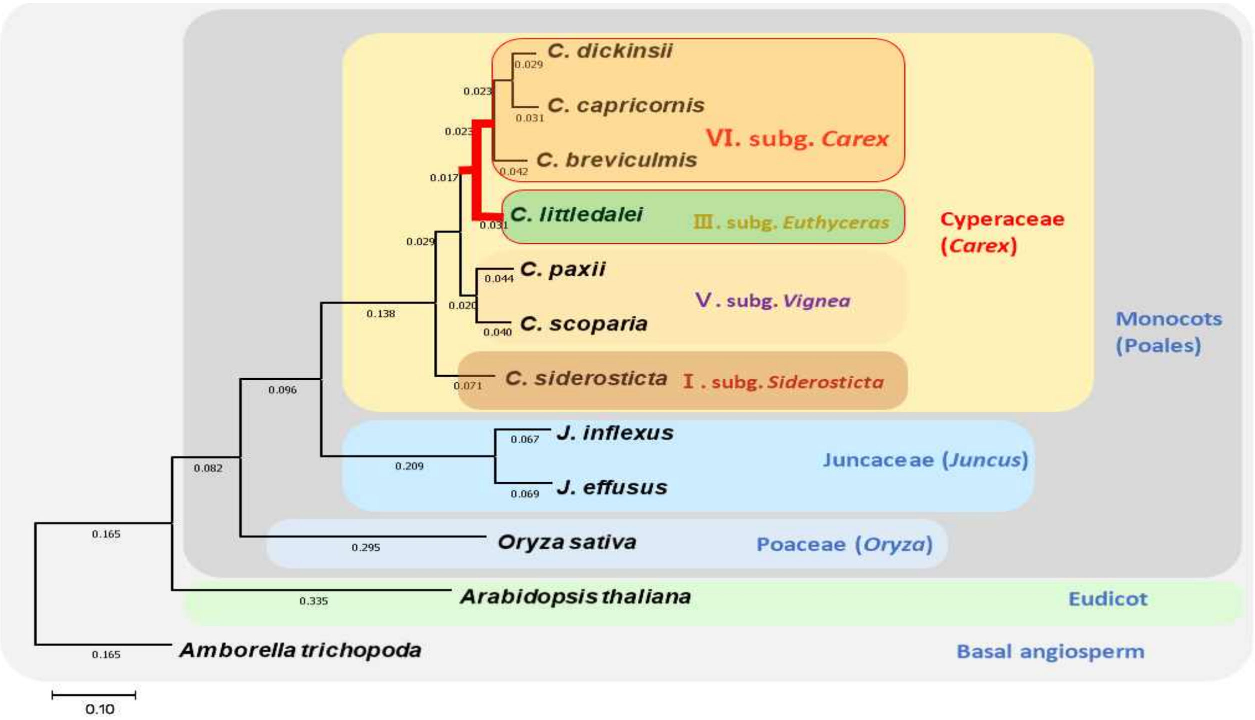


Fig. 13. Species tree based on twelve angiosperm species. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.

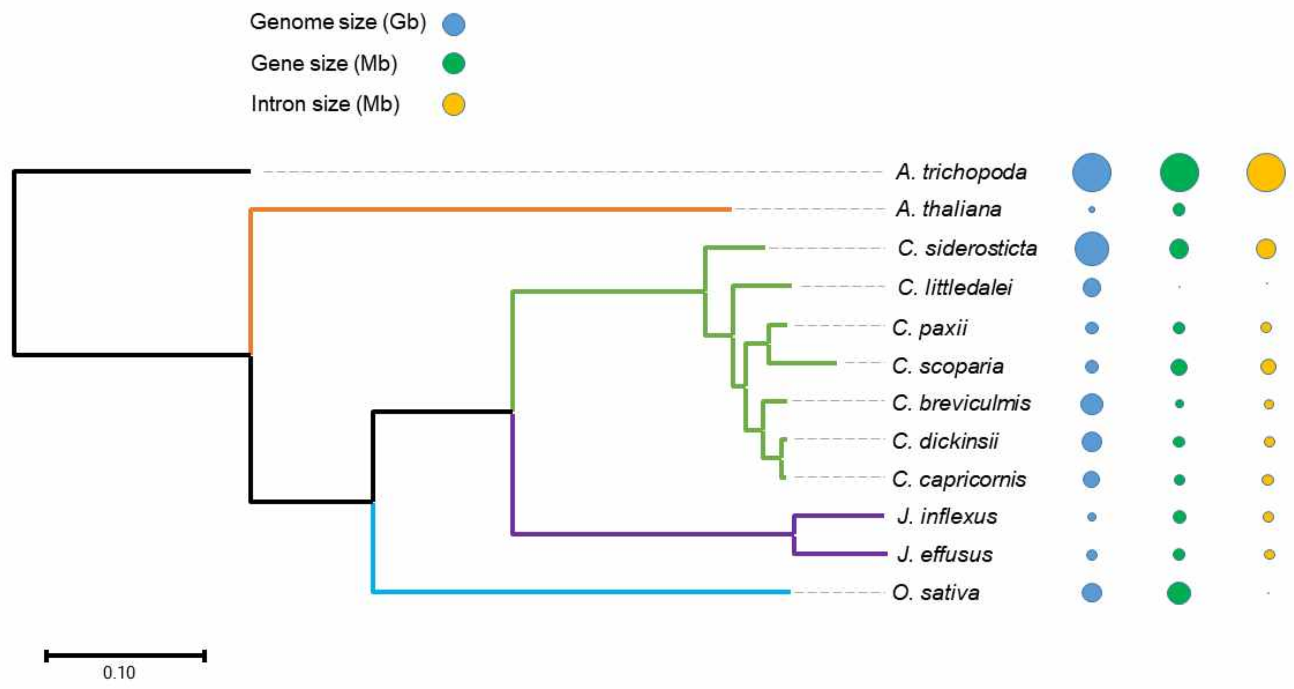


Fig. 14. Comparison of genome size, gene size, and intron size among twelve angiosperm species. Data of *Carex* from the current study was based on the BRAKER gene prediction tool with Viridiplantae data as protein evidence.

3.3 Identification of MADS−box genes and construction of phylogeny of MADS−box gene families

To explore the evolution of MADS−box genes in *Carex*, MADS−box genes in *Carex* and *Juncus* species were identified by BLASTP search and InterProScan filtering (Table 5), and a phylogenetic tree of MADS−box gene families was constructed (Fig. 15). Identification of MADS−box genes in *Arabidopsis*, *Amborella*, and rice has been well studied. Based on MADS−box genes from these three species as references, I identified MADS−box genes in the genomes of *Carex* and *Juncus* (Table 5). Approximately 900 MADS−box genes were identified from these twelve genomes (Table 5).

A phylogeny of MADS−box genes from twelve species based only on BLASTP search was constructed. For the global analysis of both Type Ⅰ and Type Ⅱ genes, genes not containing MADS−domain were excluded based on InterProScan search. Then, 785 genes remained, and the phylogenetic tree showed a clear grouping of Type Ⅰ and Type Ⅱ genes with a few exceptions (Fig. 15, Appendix Figure 1). *OsMADS64*, identified as the alpha group in the previous study, was included in the Type Ⅱ clade (Fig. 15). It is because of the potential misidentification of this gene in the previous study. *Arabidopsis AGL33*, an independent lineage in the previous study of *Arabidopsis* MADS−box gene phylogeny, was also placed in the Type Ⅱ clade (Fig. 15). The grouping between Type Ⅰ and Type Ⅱ genes is highly supported if these two genes are not considered (Fig. 15). In the previous study, Parenicova et al. recognized Mα, Mβ, Mγ, and Mδ groups in the Type I MADS-box genes of *Arabidopsis*. However, these subgroupss did not form a monophyletic clade, respectively, except for Mγ (Fig. 15). And these clades and their relationships were highly supported (Fig. 15).

Table 5. Identification of MADS-box genes for *Carex* and *Juncus*. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence

Subgenus in *Carex*	Taxon	Number of MADS-box genes detected with BlastP search (1e-5)	Number of MADS-box genes detected with BlastP and filtered with InterProScan
N/A	*Juncus effusus*	92	75
N/A	*Juncus inflexus*	75	63
Ⅰ. *Carex* subg. *Siderosticta*	*Carex siderosticta*	53	44
Ⅲ. *Carex* subg. *Euthyceras*	*Carex littledalei*	104	90
Ⅴ. *Carex* subg. *Vignea*	*Carex scoparia*	112	97
Ⅴ. *Carex* subg. *Vignea*	*Carex paxii*	57	50
Ⅵ. *Carex* subg. *Carex*	*Carex breviculmis*	43	36
Ⅵ. *Carex* subg. *Carex*	*Carex capricornis*	59	51
Ⅵ. *Carex* subg. *Carex*	*Carex dickinsii*	65	55
	Total	660	561

N/A, not applicable.

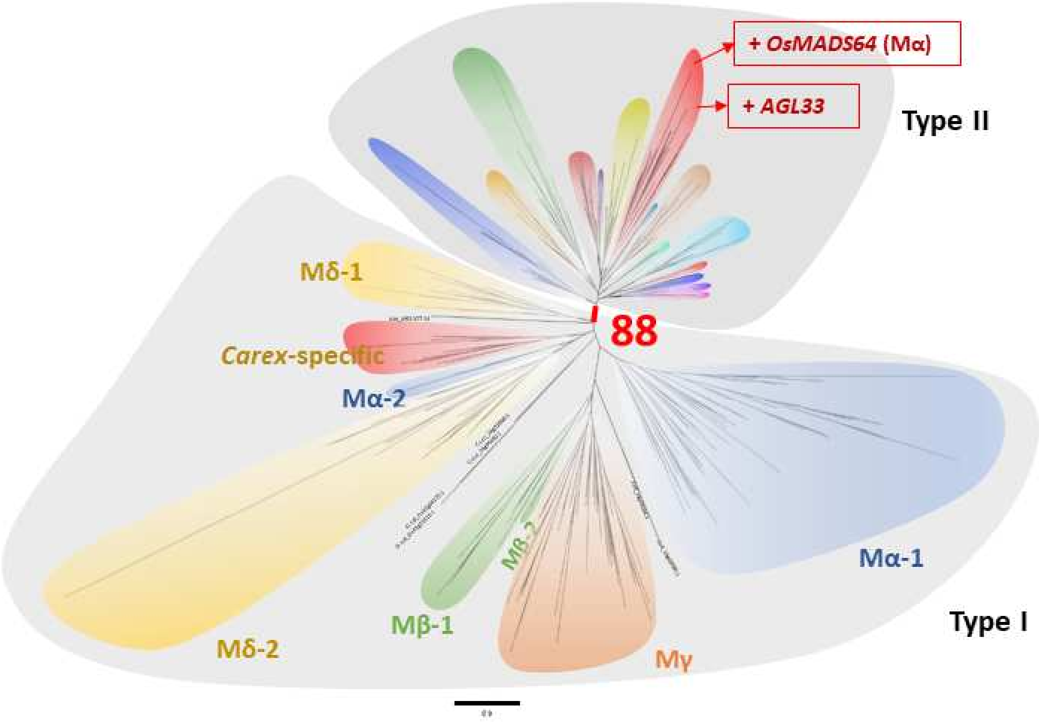
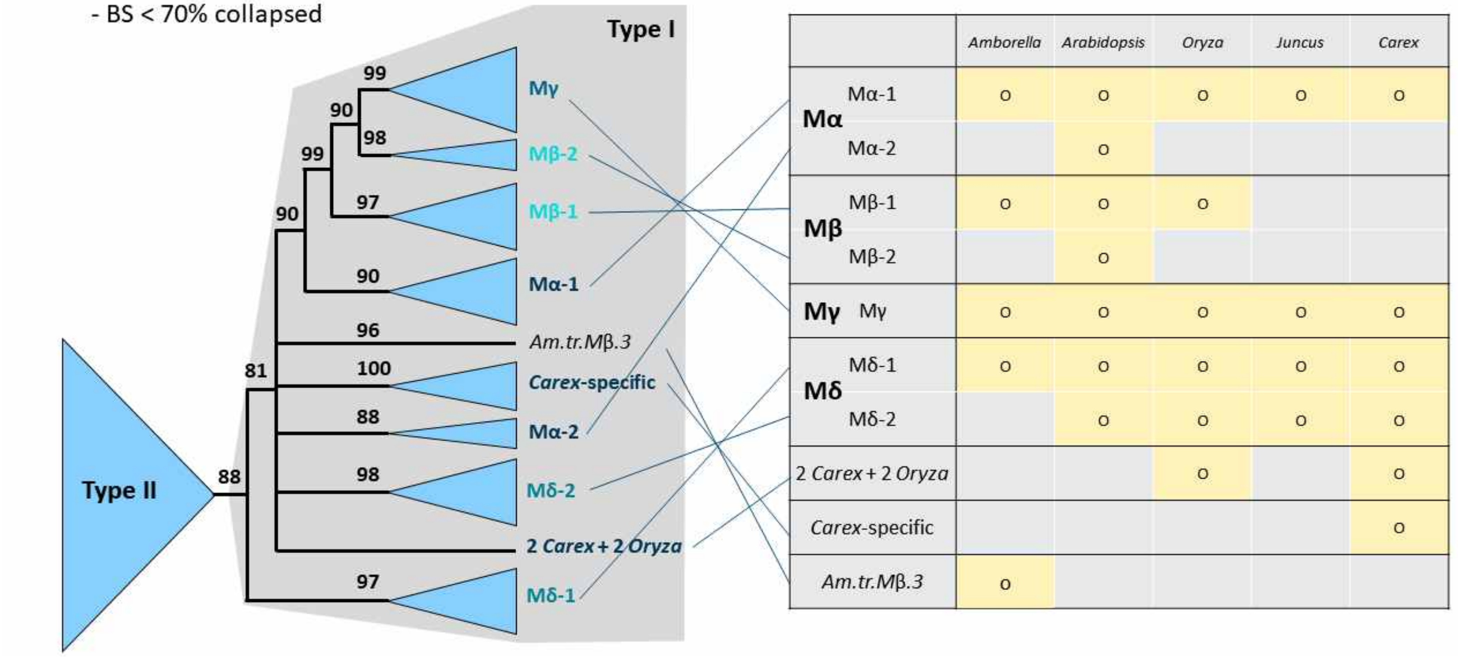


Fig. 15. Summarized phylogeny of MADS-box genes from twelve genomes. Red bold characters in red boxes indicate unexpected cases compared with Parenicova et al. (2003). Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.

I summarized the tree for Type Ⅰ genes, which shows less than 70 % bootstrap support collapsed (Fig. 16). All sequences detected with BlastP were used for phylogenetic analysis of Type Ⅰ genes. Subgroup recognition in Type Ⅰ MADS-box genes is not much studied after Parenicova et al. (2003). Once a genome has been determined in each taxon, MADS genes are analyzed based on the *Arabidopsis* MADS genes. However, their orthologies are often changed once I add more genes to the tree. For example, *Amborella* β3 was recognized during the *Amborella* genome project. But, when I added a bunch of representatives of Monocots, this gene is clearly placed outside of two β groups, and their relationship is highly supported (Fig. 16). Therefore, my results show that the four major subgroups of Type Ⅰ MADS-box genes defined based on the *Arabidopsis* genes should be divided into an additional number of subgroups when considering the evolution of the entire angiosperm MADS-box genes.

I summarized the tree for Type Ⅱ genes (Fig. 17). All sequences detected with BLASTP were used for phylogenetic analysis of Type Ⅱ genes. *Carex* genes were well-placed in the previously defined subfamilies of Type Ⅱ genes (Fig. 17). Seventeen subfamilies were recognized, but subfamilies *SQUA* and *FLC* show low bootstrap supports and AGL33, which originally formed an independent clade in *Arabidopsis*, was included in the Type Ⅱ phylogeny (Fig. 17). A clade marked as “MIXED” includes both Type I and Type II genes of monocots (Fig. 17). To increase the resolution of the tree, I extracted type Ⅱ genes from the entire MADS-box genes phylogeny and realigned. Maximum-likelihood trees based on realigned type Ⅱ genes were constructed using IQ-TREE. The phylogeny was shown with midpoint rooting using MEGA11 except for AGL33 (Fig. 18). All the Type Ⅱ subfamilies are well grouped with high bootstrap values (Fig. 18).

In each subfamily, I focused on shared-gene duplications among *Carex* species. In some lineages, *GGM13, GLO, DEF, OsMADS32,* and *AGL6* subfamilies, shared-gene duplications are not detected in *Carex*. Highly supported shared gene duplication across *Carex* species was detected in many subfamilies with high bootstrap supports (Fig. 19. A). For example, two rounds of duplications were detected in the *AGL12* subfamily. Similar duplication events were detected in *StMADS11*, *AGL17, SEP*, and *SQUA* subfamilies. However, in some cases, the duplication events were not highly supported in *AG, AGL17, TM3, SEP*, and *SQUA* subfamilies (Fig. 19. B). In the detected duplications, some of them were monocot-specific, some of them were *Juncus* plus *Carex*-specific, and some of them occurred before the diversification of *Carex* (Fig. 20).



		Amborella	*Arabidopsis*	*Oryza*	*Juncus*	*Carex*
Mα	Mα-1	o	o	o	o	o
	Mα-2		o			
Mβ	Mβ-1	o	o	o		
	Mβ-2		o			
Mγ	Mγ	o	o	o	o	o
Mδ	Mδ-1	o	o	o	o	o
	Mδ-2		o	o	o	o
2 *Carex* + 2 *Oryza*				o		o
Carex-specific						o
Am.tr.Mβ.3		o				

Fig. 16. Summarized phylogeny of Type I genes. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.

A.

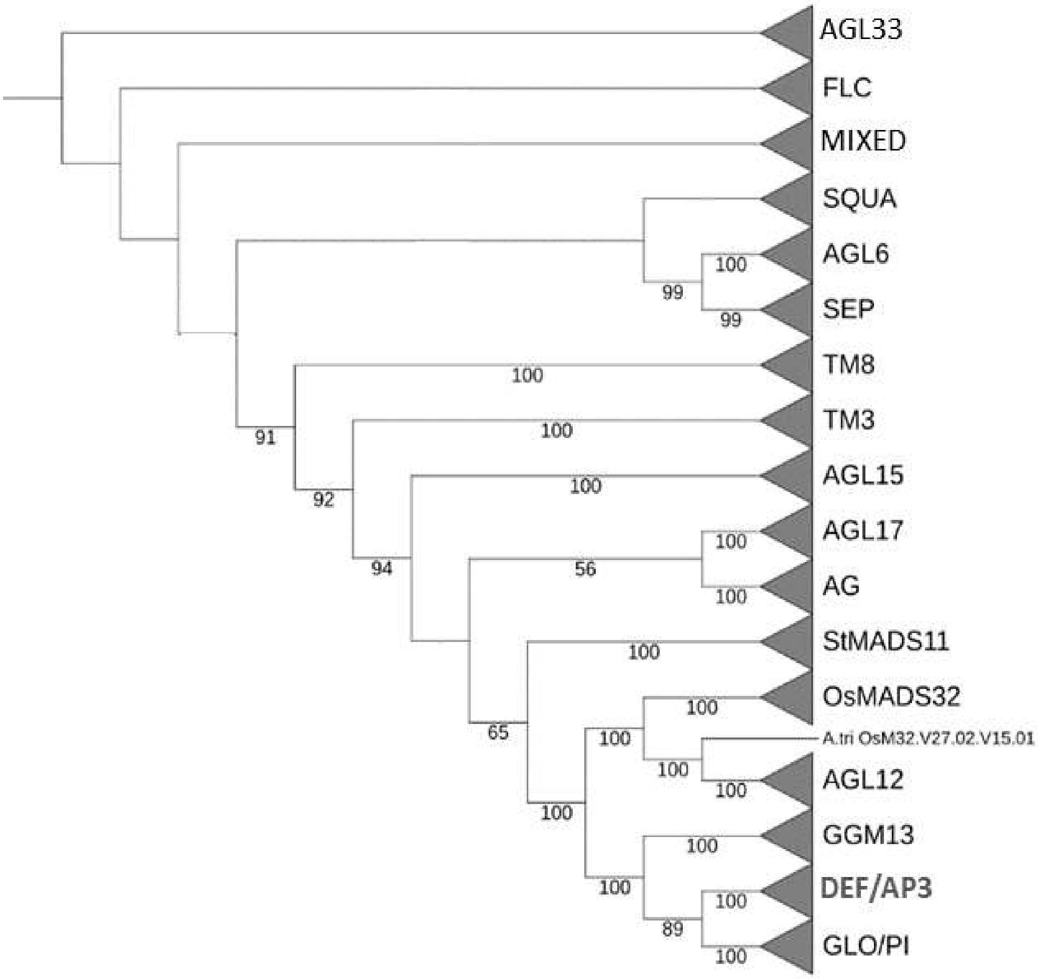


Fig. 17 (continued)

B.

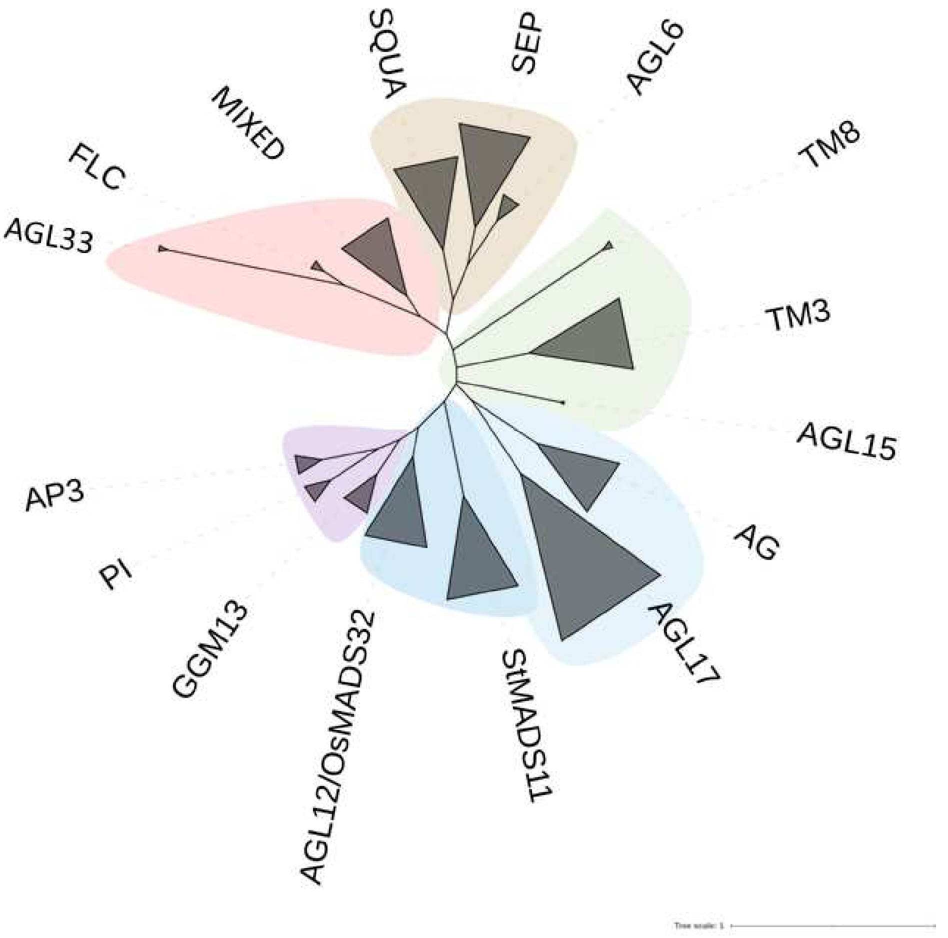


Fig. 17. Summarized phylogeny of Type Ⅱ genes. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence. **A.** A summarized cladogram showing 70% or higher bootstrap supports from the phylogeny of Type Ⅱ genes (Fig. 18) with AGL33 as an outgroup. **B.** Unrooted phylogram of Type II genes (Fig. 18). Different color areas indicate a subset of subfamilies analyzed together for better phylogenetic resolution.

Fig. 18. Phylogeny of type Ⅱ MADS-box genes from twelve genomes. Type Ⅱ genes were extracted from the entire BLASTP MADS-box genes phylogeny and realigned. Maximum-likelihood trees based on re-aligned type Ⅱ genes were constructed using the IQ-TREE. The midpoint rooting was applied. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.

Fig. 18, starting on the next page.

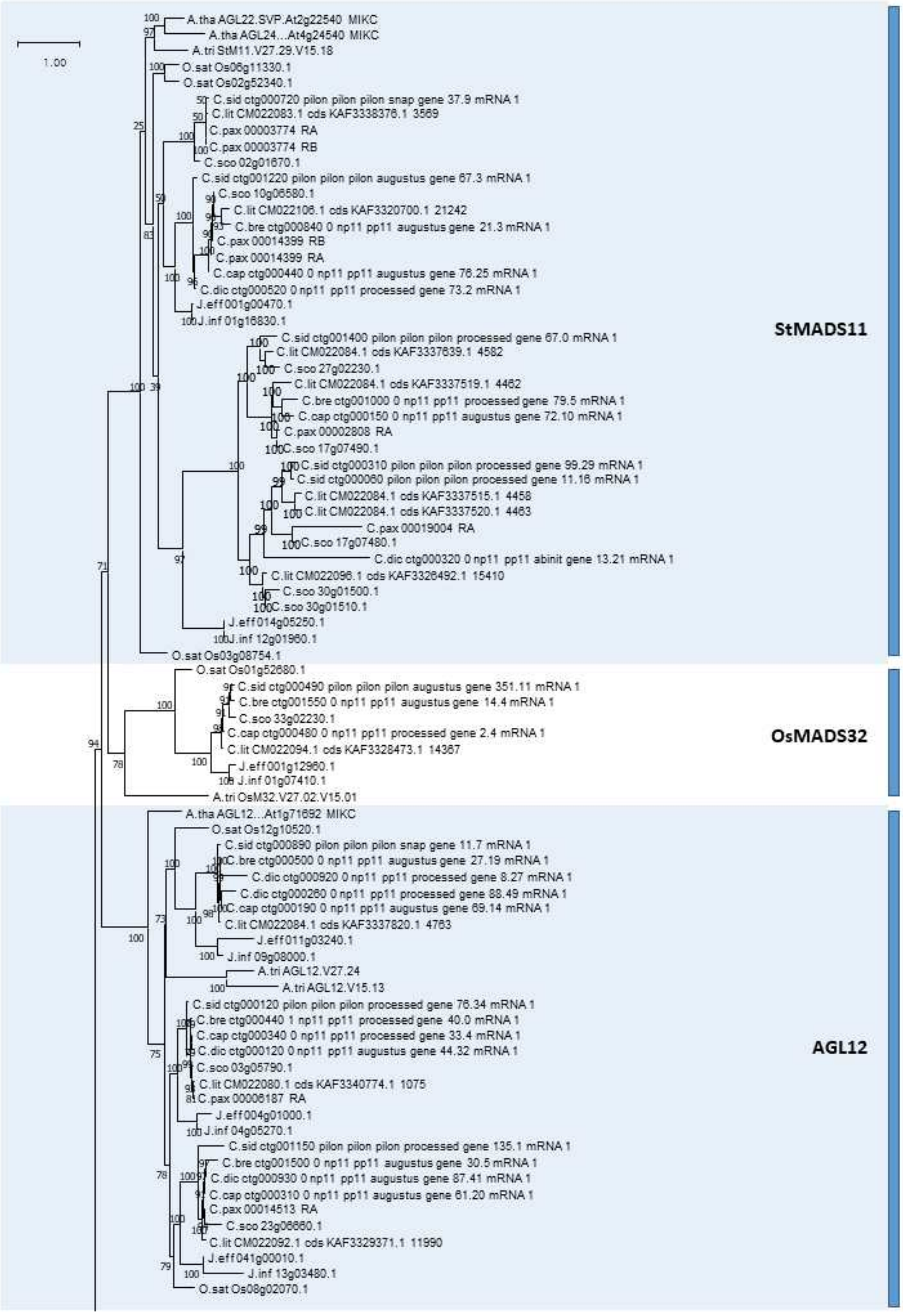


Fig. 18 (continued)

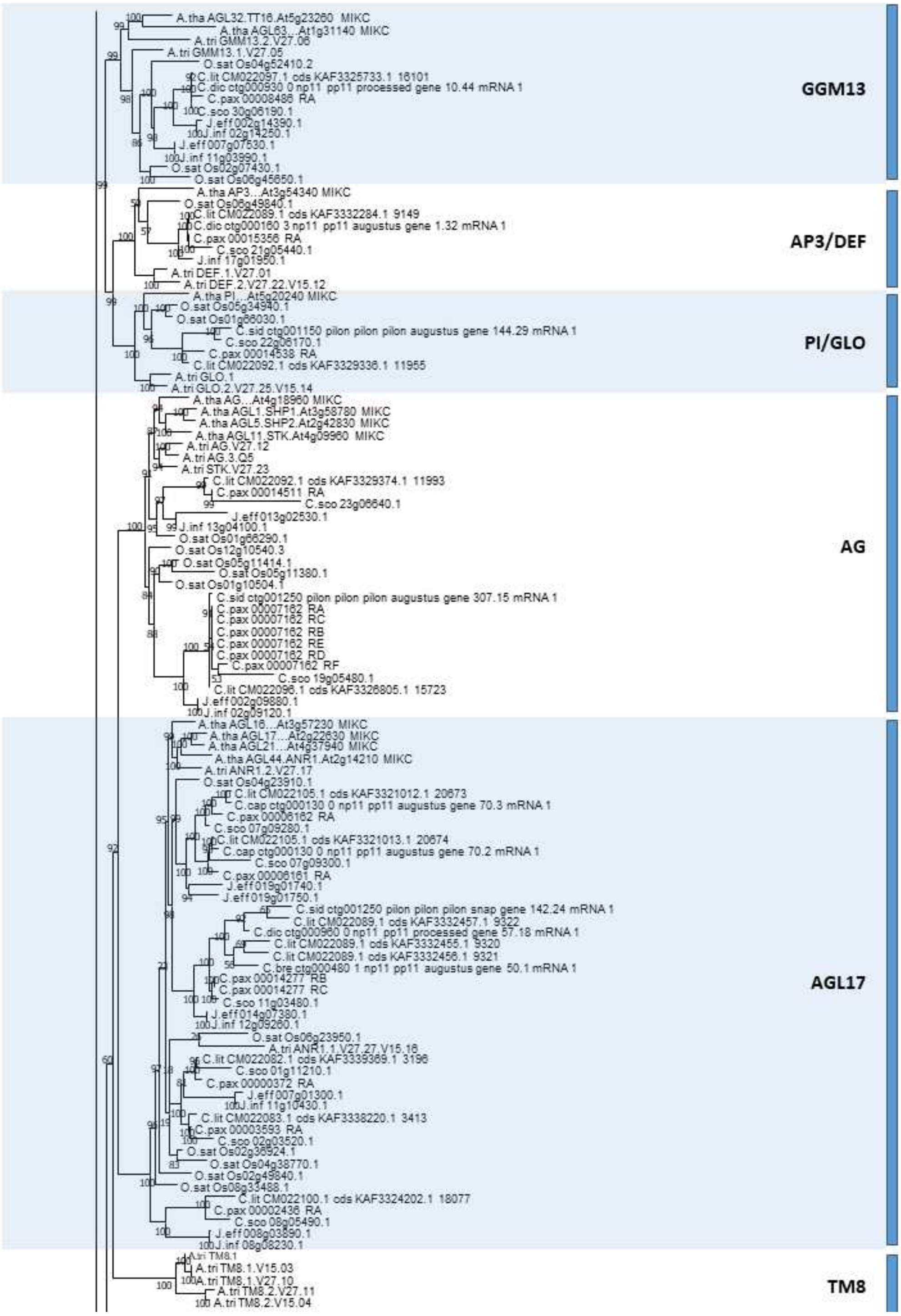


Fig. 18 (continued)

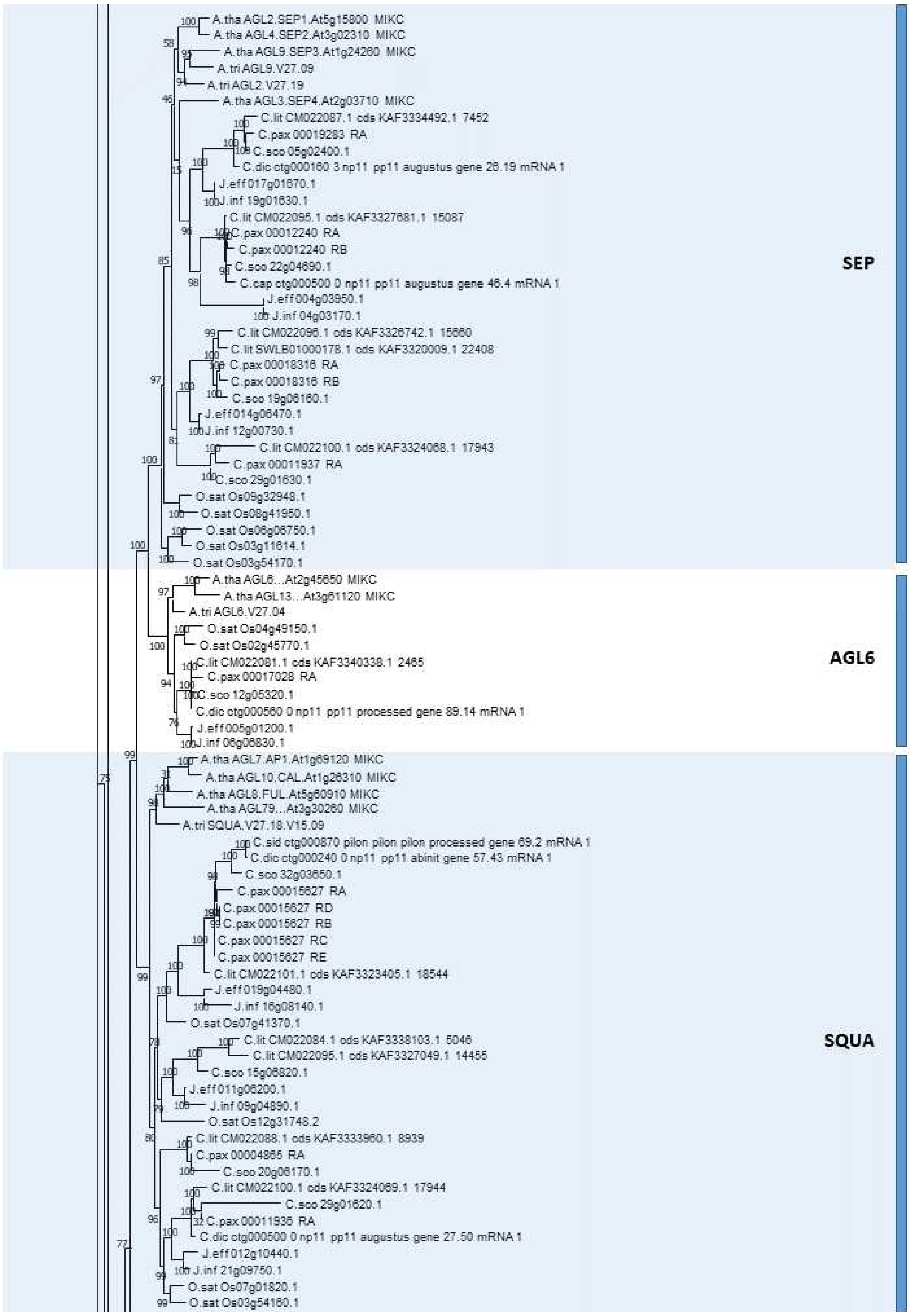


Fig. 18 (continued)

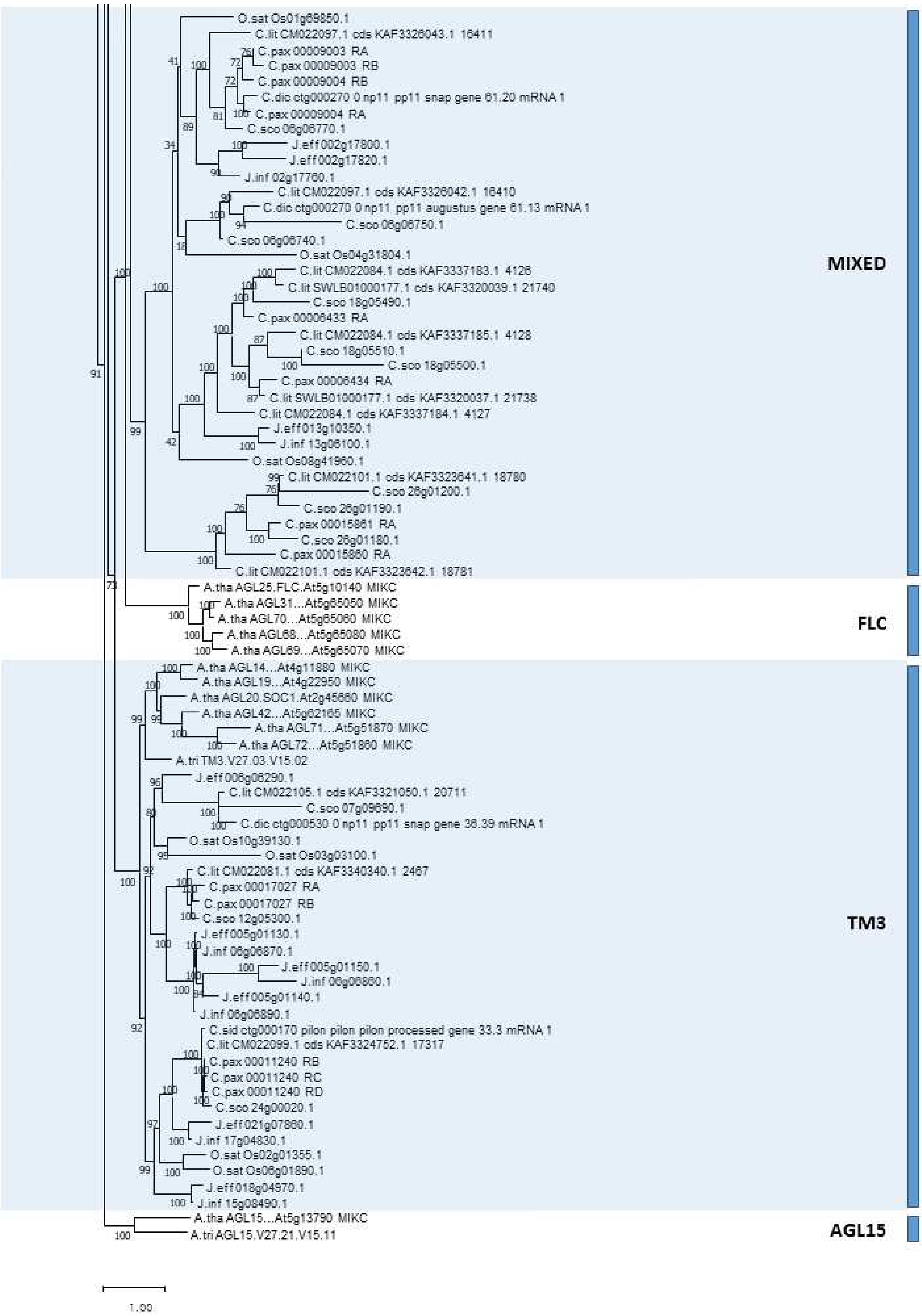

O.sat Os01g69850.1
C.lit CM022097.1 cds KAF3326043.1 16411
C.pax 00009003 RA
C.pax 00009003 RB
C.pax 00009004 RB
C.dic ctg000270 0 np11 pp11 snap gene 61.20 mRNA 1
C.pax 00009004 RA
C.sco 06g06770.1
J.eff002g17800.1
J.eff002g17820.1
J.inf 02g17760.1
C.lit CM022097.1 cds KAF3326042.1 16410
C.dic ctg000270 0 np11 pp11 augustus gene 61.13 mRNA 1
C.sco 06g06750.1
C.sco 06g06740.1
O.sat Os04g31804.1
C.lit CM022084.1 cds KAF3337183.1 4126
C.lit SWLB01000177.1 cds KAF3320039.1 21740
C.sco 18g05490.1
C.pax 00006433 RA
C.lit CM022084.1 cds KAF3337185.1 4128
C.sco 18g05510.1
C.sco 18g05500.1
C.pax 00006434 RA
C.lit SWLB01000177.1 cds KAF3320037.1 21738
C.lit CM022084.1 cds KAF3337184.1 4127
J.eff013g10350.1
J.inf 13g06100.1
O.sat Os08g41960.1
C.lit CM022101.1 cds KAF3323641.1 18780
C.sco 26g01200.1
C.sco 26g01190.1
C.pax 00015861 RA
C.sco 26g01180.1
C.pax 00015860 RA
C.lit CM022101.1 cds KAF3323642.1 18781
MIXED
A.tha AGL25.FLC.At5g10140 MIKC
A.tha AGL31...At5g65050 MIKC
A.tha AGL70...At5g65060 MIKC
A.tha AGL68...At5g65080 MIKC
A.tha AGL69...At5g65070 MIKC
FLC
A.tha AGL14...At4g11880 MIKC
A.tha AGL19...At4g22950 MIKC
A.tha AGL20.SOC1.At2g45660 MIKC
A.tha AGL42...At5g62165 MIKC
A.tha AGL71...At5g51870 MIKC
A.tha AGL72...At5g51860 MIKC
A.tri TM3.V27.03.V15.02
J.eff006g06290.1
C.lit CM022105.1 cds KAF3321050.1 20711
C.sco 07g09690.1
C.dic ctg000530 0 np11 pp11 snap gene 36.39 mRNA 1
O.sat Os10g39130.1
O.sat Os03g03100.1
C.lit CM022081.1 cds KAF3340340.1 2467
C.pax 00017027 RA
C.pax 00017027 RB
C.sco 12g05300.1
J.eff005g01130.1
J.inf 06g06870.1
J.eff005g01150.1
J.inf 06g06860.1
J.eff005g01140.1
J.inf 06g06890.1
C.sid ctg000170 pilon pilon pilon processed gene 33.3 mRNA 1
C.lit CM022099.1 cds KAF3324752.1 17317
C.pax 00011240 RB
C.pax 00011240 RC
C.pax 00011240 RD
C.sco 24g00020.1
J.eff021g07860.1
J.inf 17g04830.1
O.sat Os02g01355.1
O.sat Os06g01890.1
J.eff018g04970.1
J.inf 15g08490.1
TM3
A.tha AGL15...At5g13790 MIKC
A.tri AGL15.V27.21.V15.11
AGL15
1.00

Fig. 19. Example of shared duplications across *Carex* species. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence. **A.** Example of highly-supported shared gene duplications across *Carex* species. **B.** Example of potential shared gene duplications across *Carex* species.

A.

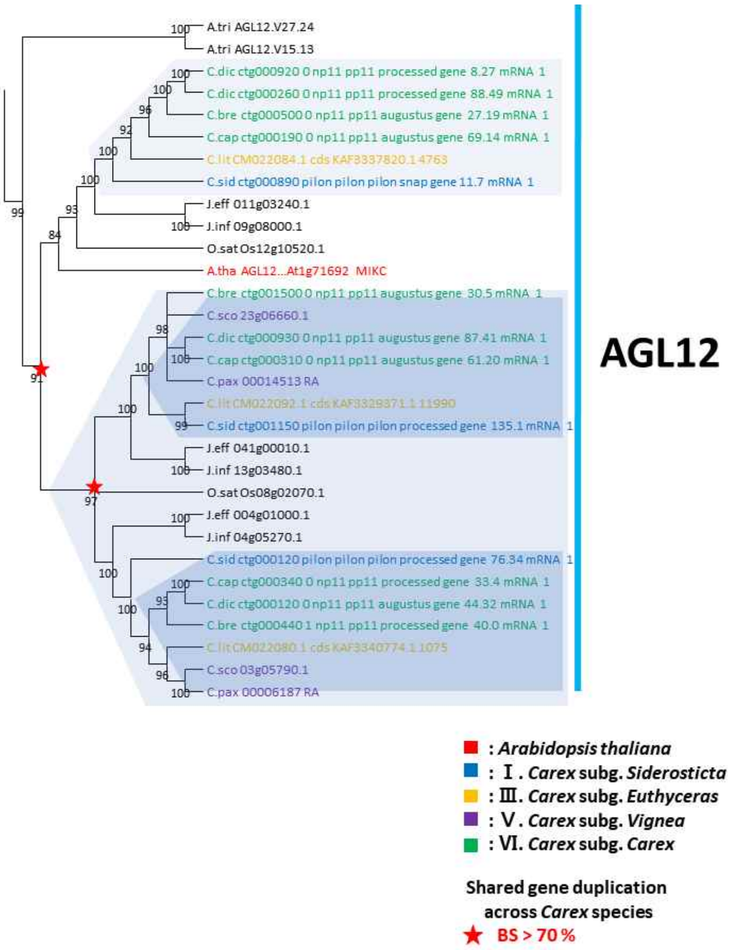
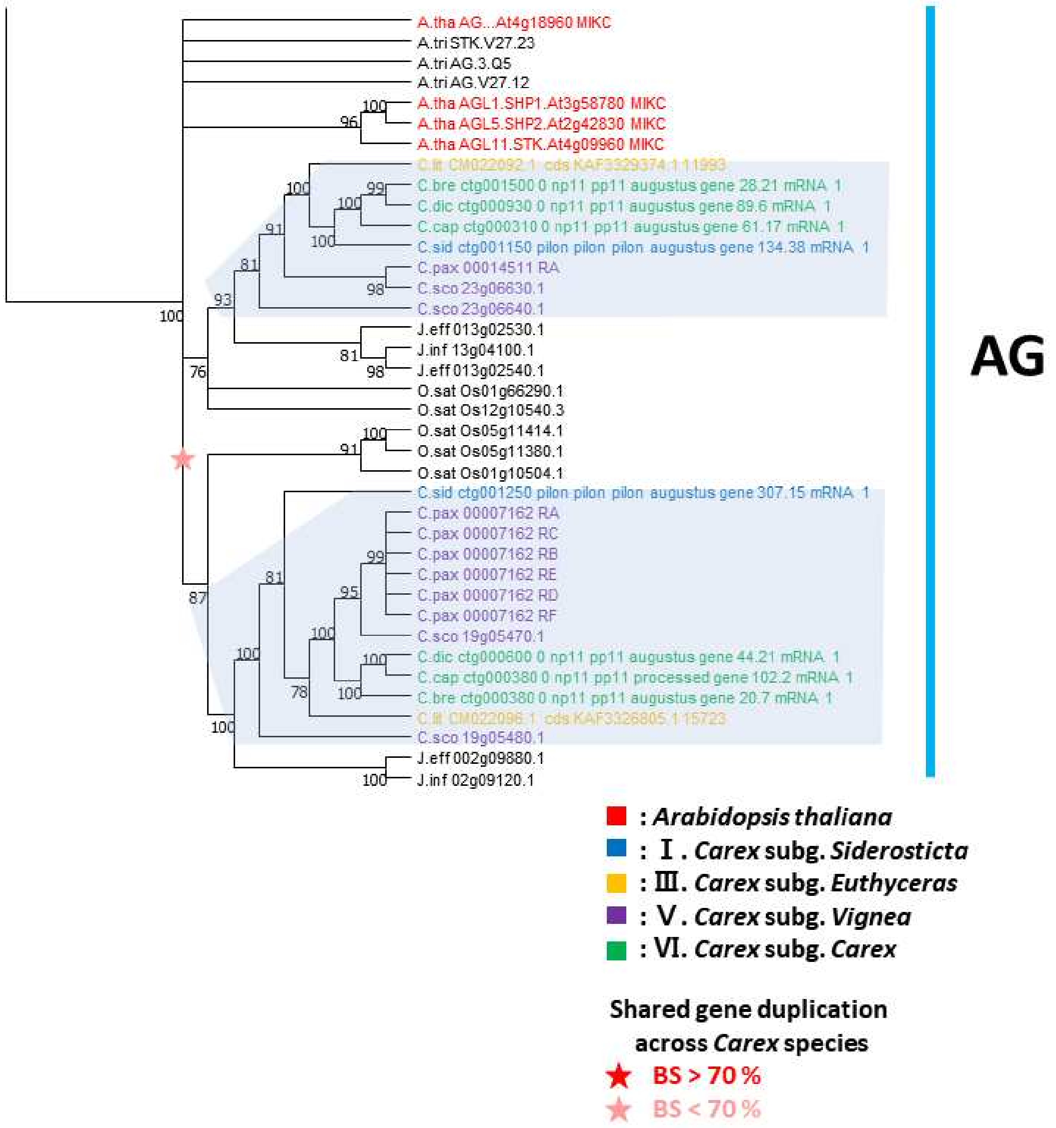


Fig. 19 (Continued)

B.


A.tha AG...At4g18960 MIKC
A.tri STK.V27.23
A.tri AG.3.Q5
A.tri AG.V27.12
A.tha AGL1.SHP1.At3g58780 MIKC
A.tha AGL5.SHP2.At2g42830 MIKC
A.tha AGL11.STK.At4g09960 MIKC
C.lit CM022092.1 cds KAF3329374.1 11993
C.bre ctg001500 0 np11 pp11 augustus gene 28.21 mRNA 1
C.dic ctg000930 0 np11 pp11 augustus gene 89.6 mRNA 1
C.cap ctg000310 0 np11 pp11 augustus gene 61.17 mRNA 1
C.sid ctg001150 pilon pilon pilon augustus gene 134.38 mRNA 1
C.pax 00014511 RA
C.sco 23g06630.1
C.sco 23g06640.1
J.eff 013g02530.1
J.inf 13g04100.1
J.eff 013g02540.1
O.sat Os01g66290.1
O.sat Os12g10540.3
O.sat Os05g11414.1
O.sat Os05g11380.1
O.sat Os01g10504.1
C.sid ctg001250 pilon pilon pilon augustus gene 307.15 mRNA 1
C.pax 00007162 RA
C.pax 00007162 RC
C.pax 00007162 RB
C.pax 00007162 RE
C.pax 00007162 RD
C.pax 00007162 RF
C.sco 19g05470.1
C.dic ctg000600 0 np11 pp11 augustus gene 44.21 mRNA 1
C.cap ctg000380 0 np11 pp11 processed gene 102.2 mRNA 1
C.bre ctg000380 0 np11 pp11 augustus gene 20.7 mRNA 1
C.lit CM022096.1 cds KAF3326805.1 15723
C.sco 19g05480.1
J.eff 002g09880.1
J.inf 02g09120.1
100
96
100
99
100
91
100
81
98
93
100
81
98
76
100
91
99
95
81
100
100
100
78
100
100
87
100
AG
🟥 : *Arabidopsis thaliana*
🟦 : Ⅰ. *Carex* subg. *Siderosticta*
🟨 : Ⅲ. *Carex* subg. *Euthyceras*
🟪 : Ⅴ. *Carex* subg. *Vignea*
🟩 : Ⅵ. *Carex* subg. *Carex*
Shared gene duplication across *Carex* species
★ BS > 70 %
★ BS < 70 %

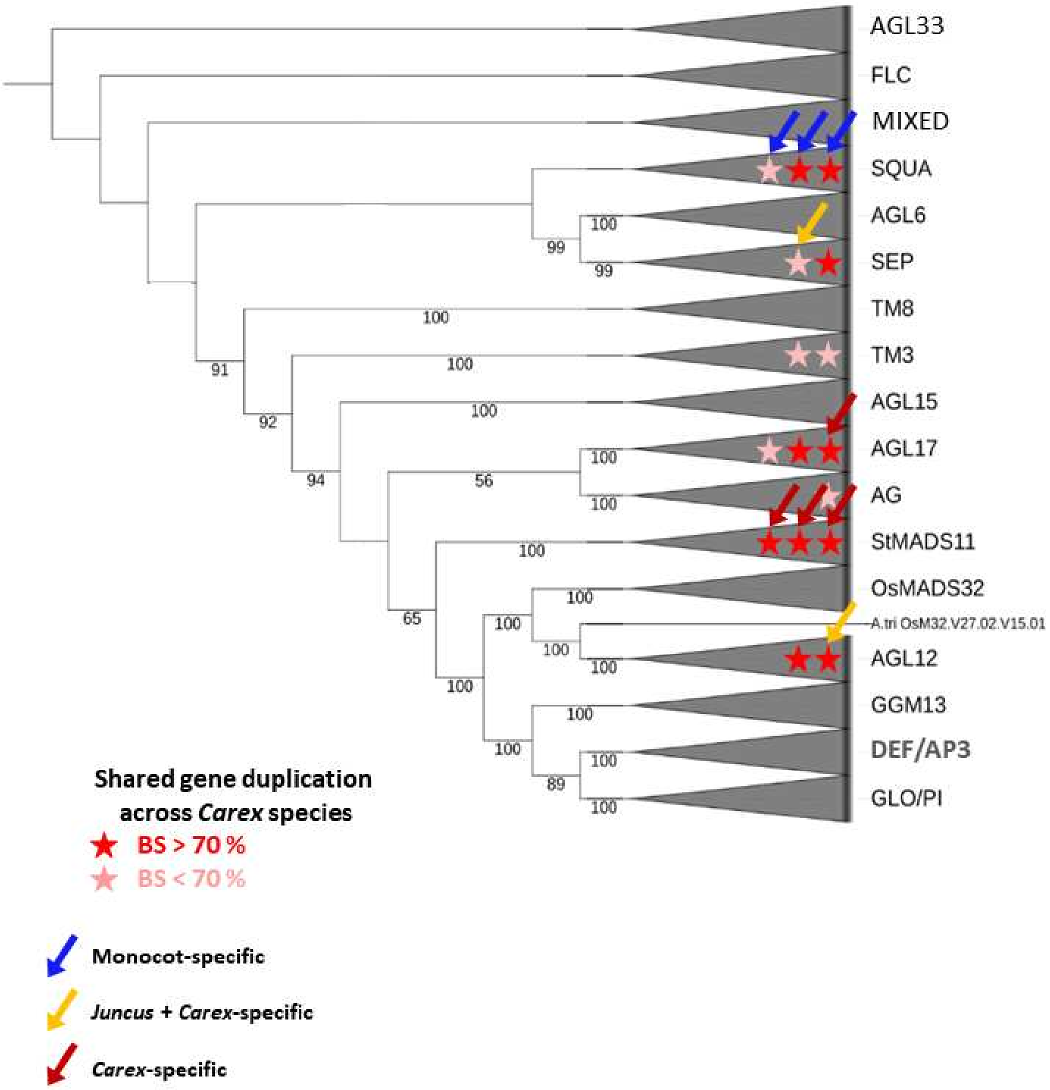


Fig. 20. Shared gene duplications of MADS-box genes in *Carex*. Three levels of clade-specific duplications are indicated as different colored arrows: monocot-specific (blue), *Juncus* + *Carex*-specific (yellow), and *Carex*-specific (brown). Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.

3.4 Difference in detected MADS-box genes according to gene prediction tools and protein evidence

I found that the gene prediction results generated were different when using different tools and different protein evidence during gene prediction. There were differences in the number of CDSs extracted from the generated gene prediction gff files, as well as in the number of identified MADS-box genes and the sequences themselves (Table 6, Table 7, Fig. 21). For *Carex dickinsii*, there are four types of data, the CDSs by the MAKER with *Carex littledalei* as protein evidence, the CDSs by the MAKER with Viridiplantae data as protein evidence, the CDSs by the BRAKER with Viridiplantae data as protein evidence, and the transcriptome sequences (Lee, 2021), so I performed a detailed comparative analysis of the MADS-box genes between them (Fig. 21). When comparing between gene prediction tools, the sequences by the BRAKER matched the transcriptome data the best (Fig. 21). In comparison between protein evidence, the number of CDSs in *Carex dickinsii* was 26,120 using *C. littledalei* and 28,046 using Viridiplante data. When using Viridiplantae data, the perfect match was 10 sequences, and when using *C. littledalei*, the perfect match was seven sequences (Fig. 21). So, using the entire green plant data, Viridiplantae, seems to produce slightly higher gene prediction results (Fig. 21). Therefore, I decided to use the BRAKER gene prediction tool with Viridiplantae data as protein evidence for the analyses of MADS-box genes families (Appendix Table 1). All sequences detected with the BLASTP search were shown in Appendix Table 1. However, since different tools and options yielded different gene prediction results (Table 6), I decided to use the combined sequences detected by the both gene prediction tools for MADS-box gene detection and functional studies of gene families (Table 8).

Table 6. A comparison of the number of CDSs predicted by MAKER with *C. littledalei* as protein evidence and BRAKER with Viridiplantae data as protein evidence

Species	MAKER_CDS	BRAKER_CDS
Carex breviculmis	21,852	16,546
Carex capricornis	20,685	19,040
Carex dickinsii	26,120	20,286
Carex paxii	30,749	22,935
Carex siderosticta	36,015	19,894

Table 7. A comparison of the number of detected MADS-box genes from CDSs predicted by MAKER with *C. littledalei* as protein evidence and BRAKER with Viridiplantae data as protein evidence

Species	MAKER		BRAKER	
	BlastP (1e-5)	BlastP & InterProScan	BlastP (1e-5)	BlastP & InterProScan
Carex breviculmis	63	30	43	36
Carex capricornis	54	28	59	51
Carex dickinsii	80	51	65	55
Carex paxii	73	63	57	50
Carex siderosticta	100	70	53	44

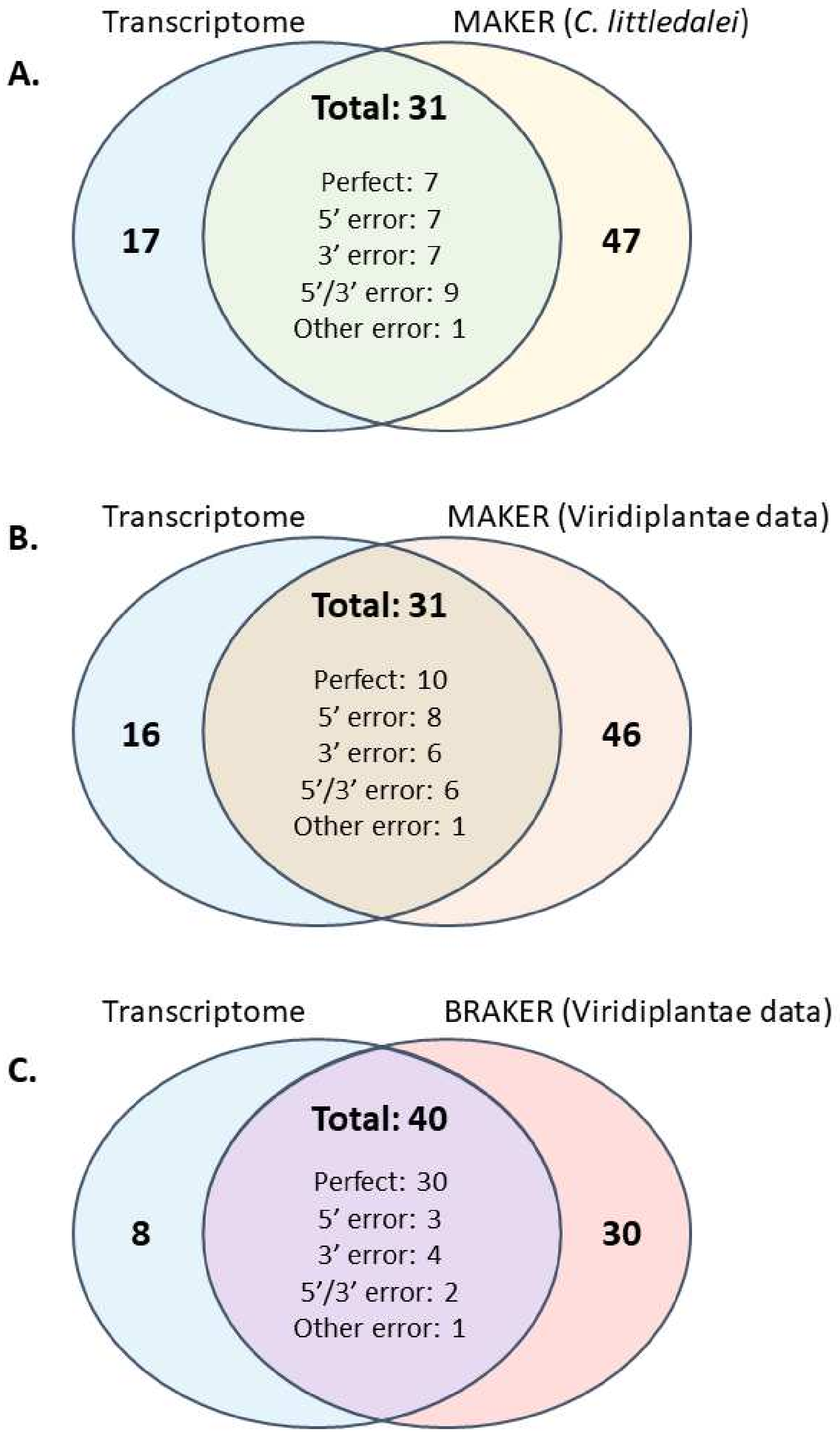

Transcriptome
MAKER (*C. littledalei*)
A.
Total: 31
Perfect: 7
5′ error: 7
3′ error: 7
5′/3′ error: 9
Other error: 1
17
47
Transcriptome
MAKER (Viridiplantae data)
B.
Total: 31
Perfect: 10
5′ error: 8
3′ error: 6
5′/3′ error: 6
Other error: 1
16
46
Transcriptome
BRAKER (Viridiplantae data)
C.
Total: 40
Perfect: 30
5′ error: 3
3′ error: 4
5′/3′ error: 2
Other error: 1
8
30

Fig. 21. Comparison of the number of MADS−box genes detected in *C. dickinsii* from the transcriptome and various gene prediction tools/protein evidence genomes: MAKER with *C. littledalei* as protein evidence (A), MAKER with Viridiplantae data as protein evidence (B), and BRAKER with Viridiplantae data as protein evidence (C).

Table 8. Detection of MADS-box genes and their subfamilies from the annotation using BRAKER (Viridiplantae data) and detection from MAKER (*C. littledalei*)

Class	Subfamily name	*A. trichopoda*	*A. thaiana*	O. sativa	*J. effusus*	*J. inflexus*	*C. siderosticta*	*C. littledalei*	*C. scoparia*	*C. paxii*	*C. breviculmis*	*C. capricornis*	*C. dickinsii*
B class	*PI/GLO*	o	o	o	x	x	M	o	o	o	o	o	o
B class	*AP3/DEF*	o	o	o	o	o	o	o	o	M	o	o	o
B sister	*GGM13*	o	o	o	o	o	M	o	o	M	o	o	o
	StMADS11	o	o	o	o	o	o	o	o	o	o	o	o
	OsMADS32	o	x	o	o	o	M	o	o	o	o	o	x
	AGL15	o	o	x	x	x	x	x	x	x	x	x	x
C class (and D class)	*AG*	o	o	o	o	o	M	o	o	o	o	o	o
	AGL17	o	o	o	o	o	o	o	o	o	o	o	o
	Only Juncus & Carex	x	x	x	o	o	o	x	o	o	o	o	o
	AGL12	o	o	o	o	o	o	o	o	o	o	o	o
A class	*SQUA*	o	o	o	o	o	M	o	o	o	o	o	o

Table 8- (Continued)

Class	Subfamily name	*A. trichopoda*	*A. thaiana*	O. sativa	*J. effusus*	*J. inflexus*	*C. siderosticta*	*C. littledalei*	*C. scoparia*	*C. paxii*	*C. breviculmis*	*C. capricornis*	*C. dickinsii*
E class	*SEP*	o	o	o	o	o	o	o	o	o	o	o	o
E sister	*AGL6*	o	o	o	o	o	M	o	o	o	o	o	o
	TM8	o	x	x	x	x	x	x	x	x	x	x	x
	TM3	o	o	o	o	o	M	o	o	o	M	o	o
	FLC	x	o	o	x	x	x	x	M	M	x	x	o

The letters in each cell represent the following;

o: detection from BRAKER (Viridiplantae data)

M: additional detection from MAKER (*C. littledalei*)

x: not detected

4. Discussion

In the preliminary analyses, all analyses were performed using the MAKER (*C. littledalei* as protein evidence) as a gene prediction tool. However, when I compared them in detail with the sequences from the PCR-based cDNA cloning in the previous study (Lee, 2021), many annotation-errors were recognized. In many cases, 5' end, 3' end, or both 5' and 3' ends differences are detected (Fig. 22). To compare the impact of annotation by different tools and options, I performed annotation by three different methods on the assembled genome of *C. dickinsii*: 1)MAKER with *C. littledalei* as protein evidence, 2) MAKER with Viridiplantae data as protein evidence, and 3) BRAKER with Viridiplantae data as protein evidence. Then, MADS-box genes from each were detected with the same method. As s result of the comparison, the genes obtained from "BRAKER with Viridiplantae data as protein evidence" best matched the MADS-box genes detected by PCR-based cDNA cloning in our previous study (Fig. 21). Therefore, I decided to use the BRAKER gene prediction tool with Viridiplantae data as protein evidence for the analyses of specific gene families, such as MADS-box gene (Appendix Table 1). All sequences detected with the BLASTP search were shown in Appendix Table 1. Different tools and options yielded different annotation results (Table 6). For comparative genomic analyses with other taxa, the best method among them should be selected and used. However, for phylogenetic or functional studies of gene families, it is reasonable to use the union of CDSs detected by different methods rather than selecting and using CDSs from one method (Table 8).

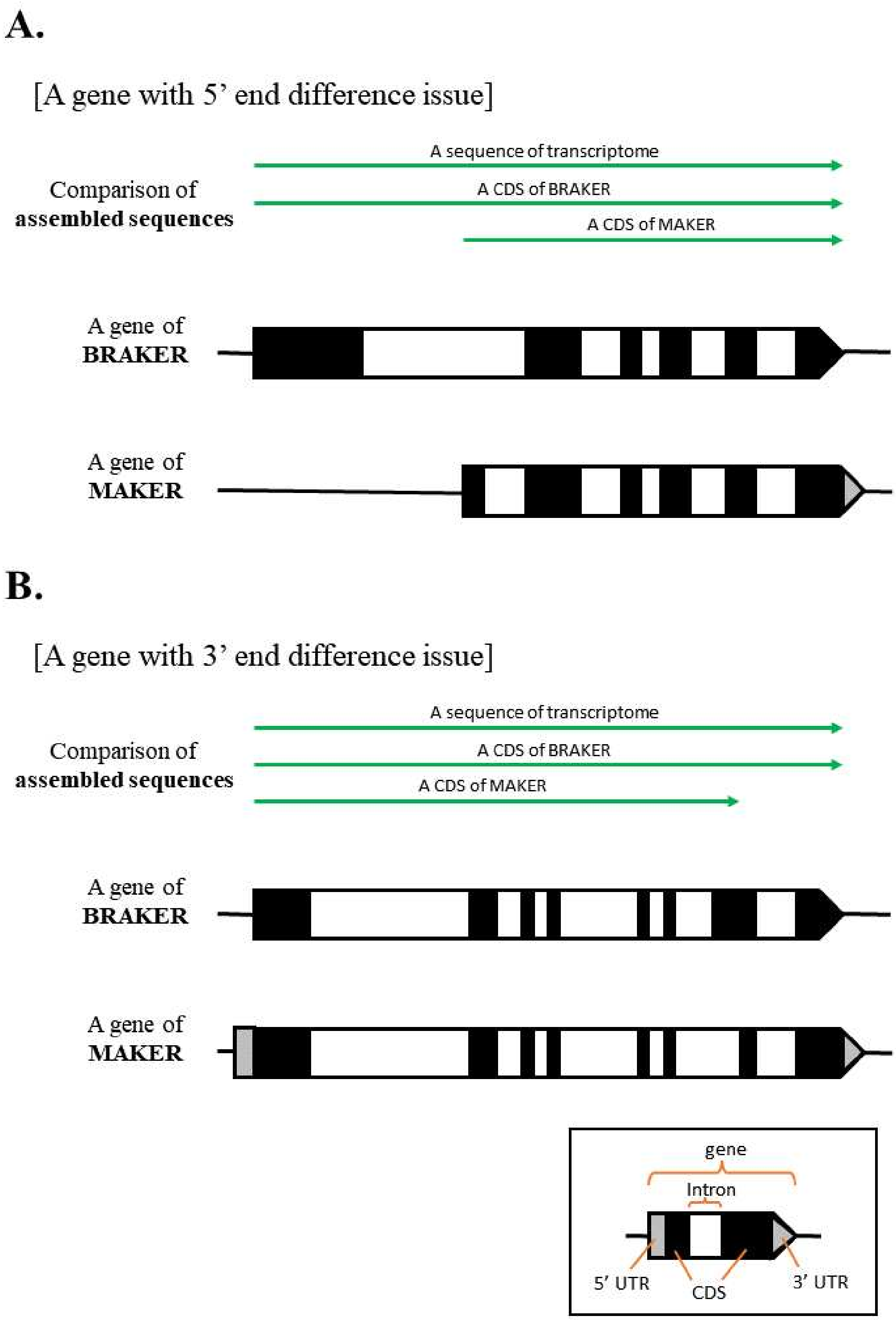

A.
[A gene with 5' end difference issue]
A sequence of transcriptome
Comparison of assembled sequences
A CDS of BRAKER
A CDS of MAKER
A gene of BRAKER
A gene of MAKER
B.
[A gene with 3' end difference issue]
A sequence of transcriptome
Comparison of assembled sequences
A CDS of BRAKER
A CDS of MAKER
A gene of BRAKER
A gene of MAKER
gene
Intron
5' UTR
CDS
3' UTR

Fig. 22 Examples of gene prediction differences between MAKER (*C. littledalei* as protein evidence) and BRAKER (Viridiplantae data as protein evidence) gene predictions. **A.** An example of 5’ end difference. **B.** An example of 3’ end difference.

5. Conclusion

Carex, included in Cyperaceae, is an important species in grassland ecosystems. Most studies of monocots have focused on the Poaceae group in Poales but have not paid much attention to the Cyperaceae. This study generated high-quality draft genomes from five representative species in *Carex*. Both long-read data sequenced with ONT and short-read data sequenced with Illumina or MGI were used to assemble the genomes. The sizes of assembled genomes ranged from 0.25 Gbp to 0.8 Gbp, and they were matched well compared to the genome size measured by flowcytometry. The total number of contigs of five assembled genomes ranged from 48 to 168. All five assembled genomes showed high complete BUSCOs values of approximately 95 %.

When I addressed orthologies of genes from six angiosperm species (*A. trichopoda, A. thaliana, O. sativa, J. effusus, C. siderosticta, and C. littledalei*), the shared orthogroups of all six species were 8,300 and 1,436 orthogroups were *Carex*-specific. The relationship of the family level of the constructed species tree was well-matched with known angiosperm phylogeny. However, different from the previous study by Villaverde et al. (2020), subg. *Carex*, which contains the largest number of species in the genus *Carex,* forms a sister to subg. *Euthyceras*. Detailed further phylogenetic analyses are needed to confirm the subgeneric relationships in *Carex*.

Carex has unique flowers whose floral structure is very different from typical angiosperm flowers. Its flowers are divided into pistillate and staminate flowers, and both have a structure called a scale. The perigynium covers gynoecium and is an especially unique floral structure. The evolutionary origin of these floral organs of *Carex* is unclear. As a fundamental step toward understanding the evolutionary origin of floral organs for *Carex*, MADS-box genes of *Carex* were detected, and *Carex*-specific MADS-box gene duplication events were investigated by phylogenetic analysis in this study. Genes of the MADS-box gene family act as transcription factors, and Type Ⅱ MADS-box genes are especially highlighted because they are involved in the identification of floral organs in angiosperms. In this study, I detected MADS-box genes of seven *Carex* and two *Juncus* species based on MADS-box genes from three angiosperms, *Amborella trichopoda*, *Arabidopsis thaliana*, and *Oryza sativa*, which are well-identified species. Phylogeny of MADS-box genes from twelve species was constructed with repeating align and IQ-TREE construction several times. The phylogenetic tree showed a clear grouping of Type Ⅰ and Type Ⅱ genes with a few exceptions. Detailed analyses of each of the Type Ⅰ and Type Ⅱ MADS-box genes were also performed. The results show that the four major subgroups of Type Ⅰ MADS-box genes defined based on the *Arabidopsis* genes should be divided into an additional number of subgroups when considering the evolution of the entire angiosperm MADS-box genes. In the analysis of the

Type Ⅱ MADS-box gene, seventeen subfamilies were recognized. In some lineages, *AGL12*, *StMADS11*, *AGL17*, *TM3*, *SEP*, and *SQUA* subfamilies, highly supported shared gene duplications across *Carex* species were detected. 'Monocot-specific' gene duplication events were found in *SQUA* subfamily, '*Juncus* + *Carex*-specific' gene duplication events were found in *SEP* and *AGL12* subfamilies. '*Carex*-specific' gene duplication events have been found in the *AGL17* and *StMADS11* subfamilies, of which *AGL17* in *Arabidopsis* promotes flowering (Han et al., 2008) and *StMADS11* in tomato is involved in the formation of flower pedicels and inflorescence meristems (Szymkowiak and Irish, 1999). Thus, in *Carex*, gene duplications in these subfamilies may have potentially influenced the evolution and diversification of *Carex*.

Identifying the complete set of a particular gene group, such as the MADS-box gene group in a genome, is critical for interpreting their expression and function. However, identifying a complete set of a particular gene group through computational analysis is an incomplete and challenging task. So, We must be aware that identifying genes with different bioinformatic methods may be different. The detection method of MADS-box genes in this study is to use the union of the results from all different gene prediction tools and options.

In this study, I determined the high-quality draft genomes of five representative species of *Carex* and performed identification and phylogenetic analyses of MADS-box genes as a first step to uncover the evolutionary origin of *Carex* flowers with unique structures. Our findings will provide insights into the evolutionary patterns of floral structure and genome evolution in *Carex*, which represents Cyperaceae.

LITERATURE CITED

Amborella Genome Project, V.A. Albert, W.B. Barbazuk, … L. Tomsho. 2013. The *Amborella* genome and the evolution of flowering plants. Science. 342(6165).

Bioinformatics, B. and S. Valencia. 2019. OmicsBox−Bioinformatics made easy. BioBam Bioinformatics. March 3, 2019.

Blum M., H. Chang, S. Chuguransky, … R.D. Finn. 2020. The InterPro protein families and domains database: 20 years on. Nucleic Acids Research. 49(D1): D344−D354.

Bruna, T., K.J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. 2021. BRAKER2: Automatic eukaryotic genome annotation with GeneMakr−EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics. 3(1): lqaa108, doi:10.1093/nargab/lqaa108.

Camacho, C., G. Coulouris, V. Avagyan, … T.L. Madden. 2008. BLAST+: architecture and applications. BMC Cioinformatics. 10: 421.

Can, M., W. Wei, H. Zi, … G. Qu. 2020. Genome sequence of *Kobresia littledalei*, the first chromosome−level genome in the family Cyperaceae. Scientific Data. 7: 175.

Cantarel B.L., I. Korf, S.M.C. Robb, … M. Yandell. 2008. MAKER: An easy−to−use annotation pipeline designed for emerging model organism genomes. Genome Research. 18: 188−196.

Chan, P. and T. Lowe. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods in molecular biology. 1962: 1-14.

Christenhusz, M. J. M., and J. W. Byng. 2016. The number of known plants species in the world and its annual increase. Phytotaxa. 261: 201-217.

Ciaffi, M., A.R. Paolacci, O.A. Tanzarella, and E. Porceddu. 2011. Molecular aspects of flower development in grasses. Sex Plant Reprod. 24: 247-282.

Coen, H.S., and E.M. Meyerowitz. 1991. The war of the whorls: Genetic interactions controlling flower development. Nature. 353: 31-37.

Cole, T.C.H., H. Hilger., and P. Stevens. 2019. Angiosperm phylogeny poster (APP) - Flowering plant systematics.

Colombo, L., J. Franken, E. Koetje, J. van Went, H.J. Dons, G.C. Angenent, and A.J. van Tunen. 1995. The petunia MADS box gene FBP11 determines ovule identity. Plant Cell. 7:1859-1868.

Egorova, T.V. 1999. The sedges (*Carex* L.) of Russia and adjacent states (within the limits of the former USSR). St. Petersburg State Chemical-Pharmaceutical Academy, St. Petersburg and Lissouri Botanical Garden, St. Louis.

Emms, D.M. and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology. 16: 157

Emms, D.M. and S. Kelly. 2017. STRIDE: Species tree root inference from gene duplication events. Molecular Biology and Evolution. 34: 3267−3278.

Emms, D.M. and S. Kelly. 2018. STAG: Species tree inference from all genes. bioRxiv. https://doi.org/10.1101/267914.

Emms, D.M. and S. Kelly. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 20:238.

Flynn J.M., R. Hubley, C. Goubert, … A.F. Smit. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. PNAS. 117: 9451−9457.

Frodin, D.G. 2004. History and concepts of big plant genera. Taxonomys. 53: 753−776.

GLOBAL *CAREX* GROUP (GCG). 2015. Making *Carex* monophyletic (Cyperaceae, tribe Cariaceae): a new broader circumscription. Botanical Journal of the Linnean Society. 179: 1−42.

GLOBAL *CAREX* GROUP (GCG). 2016. Megaphylogenetic specimen−level approaches to the *Carex* (Cyperaceae) phylogeny using ITS, ETS, and matK sequences: implications for classification. Systematic Botany. 41: 500−518.

GLOBAL *CAREX* GROUP (GCG). 2021. A framework infrageneric classification of *Carex* (Cyperaceae) and its organizing principles. Journal of Systematics and Evolution. 59: 726−762.

Gotz S., J.M. Garcia−Gomez, J. Terol, … A. Conesa. 2008. High−throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 36: 3420−3435.

Han, Ping., B. García−Ponce, G. Fonseca−Salazar, E.R.

Alvarez-Buylla, and H. Yu. 2008. AGAMOUS-LIKE 17, a novel flowering promoter, acts in a FT-independent photoperiod pathway. The Plant Journal. 55: 253-265.

Hip, A.L. 2007. Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). Evolution. 61: 2175-2194.

Hoff, K.J., S. Lange, A. Lomsadze, M. Borodovsky, and M. Stanke. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32:767-769.

Hoff, K.J., A. Lomsadze, M. Borodovsky, and M. Stanke. 2019. Whole-Genome annotation with BRAKER. Methods Mol Biol. 1962: 65-95, doi: 10.1007/978-1-4939-9173-0_5.

Hu, J., J. Fan, Z. Sun, and S. Liu. 2019. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 36: 2253-2255.

Hu, J., Z. Wang, Z. Sun, … S. Wang. 2023. An efficient error correction and accurate assembly tool for noisy long reads. bioRxiv. 2023.03.09.531669

Jones, P., D. Binns, H-Y. Chang, … S. Hunter. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30: 1236-1240.

Kang, M., A. Chanderbali, S. Lee, D.E. Soltis, P.S. Soltis, and S. Kim. 2023. High-molecular-weight DNA extraction for long-read sequencing of plant genomes: An optimaziation of standard methods. Applications in Plant Sciences. 11(3): e11528.

Kawahara, Y., M. de la Bastide, J.P. Hamilton, … T. Matsumoto. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice (N Y). 6: 4

Kearse, M., R. Moir, A. Wilson, … A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28: 1647−1649.

Kim, S., M. Yoo, V.A. Albert, J.S. Farris, P.S. Soltis, and D.E. Soltis. 2004. Phylogeny and diversification of B−function MADS−box genes in angiosperms: evolutionary and functional implications of a 260−million−year−old duplication. American Journal of Botany. 91: 2102−2118.

Kim, S., J. Koh, M. Yoo, … D.E. Soltis. 2005. Expression of floral MADS−box genes in basal angiosperms: implications for the evolution of floral regulators. The Plant Journal. 43: 724−744.

Kreczetowicz, V.I. 1935. *Carex* L. In: Komarov VL ed. Flora SSSR. Leningrad: Academy Science URSS. 3: 114−464.

Kukenthal, G. 1909. Cyperaceae−Caricoideae. In: Engler A eds. Das Pflanzenreich. Leipzig: Wilhelm Engelmann. IV. 20:1−824.

Larridon I., A.R. Zuntini, E. Leveille−Bourret, … W.J. Baker. 2021. A new classification of Cyperaceae (Poales) supported by phylogenomic data. Journal of Systematics and Evolution. 59: 852−895.

Lagesen, K., P. Hallin, E.A. Rødland, … D.W. Ussery. 2007.

RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research. 35: 3100–3108.

Lee, B. 2021. An integrative study of Cyperaceae: focused on ecpressions of floral MADS–box genes and the structural evolution of chloroplast genomes. Ph.D. diss., The graduate school of Sungshin University, Seoul.

Liang, SY and T. Koyama. 2010. *Carex* section Aulocystis Dumortier. In: Wu ZY, Raven PH, Hong DY eds. Flora of China. Acoraceae through Cyperaceae: Cyperaceae. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 23: 352–362.

Linnaeus, C. 1753.. Species plantarum. Impensis Laurentii Salvii. 1: 1–560.

Mackenzie, K.K. 1931–1935. Cyperaceae–Cariceae. North American Flora. 18: 1–478.

Manni, M., M.R. Berkeley, M. Seppey, F.A. Simão, and E.M. Zdobnov. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution. 38: 4647–4654.

Minh, B.Q., H.A. Schmidt, O. Chernomor, … R. Lanfear. 2020. IQ–TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution. 37: 1530–1534.

Mohlenbrock, R.H., P.W. Nelson (1999). Sedges: *Carex.* The Illustrated flora of Illinois. Volume 14. Southern Illinois

University Press. Carbondale.

Nam J., J. Kim, S. Lee, G. An, H. Ma, and M. Nei. 2004. Type Ⅰ MADS−box genes have experienced faster birth−and−death evolution than type Ⅱ MADS−box genes in angiosperms. Proceedings of the National Academy of Sciences of the United States of America. 101: 1910−1915.

Park, C., M. Allaby. 2017. A dictionary of environment and conversation. Oxford University Press.

Parenicova L., S. Folter, M. Kieffer,··· L. Colombo. 2003. Molecular and phylogenetic analyses of the complete MADS−box transcription factor family in *Arabidopsis*: new openings to the MADS world. Plant Cell. 15: 1538−1551.

Pelaz, S., G.S. Ditta, E. Baumann, E. Wisman, and M.F. Yanofsky. 2000. B and C floral organ identity functions require SEPALLATA MADS−box genes. Nature. 405: 200−203.

Pertea, G. and M. Pertea. 2020. GFF Utilities: GffRead and GffCompare. F1000Research. 9: ISCB Comm J−304.

Pinyopich, A., G.S. Ditta, B. Savidge, ···M.F. Yanofsky. 2003. Assesing the redundancy of MADS−box genes during carpel and ovule development. Nature. 424: 85−88.

Planta, J., Y.Y. Liang, H. Xin, ··· K.L. Childs. 2022. Chromosome−scale genome assemblies and annotations for Poales species *Carex cristatella, Carex scoparia, Juncus effusus,* and *Juncus inflexus*. G3. 12(10): jkac211.

Rice Full−Length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-Length cDNA Project Team, S. Kikuchi, … . Yasunishi. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science. 301: 376-379.

Smit, A.F.A., R. Hubely, and P. Green. 2013-2015. RepeatMakser Open-4.0. <http://www.repeatmasker.org>

Smith, D.L. and J.S. Faulkner. 1976. The inflorescence of Carex and related genera. Botanical Review. 42: 53-81.

Soltis, D.E., A.S. Chanderbali, S. Kim, M. Buzgo, and P.S. Soltis. 2007. The ABC model and its applicability to basal angiosperms. Annals of Botany. 100: 155-163.

Soltis, D.E., C.D. Bell, S. Kim, and P.S. Soltis. 2008. Origin and early evolution of angiosperms. Annals of the New York Academy of Sciences. 1133: 3-25.

Stanke, M., O., Schöffmann, B. Morgenstern, and S. Waack. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7: 62.

Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. doi: 10.1093/bioinformatics/btn013.

Starr, J.R., R.J. Bayer, and B.A. Ford. 1999. The phylogenetic position of *Carex* section Phyllostachys and its implications for phylogeny and subgeneric circumscription in *Carex*

(Cyperaceae). American Journal of Botany. 86: 563-577.

Sun, J., F. Lu, Y. Luo, L. Bie, L. Xu, and Y. Wang. 2023. OrthoVenn3: an integrated paltform for exploring and visualizing orthologous data across genomes. Nucleic Acids Research. 51(W1): W397-W403.

Szymkowiak, E.J. and E.E. Irish. 1999. Interactions between jointless and wild-type tomato tissues during development of the pedicel abscission zone and the inflorescence meristem. Plant Cell. 11: 159-175.

Tanya, Z.B., L. Reiser, D. Li, Y. Merzheritsky, R. Muller, E. Strait, and E. Huala. The *Arabidopsis* Information Resouce: Making and mining the "Gold Standard" annotated reference plant genome. Genesis. 53: 474-485.

Theißen, G. 2001. Development of floral organ identity: stories from the MADS house. Current Opinion in Plant Biology. 4: 75-85.

Theissen, G., R. Melzer. 2007. Molecular mechanisms underlying origin and diversification of the angiosperm flower. Annals of Botany. 100: 603-619.

Theißen, G., R. Melzer, and F. Rümpler. 2016. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development. 143: 3259-3271.

Villaverde, T., P. Jimenez-Mejias, M. Luceno, M.J. Waterway, S. Kim, B. Lee, … THE GLOBAL *CAREX* GROUP. 2020. A nex classification of *Carex* (Cyperaceae) subgenera supproted by HybSeq backbone phylogenetic tree. Botanical Journal of Linnean Society. 194: 141-163.

Vurture, G.W., F.J. Sedlazeck, M. Nattestad, … M.C. Scahtz. 2017. GenomeScope: fast reference-free genome profiling form short reads. Bioinformatics. 33: 2202-2204.

Waterway, M.J., T. Hoshino, and T. Masaki. 2009. Phylogeny, species richness, and ecological specialization in Cyperaceae tribe Cariceae. The Botanical Review. 75: 138-159.

Wick, R.R., L.M. Judd, C.L. Gorrie, and K.E. Holt. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics. 3(10): e000132.

Wick, R.R. and K.E. Holt. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLOS Computational Biology 18(1): e1009802.

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

ABSTRACT IN KOREAN

국문초록

이 승 연
생물학과
성신여자대학교 대학원

MADS-box 유전자는 진핵생물의 발달과 성장을 조절하는 전사인자로, 특히 ABC(D)E 모델을 통해 식물의 꽃 발달을 조절하는 데 중요한 역할을 한다. 다른 관련된 유전자들과 함께 MADS-box 유전자의 기능과 조절적인 네트워크를 밝히는 것은 피자식물의 진화를 이해하는 데 핵심적인 역할을 할 것이다. 그러나 아직 많은 주요 피자식물 계통에서 꽃의 MADS-box 유전자에 대한 연구가 수행되지 않았다. 약 2,000종이 속해있는 사초속은 피자식물 중 다섯 번째로 큰 속이며, 심피를 감싸는 주머니같은 기관인 과포(perigynium)를 포함한 독특한 꽃의 형태를 가진다. 벼목 내에서 MADS-box 유전자의 검출과 그 기능적 연구는 주로 사초과보다는 여러 모델 식물 종들을 포함하는 벼과에 집중되어 왔다. 본 연구에서는 사초과를 대표하는 사초속에서 MADS-box 유전자들을 식별했다. 첫 번째로 다섯 종의 사초속 식물들(대사초, 대구사초, 청사초, 양뿔사초, 도깨비사초)의 전체 유전체의 초안을 결정했다. 결정된 유전체들은 48~168개의 contigs를 가지는 고품질의 유전체이며, 94.5 % 이상의 BUSCO 값을 보인다. 다음으로 본 연구에서 결정한 다섯 종의 사초속 식물들과 더불어 이전에 보고된 두 개의 사초속 유전

체들과 벼목에 속하는 다른 대표 식물로써 두 종의 골풀과 식물들을 함께 분석하였다. 본 연구에서 결정한 사초속 식물 5종을 포함한 12 종의 종 계통수를 그린 결과 기존에 잘 알려진 피자식물의 계통수와 일치하였다. 사초속 내부에서는 subg. *Carex*가 subg. *Euthyceras*와 자매그룹을 형성하였다. 여섯 종의 피자식물(*A. trichopoda, A. thaliana, O. sativa, J. effusus, C. siderosticta*, and *C. littledalei*)의 공통 orthogroups은 8,300개이며, 1,436개의 orthogroups이 사초속 특이적임을 확인하였다. MADS-box 유전자의 검출에는 잘 연구된 *Amborella trichopoda*, *Arabidopsis thaliana*, 그리고 *Oryza sativa*를 대표 식물로써 사용하였다. 12종의 유전체로부터 대략 900개의 MADS-box 유전자들이 식별되었다. 단백질 서열로 align한 후 DNA matrix로 maximum-likelihood 계통수를 생성하였고 Type Ⅰ과 Type Ⅱ 유전자들 각각의 세부 분석을 진행하였다. 전체 MADS-box 유전자들의 계통수는 몇 개의 예외 사항을 제외하고 Type Ⅰ과 Type Ⅱ 유전자들은 각각 높은 지지도의 분계군으로 묶였다. Type Ⅰ 분석 결과는 전체 피자식물의 MADS-box 유전자들의 진화를 고려할 때는 *Arabidopsis* 연구를 바탕으로 한 네 개의 Type Ⅰ subgruops은 더 많은 subgroups로 나눠야 함을 제시한다. Type Ⅱ 유전자들만의 계통 분석 결과 17개의 subfamilies가 인식되었다. *AGL12, StMADS11, AGL17, TM3, SEP*, 그리고 *SQUA* subfamilies에서는 각각 사초속의 기원 이후에 유전자 중복이 일어난 것이 확인되었고, 이러한 중복 현상은 유전자 계통수에서 강력하게 지지 되었다. 특히 *AGL17*과 *StMADS11*에서는 각각 사초속 특이적 중복이 발견되었다. 완성된 유전체 서열로부터 MADS-box 유전자들을 검출하는 방법들 또한 논의되었다. 본 연구는 사초과의 사초속 식물들을 대표하는 식물들의 전체 유전체를 밝혀냈고, 이들의 MADS-box 유전자들을 검출하여 이들의 유전자 진화의 역사를 제시하였다. 이들 유전자들에 대한 향후 연구는 대부분의 다른 피자식물들과는 상이한

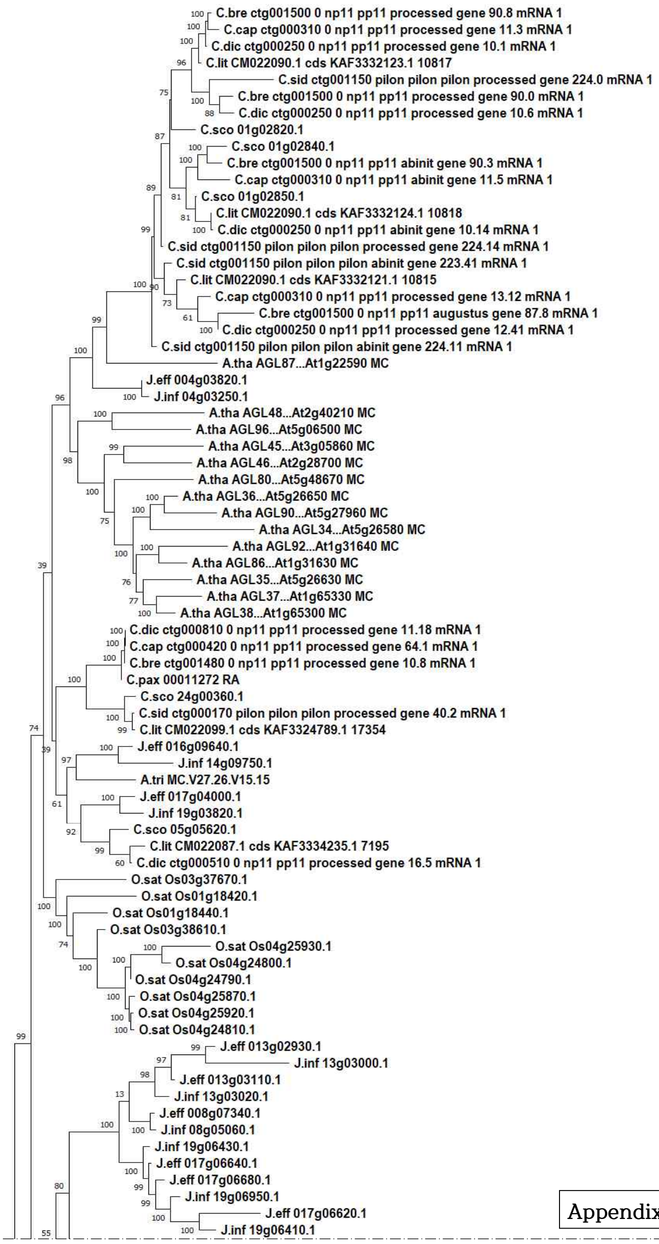
구조를 갖는 사초속 식물 꽃의 각 부분에 대한 기원과 형태적 진화를 밝히는 데 기여할 것으로 기대된다.

APPENDICES

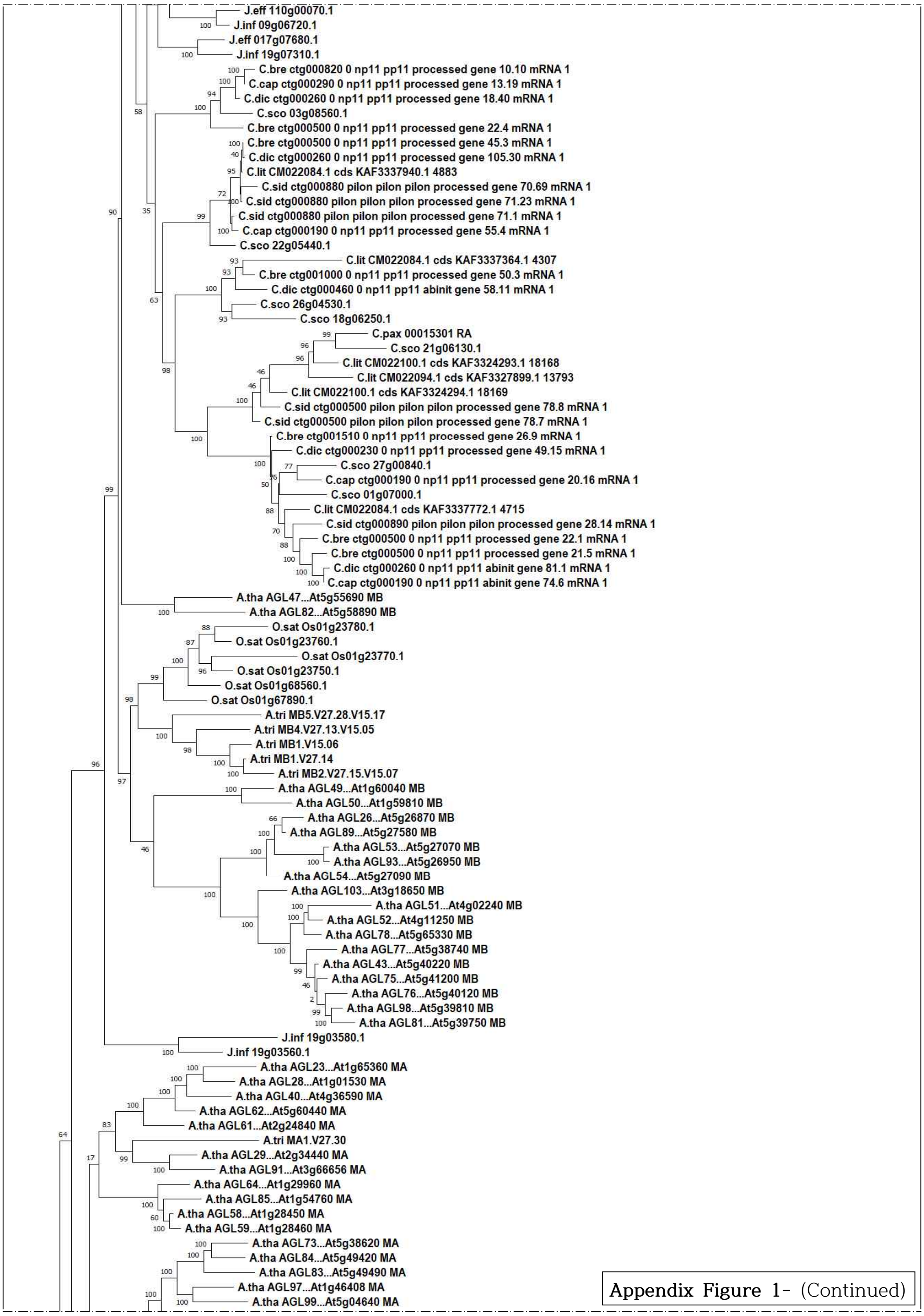
Appendix Figure 1. Phylogeny of MADS−box genes from twelve genomes.

Appendix Table 1. List of MADS−box genes detected in this study.

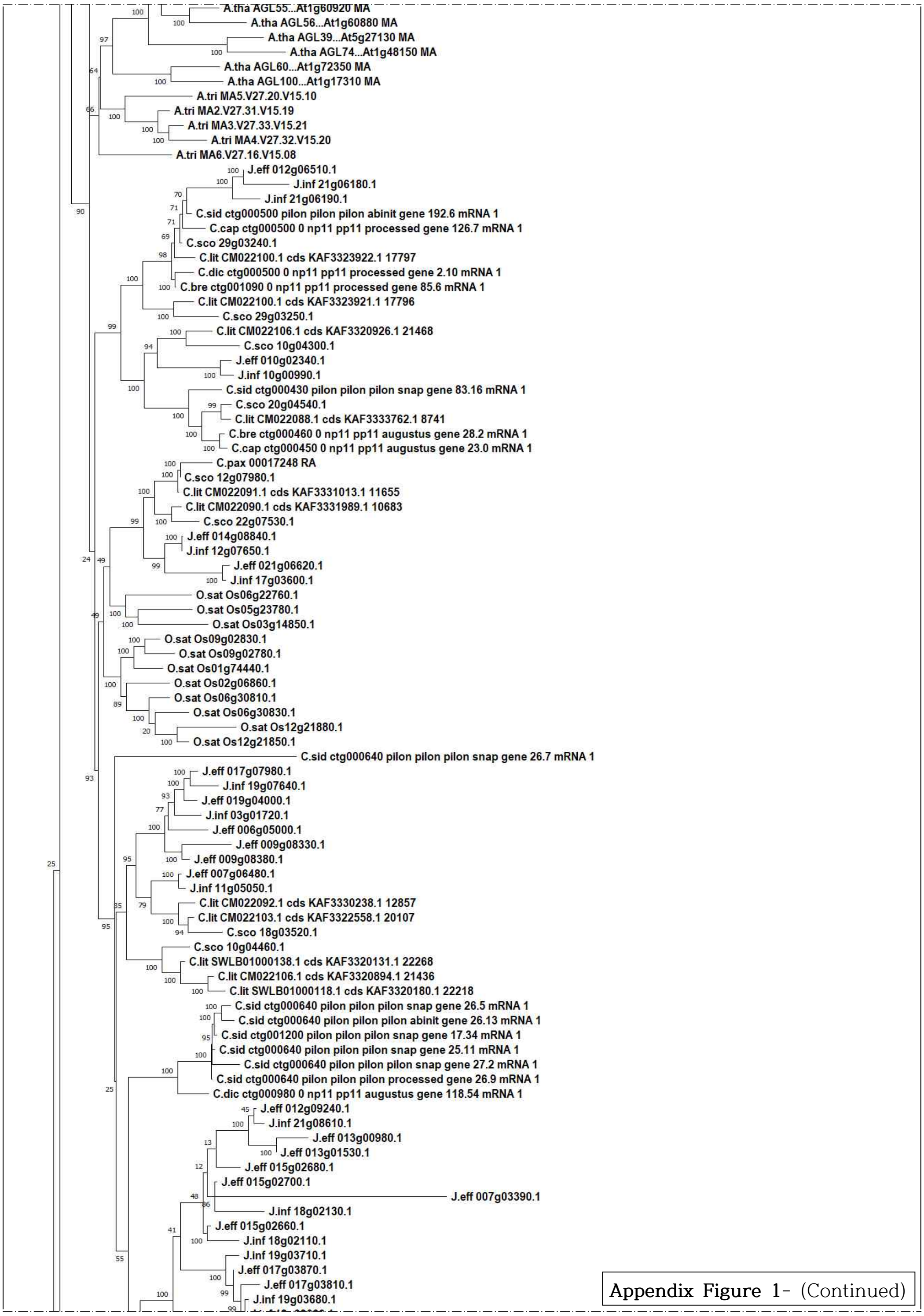
Appendix Figure 1. Phylogeny of MADS-box genes from twelve genomes. The genes were searched based on BlastP (1e-5) and InterProScan. Data of *Carex* from the current study was based on the MAKER gene prediction tool with *C. littledalei* as protein evidence.



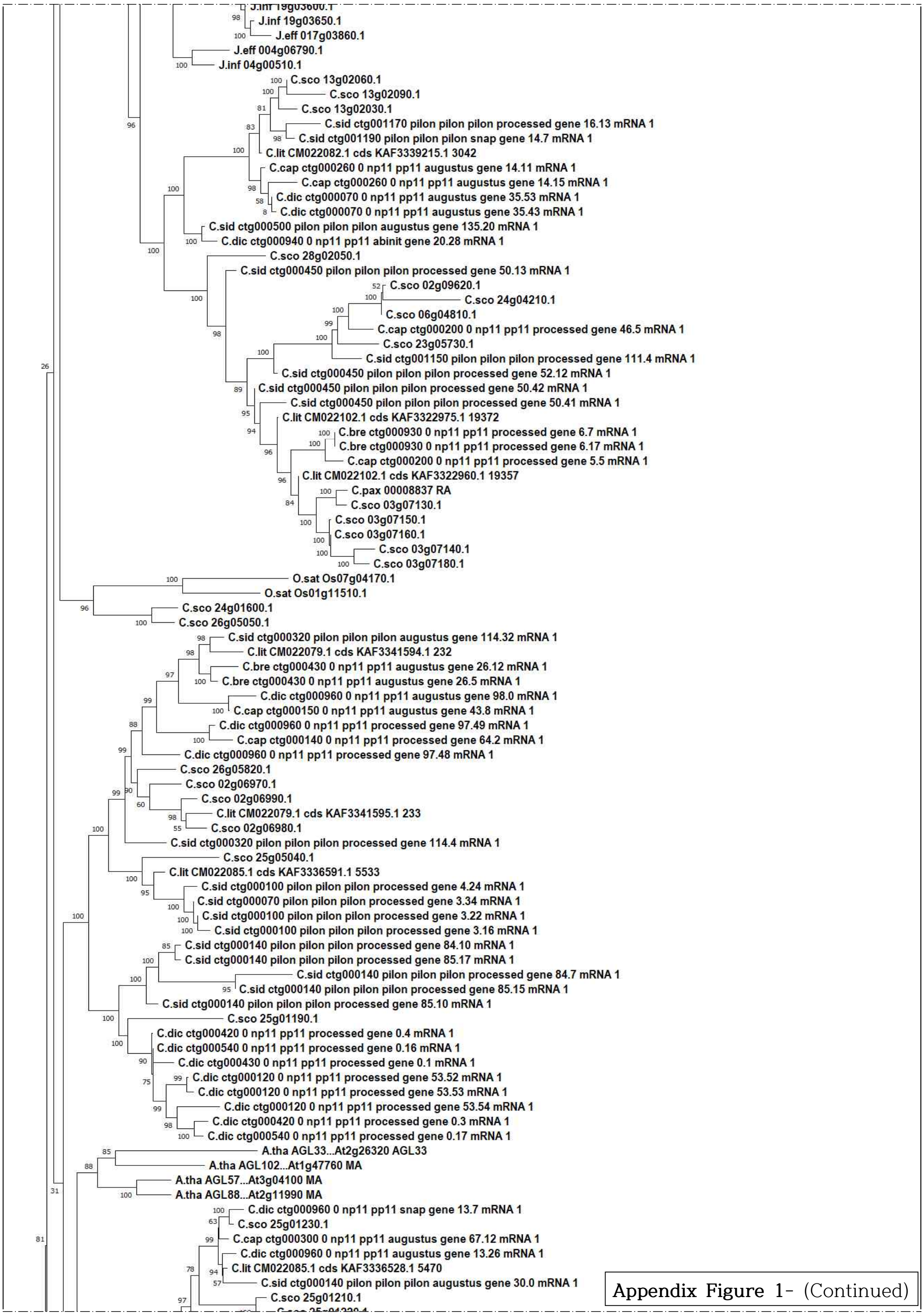
Appendix Figure 1- (Continued)



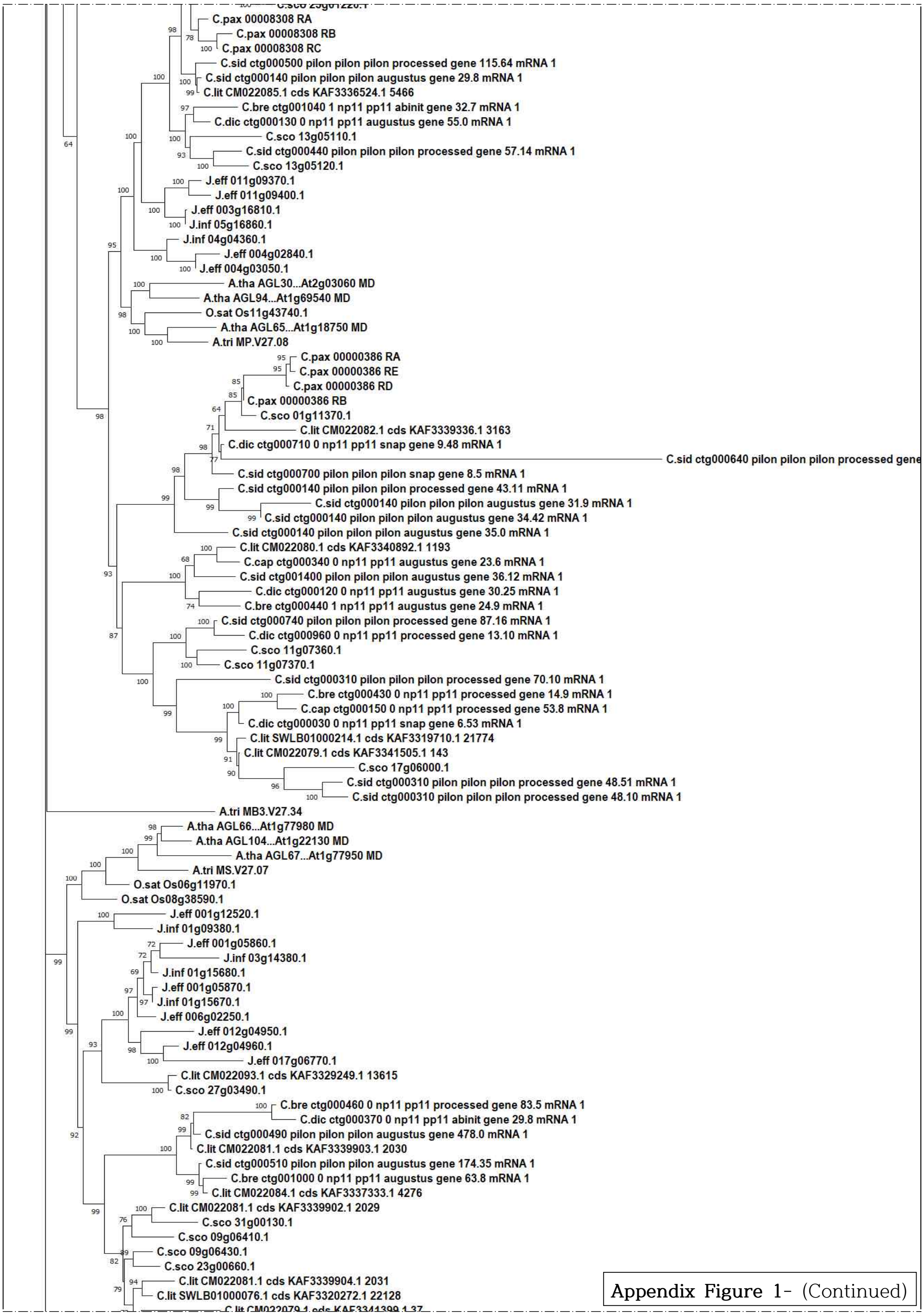
Appendix Figure 1- (Continued)



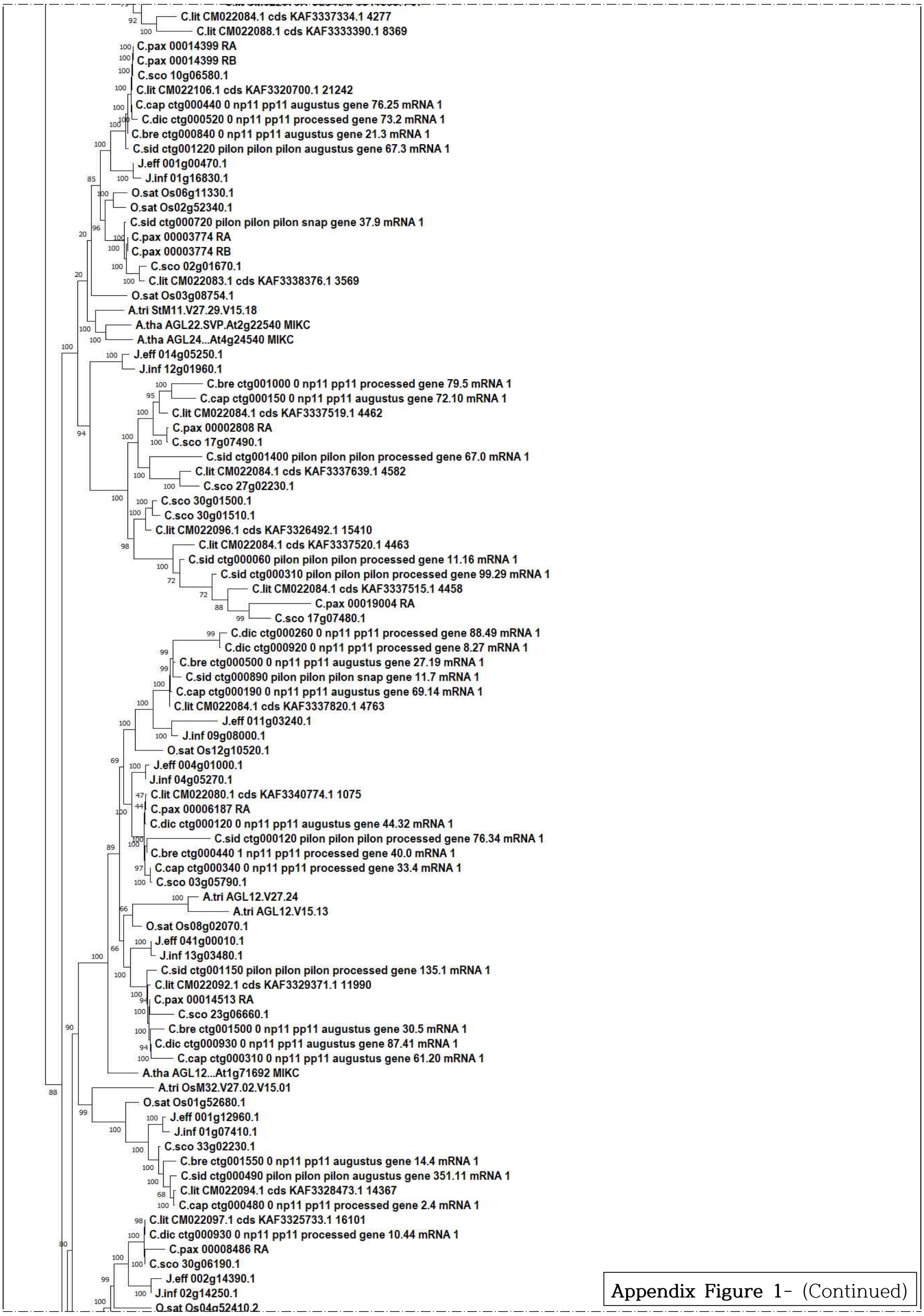
Appendix Figure 1- (Continued)



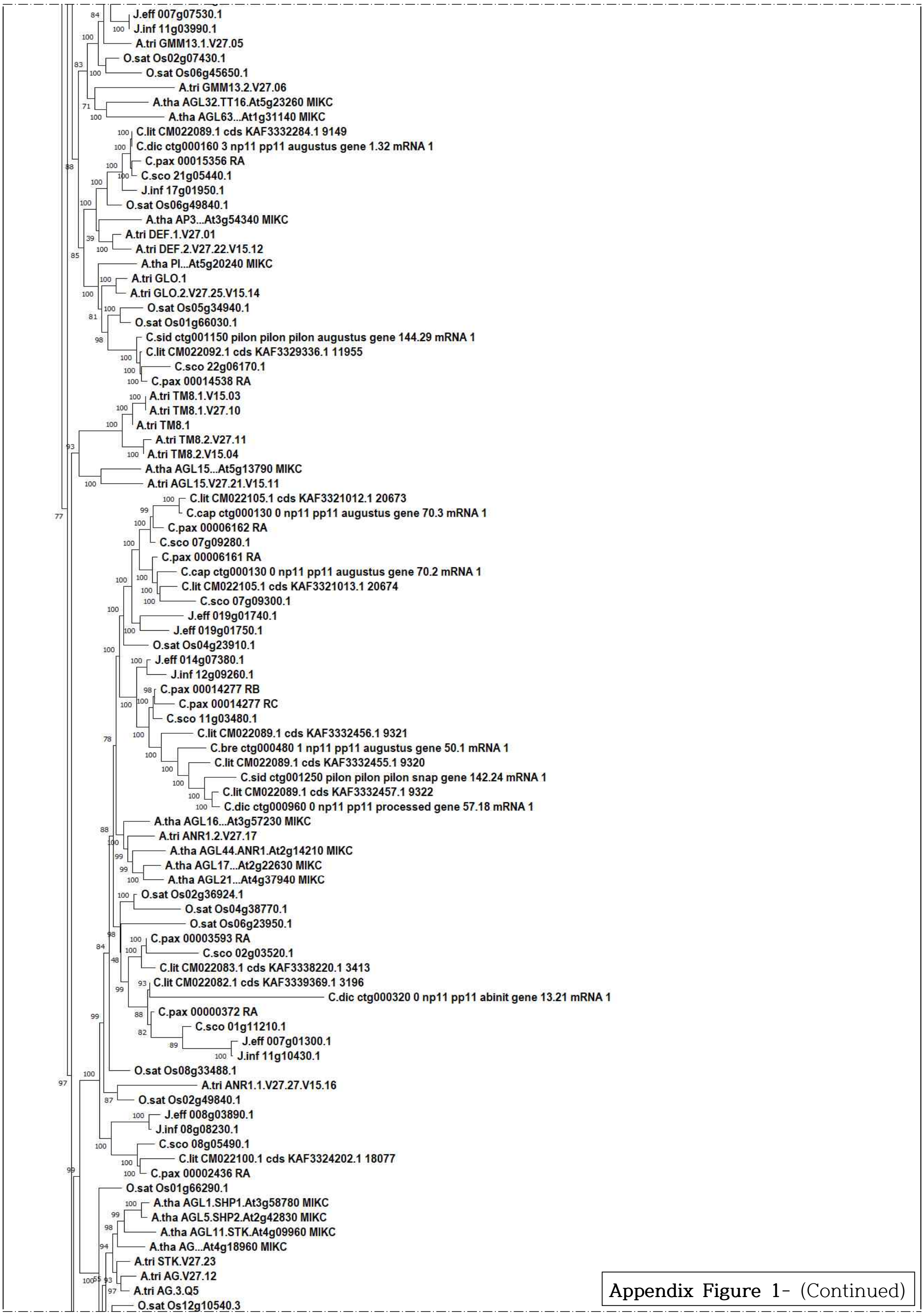
Appendix Figure 1- (Continued)



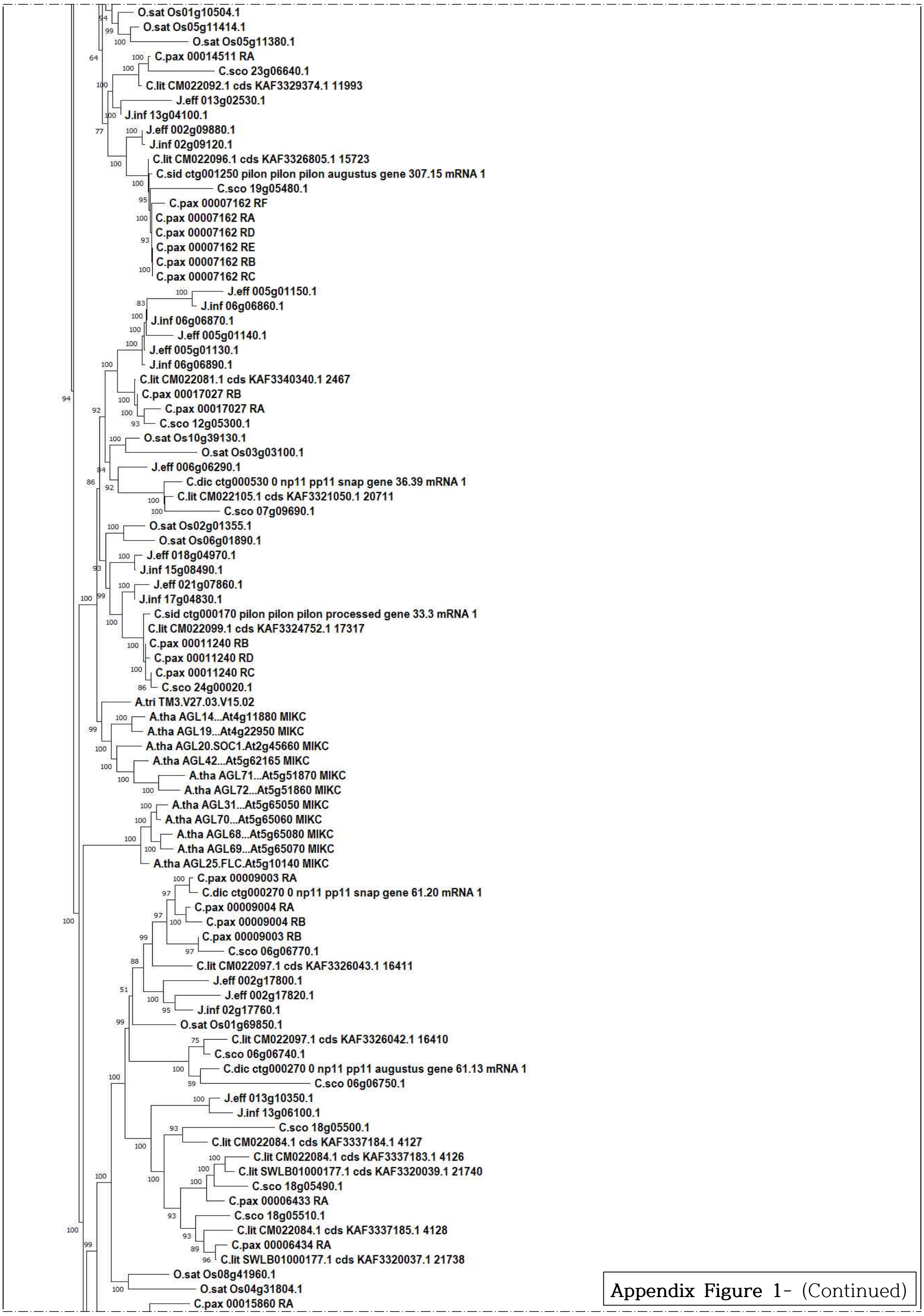
Appendix Figure 1- (Continued)



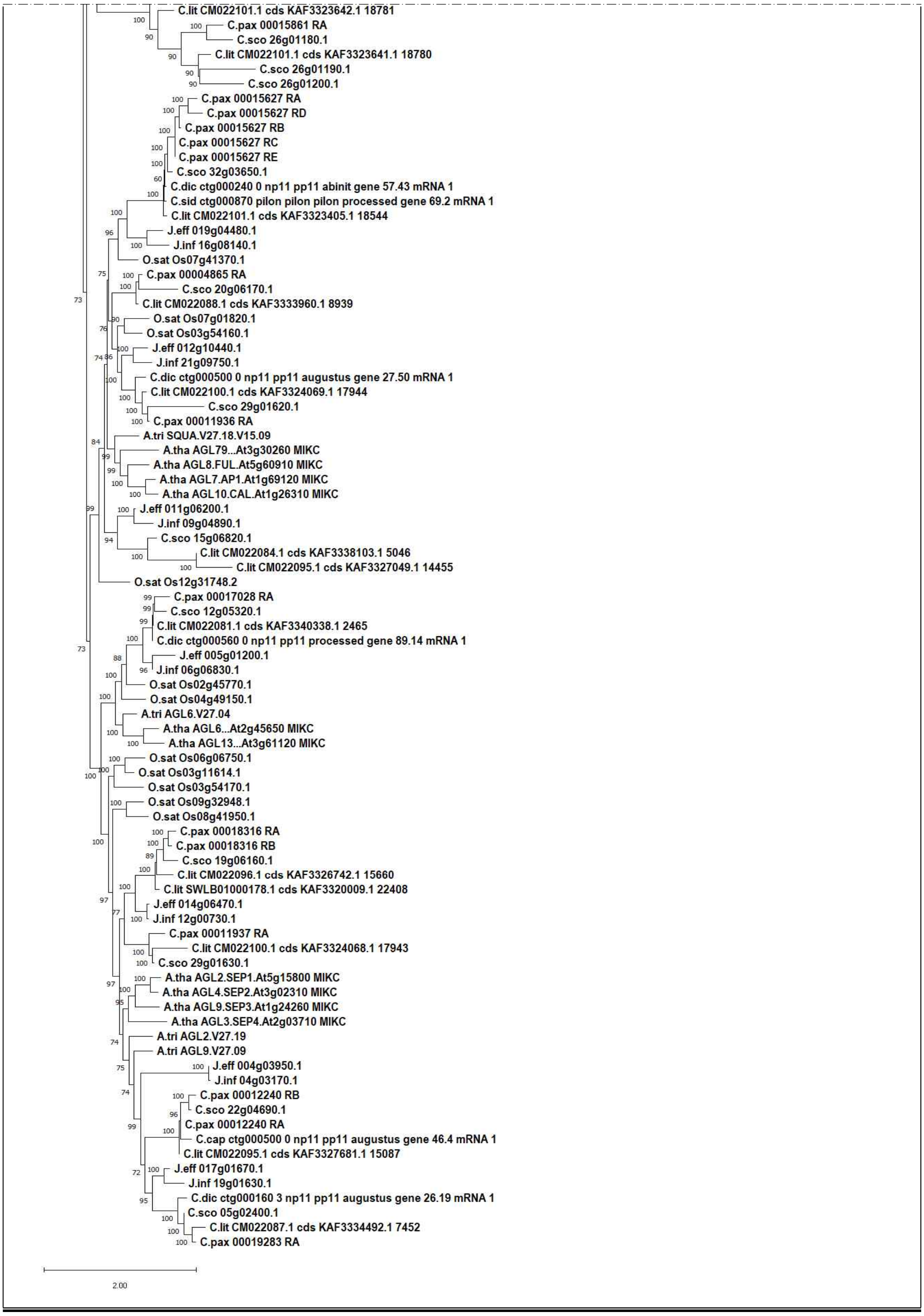
Appendix Figure 1- (Continued)



Appendix Figure 1- (Continued)



Appendix Figure 1- (Continued)


C.lit CM022101.1 cds KAF3323642.1 18781
C.pax 00015861 RA
C.sco 26g01180.1
C.lit CM022101.1 cds KAF3323641.1 18780
C.sco 26g01190.1
C.sco 26g01200.1
C.pax 00015627 RA
C.pax 00015627 RD
C.pax 00015627 RB
C.pax 00015627 RC
C.pax 00015627 RE
C.sco 32g03650.1
C.dic ctg000240 0 np11 pp11 abinit gene 57.43 mRNA 1
C.sid ctg000870 pilon pilon pilon processed gene 69.2 mRNA 1
C.lit CM022101.1 cds KAF3323405.1 18544
J.eff 019g04480.1
J.inf 16g08140.1
O.sat Os07g41370.1
C.pax 00004865 RA
C.sco 20g06170.1
C.lit CM022088.1 cds KAF3333960.1 8939
O.sat Os07g01820.1
O.sat Os03g54160.1
J.eff 012g10440.1
J.inf 21g09750.1
C.dic ctg000500 0 np11 pp11 augustus gene 27.50 mRNA 1
C.lit CM022100.1 cds KAF3324069.1 17944
C.sco 29g01620.1
C.pax 00011936 RA
A.tri SQUA.V27.18.V15.09
A.tha AGL79...At3g30260 MIKC
A.tha AGL8.FUL.At5g60910 MIKC
A.tha AGL7.AP1.At1g69120 MIKC
A.tha AGL10.CAL.At1g26310 MIKC
J.eff 011g06200.1
J.inf 09g04890.1
C.sco 15g06820.1
C.lit CM022084.1 cds KAF3338103.1 5046
C.lit CM022095.1 cds KAF3327049.1 14455
O.sat Os12g31748.2
C.pax 00017028 RA
C.sco 12g05320.1
C.lit CM022081.1 cds KAF3340338.1 2465
C.dic ctg000560 0 np11 pp11 processed gene 89.14 mRNA 1
J.eff 005g01200.1
J.inf 06g06830.1
O.sat Os02g45770.1
O.sat Os04g49150.1
A.tri AGL6.V27.04
A.tha AGL6...At2g45650 MIKC
A.tha AGL13...At3g61120 MIKC
O.sat Os06g06750.1
O.sat Os03g11614.1
O.sat Os03g54170.1
O.sat Os09g32948.1
O.sat Os08g41950.1
C.pax 00018316 RA
C.pax 00018316 RB
C.sco 19g06160.1
C.lit CM022096.1 cds KAF3326742.1 15660
C.lit SWLB01000178.1 cds KAF3320009.1 22408
J.eff 014g06470.1
J.inf 12g00730.1
C.pax 00011937 RA
C.lit CM022100.1 cds KAF3324068.1 17943
C.sco 29g01630.1
A.tha AGL2.SEP1.At5g15800 MIKC
A.tha AGL4.SEP2.At3g02310 MIKC
A.tha AGL9.SEP3.At1g24260 MIKC
A.tha AGL3.SEP4.At2g03710 MIKC
A.tri AGL2.V27.19
A.tri AGL9.V27.09
J.eff 004g03950.1
J.inf 04g03170.1
C.pax 00012240 RB
C.sco 22g04690.1
C.pax 00012240 RA
C.cap ctg000500 0 np11 pp11 augustus gene 46.4 mRNA 1
C.lit CM022095.1 cds KAF3327681.1 15087
J.eff 017g01670.1
J.inf 19g01630.1
C.dic ctg000160 3 np11 pp11 augustus gene 26.19 mRNA 1
C.sco 05g02400.1
C.lit CM022087.1 cds KAF3334492.1 7452
C.pax 00019283 RA
2.00

Appendix Table 1. List of MADS-box genes detected in this study. The sequences of five *Carex* species (*C. siderosticta, C. paxii, C. breviculmis, C. capricornis, and C. dickinsii*) are based on the BRAKER gene prediction tool with Viridiplantae data as protein evidence

Type	Subfamily name	Species	Gene name
Type Ⅱ	SQUA	*A. trichopoda*	Am.tr.SQUA.V27.18.V15.09
		A. thaliana	Ar.th.AT1G26310.1_CAL1,AGL10,CAL
			Ar.th.AT1G69120.1_AP1,AtAP1,AGL7
			Ar.th.AT3G30260.1_AGL79
			Ar.th.AT5G60910.1_FUL,AGL8
		O. sativa	Or.sa.Os07g01820.1_OsMADS15
			Or.sa.Os03g54160.1_OsMADS14
			Or.sa.Os12g31748.2_OsMADS20
			Or.sa.Os07g41370.1_OsMADS18
		J. effusus	Ju.ef.Jeff012g10440.1
			Ju.ef.Jeff011g06200.1
			Ju.ef.Jeff019g04480.1
		J. inflexus	Ju.in.Jinf21g09750.1
			Ju.in.Jinf09g04890.1
			Ju.in.Jinf16g08140.1
		C. siderosticta	Ca.si.g14054.t1
			Ca.si.g13649.t1
		C. littledalei	Ca.li.CM022100.1_KAF3324069.1_17944
			Ca.li.CM022088.1_KAF3333960.1_8939
			Ca.li.CM022084.1_KAF3338103.1_5046
			Ca.li.CM022095.1_KAF3327049.1_14455
			Ca.li.CM022101.1_KAF3323405.1_18544
		C. paxii	Ca.pa.g11318.t1
			Ca.pa.g11318.t2
			Ca.pa.g21106.t1
			Ca.pa.g8664.t1
			Ca.pa.g2773.t1
		C. scorparia	Ca.sc.Csco29g01610.1
			Ca.sc.Csco29g01620.1
			Ca.sc.Csco20g06170.1
			Ca.sc.Csco15g06820.1
			Ca.sc.Csco32g03650.1
		C. breviculmis	Ca.br.g9556.t1
			Ca.br.g2122.t1
		C. capricornis	Ca.ca.g17801.t1
			Ca.ca.g17801.t2
			Ca.ca.g14789.t1
			Ca.ca.g15887.t1
			Ca.ca.g11159.t1
		C. dickinsii	Ca.di.g8161.t1
			Ca.di.g18199.t1
			Ca.di.g14191.t1
			Ca.di.g4478.t1
	AP3/DEF	*A. trichopoda*	Am.tr.DEF.1.V27.01
			Am.tr.DEF.2.V27.22.V15.12
		A. thaliana	Ar.th.AT3G54340.1_AP3,ATAP3

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
		O. sativa	Or.sa.Os06g49840.1_OsMADS16
		J. effusus	—
		J. inflexus	Ju.in.Jinf17g01950.1
		C. siderosticta	Ca.si.g17874.t1
		C. littledalei	Ca.li.CM022089.1_KAF3332284.1_9149
		C. paxii	—
		C. scorparia	Ca.sc.Csco21g05440.1
		C. breviculmis	Ca.br.g3722.t1
		C. capricornis	Ca.ca.g16714.t1
		C. dickinsii	Ca.di.g1712.t1
	PI/GLO	*A. trichopoda*	Am.tr.GLO.1 Am.tr.GLO.2.V27.25.V15.14
		A. thaliana	Ar.th.AT5G20240.1_PI
		O. sativa	Or.sa.Os05g34940.1_OsMADS4 Or.sa.Os01g66030.1_OsMADS2
		J. effusus	—
		J. inflexus	—
		C. siderosticta	—
		C. littledalei	Ca.li.CM022092.1_KAF3329336.1_11955
		C. paxii	Ca.pa.g517.t1
		C. scorparia	Ca.sc.Csco22g06170.1
		C. breviculmis	Ca.br.g13056.t1
		C. capricornis	Ca.ca.g8792.t1
		C. dickinsii	Ca.di.g15679.t1
	GGM13	*A. trichopoda*	Am.tr.GMM13.1.V27.05 Am.tr.GMM13.2.V27.06
		A. thaliana	Ar.th.AT1G31140.2_GOA,AGL63 Ar.th.AT5G23260.4_TT16,ABS,AGL32
		O. sativa	Or.sa.Os02g07430.1_OsMADS29 Or.sa.Os06g45650.1_OsMADS30 Or.sa.Os04g52410.2_OsMADS31
		J. effusus	Ju.ef.Jeff002g14390.1 Ju.ef.Jeff007g07530.1
		J. inflexus	Ju.in.Jinf02g14250.1 Ju.in.Jinf11g03990.1
		C. siderosticta	—
		C. littledalei	Ca.li.CM022097.1_KAF3325733.1_16101
		C. paxii	—
		C. scorparia	Ca.sc.Csco30g06190.1
		C. breviculmis	Ca.br.g7223.t1
		C. capricornis	Ca.ca.g12586.t1
		C. dickinsii	—
	AG	*A. trichopoda*	Am.tr.AG.V27.12 Am.tr.AG.3.Q5 Am.tr.STK.V27.23
		A. thaliana	Ar.th.AT2G42830.2_AGL5,SHP2 Ar.th.AT3G58780.4_SHP1,AGL1 Ar.th.AT4G09960.4_STK,AGL11 Ar.th.AT4G18960.1_AG
		O. sativa	Or.sa.Os05g11414.1_OsMADS58 Or.sa.Os05g11380.1_OsMADS66 Or.sa.Os01g10504.1_OsMADS3 Or.sa.Os01g66290.1_OsMADS21 Or.sa.Os12g10540.3_OsMADS13

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
		J. effusus	Ju.ef.Jeff002g09880.1
			Ju.ef.Jeff013g02530.1
		J. inflexus	Ju.in.Jinf02g09120.1
			Ju.in.Jinf13g04100.1
		C. siderosticta	—
		C. littledalei	Ca.li.CM022096.1_KAF3326805.1_15723
			Ca.li.CM022092.1_KAF3329374.1_11993
		C. paxii	Ca.pa.g17856.t1
		C. scorparia	Ca.sc.Csco19g05480.1
			Ca.sc.Csco23g06640.1
			Ca.sc.Csco23g06630.1
		C. breviculmis	Ca.br.g1508.t1
		C. capricornis	Ca.ca.g12106.t1
			Ca.ca.g8823.t1
		C. dickinsii	Ca.di.g12716.t1
			Ca.di.g15710.t1
	SEP	*A. trichopoda*	Am.tr.AGL2.V27.19
			Am.tr.AGL9.V27.09
		A. thaliana	Ar.th.AT5G15800.2_AGL2,SEP1
			Ar.th.AT3G02310.1_AGL4,SEP2
			Ar.th.AT1G24260.2_AGL9,SEP3
			Ar.th.AT2G03710.1_SEP4,AGL3
		O. sativa	Or.sa.Os09g32948.1_OsMADS8
			Or.sa.Os08g41950.1_OsMADS7
			Or.sa.Os06g06750.1_OsMADS5
			Or.sa.Os03g11614.1_OsMADS1
			Or.sa.Os03g54170.1_OsMADS34
		J. effusus	Ju.ef.Jeff014g06470.1
			Ju.ef.Jeff017g01670.1
			Ju.ef.Jeff004g03940.1
			Ju.ef.Jeff004g03950.1
		J. inflexus	Ju.in.Jinf12g00730.1
			Ju.in.Jinf19g01630.1
			Ju.in.Jinf04g03180.1
			Ju.in.Jinf04g03170.1
		C. siderosticta	Ca.si.g18466.t1
			Ca.si.g2760.t1
			Ca.si.g2760.t2
			Ca.si.g9478.t1
			Ca.si.g9478.t2
		C. littledalei	Ca.li.CM022096.1_KAF3326742.1_15660
			Ca.li.SWLB01000178.1_KAF3320009.1_22408
			Ca.li.CM022087.1_KAF3334492.1_7452
			Ca.li.CM022100.1_KAF3324068.1_17943
			Ca.li.CM022095.1_KAF3327681.1_15087
		C. paxii	Ca.pa.g19278.t1
			Ca.pa.g2631.t1
			Ca.pa.g11319.t1
		C. scorparia	Ca.sc.Csco19g06160.1
			Ca.sc.Csco05g02400.1
			Ca.sc.Csco29g01630.1
			Ca.sc.Csco22g04690.1
		C. breviculmis	Ca.br.g1552.t1
			Ca.br.g9557.t1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
		C. capricornis	Ca.ca.g12053.t1
			Ca.ca.g9908.t1
			Ca.ca.g9908.t2
			Ca.ca.g17802.t1
			Ca.ca.g17432.t1
		C. dickinsii	Ca.di.g12772.t1
			Ca.di.g12772.t2
			Ca.di.g1850.t1
			Ca.di.g1850.t2
			Ca.di.g8160.t1
			Ca.di.g9156.t1
	AGL6	*A. trichopoda*	Am.tr.AGL6.V27.04
		A. thaliana	Ar.th.AT2G45650.1_AGL6,RSB1
			Ar.th.AT3G61120.1_AGL13
		O. sativa	Or.sa.Os04g49150.1_OsMADS17
			Or.sa.Os02g45770.1_OsMADS6
		J. effusus	Ju.ef.Jeff005g01200.1
		J. inflexus	Ju.in.Jinf06g06830.1
		C. siderosticta	—
		C. littledalei	Ca.li.CM022081.1_KAF3340338.1_2465
		C. paxii	Ca.pa.g12910.t1
		C. scorparia	Ca.sc.Csco12g05320.1
		C. breviculmis	Ca.br.g11144.t1
		C. capricornis	Ca.ca.g2268.t1
		C. dickinsii	Ca.di.g11120.t1
	StMADS11	*A. trichopoda*	Am.tr.StM11.V27.29.V15.18
		A. thaliana	Ar.th.AT2G22540.1_AGL22,FAQ1,SVP
			Ar.th.AT4G24540.1_AGL24
		O. sativa	Or.sa.Os06g11330.1_OsMADS55
			Or.sa.Os02g52340.1_OsMADS22
			Or.sa.Os03g08754.1_OsMADS47
		J. effusus	Ju.ef.Jeff012g10430.1
			Ju.ef.Jeff013g02540.1
			Ju.ef.Jeff014g05250.1
			Ju.ef.Jeff001g00470.1
		J. inflexus	Ju.in.Jinf13g03490.1
			Ju.in.Jinf21g09740.1
			Ju.in.Jinf12g01960.1
			Ju.in.Jinf01g16830.1
		C. siderosticta	Ca.si.g4160.t1
			Ca.si.g16601.t1
			Ca.si.g10929.t1
			Ca.si.g10929.t2
			Ca.si.g11359.t1
		C. littledalei	Ca.li.CM022084.1_KAF3337639.1_4582
			Ca.li.CM022084.1_KAF3337519.1_4462
			Ca.li.CM022096.1_KAF3326492.1_15410
			Ca.li.CM022084.1_KAF3337515.1_4458
			Ca.li.CM022084.1_KAF3337520.1_4463
			Ca.li.CM022106.1_KAF3320700.1_21242
			Ca.li.CM022083.1_KAF3338376.1_3569
		C. paxii	Ca.pa.g5207.t1
			Ca.pa.g12103.t1
			Ca.pa.g15151.t2

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.pa.g15151.t1
		C. scorparia	Ca.sc.Csco27g02230.1
			Ca.sc.Csco17g07490.1
			Ca.sc.Csco30g01500.1
			Ca.sc.Csco30g01510.1
			Ca.sc.Csco17g07480.1
			Ca.sc.Csco25g01190.1
			Ca.sc.Csco10g06580.1
			Ca.sc.Csco02g01670.1
		C. breviculmis	Ca.br.g5797.t1
			Ca.br.g11780.t1
		C. capricornis	Ca.ca.g2883.t1
			Ca.ca.g14235.t1
			Ca.ca.g10486.t1
		C. dickinsii	Ca.di.g3674.t1
			Ca.di.g6436.t1
			Ca.di.g9359.t1
			Ca.di.g15072.t1
	OsMADS32	*A. trichopoda*	Am.tr.OsM32.V27.02.V15.01
		A. thaliana	—
		O. sativa	Or.sa.Os01g52680.1_OsMADS32
		J. effusus	Ju.ef.Jeff001g12960.1
		J. inflexus	Ju.in.Jinf01g07410.1
		C. siderosticta	—
		C. littledalei	Ca.li.CM022094.1_KAF3328473.1_14367
		C. paxii	Ca.pa.g19800.t1
		C. scorparia	Ca.sc.Csco33g02230.1
		C. breviculmis	Ca.br.g14618.t1
		C. capricornis	Ca.ca.g16029.t1
		C. dickinsii	—
	AGL12	*A. trichopoda*	Am.tr.AGL12.V27.24
			Am.tr.AGL12.V15.13
		A. thaliana	Ar.th.AT1G71692.1_XAL1,AGL12
		O. sativa	Or.sa.Os12g10520.1_OsMADS33
			Or.sa.Os08g02070.1_OsMADS26
		J. effusus	Ju.ef.Jeff011g03240.1
			Ju.ef.Jeff041g00010.1
			Ju.ef.Jeff004g01000.1
		J. inflexus	Ju.in.Jinf09g08000.1
			Ju.in.Jinf13g03480.1
			Ju.in.Jinf04g05270.1
		C. siderosticta	Ca.si.g14305.t1
			Ca.si.g15770.t1
			Ca.si.g442.t1
		C. littledalei	Ca.li.CM022084.1_KAF3337820.1_4763
			Ca.li.CM022092.1_KAF3329371.1_11990
			Ca.li.CM022080.1_KAF3340774.1_1075
		C. paxii	Ca.pa.g494.t1
			Ca.pa.g1733.t1
		C. scorparia	Ca.sc.Csco23g06660.1
			Ca.sc.Csco03g05790.1
		C. breviculmis	Ca.br.g13034.t1
			Ca.br.g3099.t1
		C. capricornis	Ca.ca.g8821.t1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.ca.g10202.t1
		C. dickinsii	Ca.di.g15707.t1
	AGL15	*A. trichopoda*	Am.tr.AGL15.V27.21.V15.11
		A. thaliana	Ar.th.AT5G13790.2_AGL15
			Ar.th.AT3G57390.1_AGL18
		O. sativa	—
		J. effusus	—
		J. inflexus	—
		C. siderosticta	—
		C. littledalei	—
		C. paxii	—
		C. scorparia	Ca.sc.Csco19g05470.1
		C. breviculmis	—
		C. capricornis	—
		C. dickinsii	—
	AGL17	*A. trichopoda*	Am.tr.ANR1.1.V27.27.V15.16
			Am.tr.ANR1.2.V27.17
		A. thaliana	Ar.th.AT2G14210.2_ANR1,AtANR1,AGL44
			Ar.th.AT2G22630.1_AGL17
			Ar.th.AT4G37940.1_AGL21
			Ar.th.AT3G57230.1_AGL16
		O. sativa	Or.sa.Os08g33488.1_OsMADS23
			Or.sa.Os02g49840.1_OsMADS57
			Or.sa.Os06g23950.1_OsMADS59
			Or.sa.Os04g38770.1_OsMADS61
			Or.sa.Os02g36924.1_OsMADS27
			Or.sa.Os04g23910.1_OsMADS25
		J. effusus	Ju.ef.Jeff016g07270.1
			Ju.ef.Jeff007g01300.1
			Ju.ef.Jeff014g07380.1
			Ju.ef.Jeff019g01740.1
			Ju.ef.Jeff019g01750.1
			Ju.ef.Jeff008g03890.1
		J. inflexus	Ju.in.Jinf11g10430.1
			Ju.in.Jinf12g09260.1
			Ju.in.Jinf08g08230.1
		C. siderosticta	Ca.si.g10426.t1
		C. littledalei	Ca.li.CM022092.1_KAF3330189.1_12808
			Ca.li.CM022083.1_KAF3338220.1_3413
			Ca.li.CM022082.1_KAF3339369.1_3196
			Ca.li.CM022089.1_KAF3332455.1_9320
			Ca.li.CM022089.1_KAF3332456.1_9321
			Ca.li.CM022089.1_KAF3332457.1_9322
			Ca.li.CM022083.1_KAF3338430.1_3623
			Ca.li.CM022105.1_KAF3321012.1_20673
			Ca.li.CM022105.1_KAF3321013.1_20674
			Ca.li.CM022100.1_KAF3324202.1_18077
		C. paxii	Ca.pa.g14966.t1
			Ca.pa.g14966.t2
			Ca.pa.g3914.t1
			Ca.pa.g11978.t1
			Ca.pa.g1705.t1
			Ca.pa.g1706.t1
		C. scorparia	Ca.sc.Csco08g05480.1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.sc.Csco02g03520.1
			Ca.sc.Csco01g11210.1
			Ca.sc.Csco20g06180.1
			Ca.sc.Csco06g06750.1
			Ca.sc.Csco11g03480.1
			Ca.sc.Csco07g09300.1
			Ca.sc.Csco07g09280.1
			Ca.sc.Csco08g05490.1
		C. breviculmis	Ca.br.g3852.t1
			Ca.br.g9427.t1
		C. capricornis	Ca.ca.g10395.t1
			Ca.ca.g11559.t1
			Ca.ca.g1940.t1
			Ca.ca.g6615.t1
		C. dickinsii	Ca.di.g15203.t1
			Ca.di.g17772.t1
			Ca.di.g16937.t1
			Ca.di.g9713.t1
			Ca.di.g6018.t1
	TM3	*A. trichopoda*	Am.tr.TM3.V27.03.V15.02
		A. thaliana	Ar.th.AT2G45660.1_AGL20,SOC1,ATSOC1
			Ar.th.AT5G51860.1_AGL72
			Ar.th.AT5G51870.3_AGL71
			Ar.th.AT5G62165.1_AGL42,FYF
			Ar.th.AT4G11880.1_AGL14,XAL2
			Ar.th.AT4G22950.1_GL19,AGL19
		O. sativa	Or.sa.Os10g39130.1_OsMADS56
			Or.sa.Os03g03100.1_OsMADS50
		J. effusus	Ju.ef.Jeff007g03390.1
			Ju.ef.Jeff005g01140.1
			Ju.ef.Jeff005g01150.1
			Ju.ef.Jeff005g01170.1
			Ju.ef.Jeff005g01130.1
			Ju.ef.Jeff021g07860.1
			Ju.ef.Jeff018g04970.1
			Ju.ef.Jeff006g06290.1
		J. inflexus	Ju.in.Jinf06g06850.1
			Ju.in.Jinf06g06860.1
			Ju.in.Jinf06g06870.1
			Ju.in.Jinf06g06890.1
			Ju.in.Jinf17g04830.1
			Ju.in.Jinf15g08490.1
		C. siderosticta	Ca.si.g12342.t1
			Ca.si.g1279.t1
		C. littledalei	Ca.li.CM022081.1_KAF3340340.1_2467
			Ca.li.CM022099.1_KAF3324752.1_17317
			Ca.li.CM022105.1_KAF3321050.1_20711
		C. paxii	Ca.pa.g1121.t1
		C. scorparia	Ca.sc.Csco12g05300.1
			Ca.sc.Csco24g00020.1
			Ca.sc.Csco07g09690.1
		C. breviculmis	—
		C. capricornis	Ca.ca.g2270.t1
			Ca.ca.g13329.t1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.ca.g1900.t1
		C. dickinsii	Ca.di.g11121.t1
			Ca.di.g14087.t1
	TM8	*A. trichopoda*	Am.tr.TM8.1
			Am.tr.TM8.1.V15.03
			Am.tr.TM8.1.V27.10
			Am.tr.TM8.2.V27.11
			Am.tr.TM8.2.V15.04
		A. thaliana	—
		O. sativa	—
		J. effusus	—
		J. inflexus	—
		C. siderosticta	—
		C. littledalei	—
		C. paxii	—
		C. scorparia	—
		C. breviculmis	—
		C. capricornis	—
		C. dickinsii	—
	FLC	*A. trichopoda*	Ar.th.AT1G77080.8_FLM,AGL27,MAF1
			Ar.th.AT5G65050.3_MAF2,AGL31
			Ar.th.AT5G65060.1_AGL70,FCL3,MAF3
			Ar.th.AT5G65070.3_MAF4,FCL4,AGL69
			Ar.th.AT5G65080.2_MAF5,AGL68
			Ar.th.AT5G10140.1_FLC,RSB6,AGL25,FLF
		A. thaliana	—
		O. sativa	—
		J. effusus	—
		J. inflexus	—
		C. siderosticta	—
		C. littledalei	—
		C. paxii	—
		C. scorparia	—
		C. breviculmis	—
		C. capricornis	—
		C. dickinsii	—
Type I	Only monocots	*A. trichopoda*	—
		A. thaliana	—
		O. sativa	Or.sa.Os01g69850.1_OsMADS65
			Or.sa.Os08g41960.1_OsMADS37
			Or.sa.Os04g31804.1_OsMADS64
		J. effusus	Ju.ef.Jeff013g10350.1
			Ju.ef.Jeff002g17800.1
			Ju.ef.Jeff002g17820.1
		J. inflexus	Ju.in.Jinf13g06100.1
			Ju.in.Jinf02g17760.1
		C. siderosticta	Ca.si.g5132.t1
			Ca.si.g13843.t1
			Ca.si.g13844.t1
		C. littledalei	Ca.li.SWLB01000177.1_KAF3320038.1_21739
			Ca.li.CM022084.1_KAF3337184.1_4127
			Ca.li.CM022084.1_KAF3337183.1_4126
			Ca.li.SWLB01000177.1_KAF3320039.1_21740
			Ca.li.CM022084.1_KAF3337185.1_4128

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.li.SWLB01000177.1_KAF3320037.1_21738
			Ca.li.CM022097.1_KAF3326042.1_16410
			Ca.li.CM022097.1_KAF3326043.1_16411
			Ca.li.CM022101.1_KAF3323641.1_18780
			Ca.li.CM022101.1_KAF3323642.1_18781
		C. paxii	Ca.pa.g2005.t1
			Ca.pa.g2006.t1
			Ca.pa.g2006.t2
			Ca.pa.g7345.t1
			Ca.pa.g7346.t1
			Ca.pa.g7347.t1
			Ca.pa.g3026.t1
			Ca.pa.g3025.t1
		C. scorparia	Ca.sc.Csco18g05500.1
			Ca.sc.Csco18g05510.1
			Ca.sc.Csco18g05490.1
			Ca.sc.Csco06g06740.1
			Ca.sc.Csco06g06770.1
			Ca.sc.Csco26g01200.1
			Ca.sc.Csco26g01190.1
			Ca.sc.Csco26g01180.1
		C. breviculmis	Ca.br.g4775.t1
		C. capricornis	Ca.ca.g9454.t1
			Ca.ca.g3582.t1
			Ca.ca.g3583.t1
		C. dickinsii	Ca.di.g10667.t1
			Ca.di.g6153.t1
			Ca.di.g6157.t1
			Ca.di.g4265.t1
			Ca.di.g4266.t1
	Other	*A. trichopoda*	Am.tr.MP.V27.08
			Am.tr.MS.V27.07
			Am.tr.MA1.V27.30
			Am.tr.MA2.V27.31.V15.19
			Am.tr.MA3.V27.33.V15.21
			Am.tr.MA4.V27.32.V15.20
			Am.tr.MA5.V27.20.V15.10
			Am.tr.MA6.V27.16.V15.08
			Am.tr.MB1.V27.14
			Am.tr.MB1.V15.06
			Am.tr.MB2.V27.15.V15.07
			Am.tr.MB3.V27.34
			Am.tr.MB4.V27.13.V15.05
			Am.tr.MB5.V27.28.V15.17
			Am.tr.MC.V27.26.V15.15
		A. thaliana	Ar.th.AT1G01530.1_AGL28
			Ar.th.AT1G17310.2_AGL100
			Ar.th.AT1G18750.3_AGL65
			Ar.th.AT1G22130.1_AGL104
			Ar.th.AT1G22590.2_AGL87
			Ar.th.AT1G28450.1_AGL58
			Ar.th.AT1G28460.1_AGL59
			Ar.th.AT1G31630.1_AGL86
			Ar.th.AT1G31640.1_AGL92

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ar.th.AT1G46408.1_AGL97
			Ar.th.AT1G47760.1_AGL102
			Ar.th.AT1G48150.1_AGL74
			Ar.th.AT1G54760.1_AGL85
			Ar.th.AT1G59810.1_AGL50
			Ar.th.AT1G60040.1_AGL49
			Ar.th.AT1G60880.1_AGL56
			Ar.th.AT1G60920.1_AGL55
			Ar.th.AT1G65300.1_PHE2,AGL38
			Ar.th.AT1G65330.1_PHE1,AGL37
			Ar.th.AT1G65360.1_AGL23
			Ar.th.AT1G69540.2_AGL94
			Ar.th.AT1G72350.1_AGL60
			Ar.th.AT1G77950.1_AGL67
			Ar.th.AT1G77980.1_AGL66
			Ar.th.AT2G03060.2_AGL30
			Ar.th.AT2G15660.1_AGL95
			Ar.th.AT2G24840.1_DIA,AGL61
			Ar.th.AT2G26320.1_AGL33
			Ar.th.AT2G28700.1_AGL46
			Ar.th.AT2G34440.1_AGL29
			Ar.th.AT2G40210.1_AGL48
			Ar.th.AT3G04100.1_AGL57
			Ar.th.AT3G05860.1_AGL45
			Ar.th.AT3G18650.1_AGL103
			Ar.th.AT3G66656.1_AGL91
			Ar.th.AT4G02235.1_AGL51
			Ar.th.AT4G11250.1_AGL52
			Ar.th.AT4G36590.1_AGL40
			Ar.th.AT5G04640.1_AGL99
			Ar.th.AT5G06500.1_AGL96
			Ar.th.AT5G26580.1_AGL34
			Ar.th.AT5G26630.1_AGL35
			Ar.th.AT5G26650.1_AGL36
			Ar.th.AT5G26950.1_AGL93
			Ar.th.AT5G27050.1_AGL101
			Ar.th.AT5G27070.1_AGL53
			Ar.th.AT5G27090.1_AGL54
			Ar.th.AT5G27130.1_AGL39
			Ar.th.AT5G27580.1_AGL89
			Ar.th.AT5G27960.1_AGL90
			Ar.th.AT5G38620.1_AGL73
			Ar.th.AT5G38740.1_AGL77
			Ar.th.AT5G39750.1_EMB3008,AGL81
			Ar.th.AT5G39810.1_AGL98
			Ar.th.AT5G40120.1_AGL76
			Ar.th.AT5G40220.1_AGL43
			Ar.th.AT5G41200.1_AGL75
			Ar.th.AT5G48670.1_FEM111,AGL80
			Ar.th.AT5G49420.1_AGL84
			Ar.th.AT5G49490.1_AGL83
			Ar.th.AT5G55690.1_AGL47
			Ar.th.AT5G58890.1_AGL82
			Ar.th.AT5G60440.1_AGL62

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ar.th.AT5G65330.1_AGL78
		O. sativa	Or.sa.Os01g11510.1_OsMADS91
			Or.sa.Os01g18420.1_OsMADS88
			Or.sa.Os01g18440.1_OsMADS89
			Or.sa.Os01g23750.1_OsMADS92
			Or.sa.Os01g23760.1_OsMADS93
			Or.sa.Os01g23770.1_OsMADS94
			Or.sa.Os01g23780.1_OsMADS95
			Or.sa.Os01g67890.1_OsMADS96
			Or.sa.Os01g68420.1_OsMADS97
			Or.sa.Os01g68560.1_OsMADS98
			Or.sa.Os01g74440.1_OsMADS79
			Or.sa.Os02g01355.1_putative
			Or.sa.Os02g06860.1_OsMADS80
			Or.sa.Os02g07430.1_OsMADS29
			Or.sa.Os03g14850.1_OsMADS72
			Or.sa.Os03g37670.1_OsMADS86
			Or.sa.Os03g38610.1_OsMADS87
			Or.sa.Os04g24790.1_OsMADS81
			Or.sa.Os04g24800.1_OsMADS82
			Or.sa.Os04g24810.1_OsMADS83
			Or.sa.Os04g25870.1_OsMADS84
			Or.sa.Os04g25920.1_OsMADS85
			Or.sa.Os04g25930.1_OsMADS99
			Or.sa.Os05g23780.1_OsMADS70
			Or.sa.Os06g01890.1_putative
			Or.sa.Os06g11970.1_OsMADS63
			Or.sa.Os06g22760.1_OsMADS71
			Or.sa.Os06g30810.1_OsMADS75
			Or.sa.Os06g30830.1_OsMADS76
			Or.sa.Os07g04170.1_OsMADS90
			Or.sa.Os08g38590.1_OsMADS62
			Or.sa.Os09g02780.1_OsMADS77
			Or.sa.Os09g02830.1_OsMADS78
			Or.sa.Os11g43740.1_OsMADS68
			Or.sa.Os12g21850.1_OsMADS73
			Or.sa.Os12g21880.1_OsMADS74
		J. effusus	Ju.ef.Jeff009g08380.1
			Ju.ef.Jeff010g02340.1
			Ju.ef.Jeff011g09370.1
			Ju.ef.Jeff011g09400.1
			Ju.ef.Jeff012g04950.1
			Ju.ef.Jeff012g04960.1
			Ju.ef.Jeff012g06510.1
			Ju.ef.Jeff012g09240.1
			Ju.ef.Jeff013g00980.1
			Ju.ef.Jeff013g01530.1
			Ju.ef.Jeff013g02860.1
			Ju.ef.Jeff013g02930.1
			Ju.ef.Jeff013g03110.1
			Ju.ef.Jeff014g08840.1
			Ju.ef.Jeff015g02660.1
			Ju.ef.Jeff015g02680.1
			Ju.ef.Jeff015g02700.1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ju.ef.Jeff016g09640.1
			Ju.ef.Jeff017g03810.1
			Ju.ef.Jeff017g03860.1
			Ju.ef.Jeff017g03870.1
			Ju.ef.Jeff017g04000.1
			Ju.ef.Jeff017g06620.1
			Ju.ef.Jeff017g06630.1
			Ju.ef.Jeff017g06640.1
			Ju.ef.Jeff017g06680.1
			Ju.ef.Jeff017g06770.1
			Ju.ef.Jeff017g07680.1
			Ju.ef.Jeff017g07980.1
			Ju.ef.Jeff019g04000.1
			Ju.ef.Jeff021g05170.1
			Ju.ef.Jeff021g05180.1
			Ju.ef.Jeff021g06620.1
			Ju.ef.Jeff110g00070.1
		J. inflexus	Ju.in.Jinf01g09380.1
			Ju.in.Jinf01g15670.1
			Ju.in.Jinf01g15680.1
			Ju.in.Jinf03g01720.1
			Ju.in.Jinf03g03200.1
			Ju.in.Jinf03g14380.1
			Ju.in.Jinf04g00510.1
			Ju.in.Jinf04g03250.1
			Ju.in.Jinf04g04360.1
			Ju.in.Jinf05g16860.1
			Ju.in.Jinf08g05060.1
			Ju.in.Jinf09g06720.1
			Ju.in.Jinf10g00990.1
			Ju.in.Jinf11g05050.1
			Ju.in.Jinf11g10420.1
			Ju.in.Jinf12g07650.1
			Ju.in.Jinf13g03000.1
			Ju.in.Jinf13g03020.1
			Ju.in.Jinf13g03320.1
			Ju.in.Jinf14g09750.1
			Ju.in.Jinf16g05770.1
			Ju.in.Jinf17g03600.1
			Ju.in.Jinf18g02110.1
			Ju.in.Jinf18g02130.1
			Ju.in.Jinf19g03560.1
			Ju.in.Jinf19g03580.1
			Ju.in.Jinf19g03600.1
			Ju.in.Jinf19g03650.1
			Ju.in.Jinf19g03680.1
			Ju.in.Jinf19g03710.1
			Ju.in.Jinf19g03820.1
			Ju.in.Jinf19g06410.1
			Ju.in.Jinf19g06420.1
			Ju.in.Jinf19g06430.1
			Ju.in.Jinf19g06950.1
			Ju.in.Jinf19g07310.1
			Ju.in.Jinf19g07640.1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ju.in.Jinf21g06180.1
			Ju.in.Jinf21g06190.1
			Ju.in.Jinf21g08610.1
		C. siderosticta	Ca.si.g10117.t1
			Ca.si.g10344.t1
			Ca.si.g10586.t1
			Ca.si.g10587.t1
			Ca.si.g10588.t1
			Ca.si.g10589.t1
			Ca.si.g10590.t1
			Ca.si.g10591.t1
			Ca.si.g10591.t2
			Ca.si.g10592.t1
			Ca.si.g11161.t1
			Ca.si.g1310.t1
			Ca.si.g14351.t1
			Ca.si.g14354.t1
			Ca.si.g15263.t1
			Ca.si.g16012.t1
			Ca.si.g16013.t1
			Ca.si.g16014.t1
			Ca.si.g16015.t1
			Ca.si.g16263.t1
			Ca.si.g16300.t1
			Ca.si.g16424.t1
			Ca.si.g1746.t1
			Ca.si.g17600.t1
			Ca.si.g2556.t1
			Ca.si.g600.t1
			Ca.si.g616.t1
			Ca.si.g6838.t1
			Ca.si.g9258.t1
			Ca.si.g9259.t1
			Ca.si.g9418.t1
			Ca.si.g9600.t1
			Ca.si.g9601.t1
		C. littledalei	Ca.li.CM022085.1_KAF3336528.1_5470
			Ca.li.CM022085.1_KAF3336591.1_5533
			Ca.li.CM022087.1_KAF3334235.1_7195
			Ca.li.CM022087.1_KAF3334432.1_7392
			Ca.li.CM022088.1_KAF3333390.1_8369
			Ca.li.CM022088.1_KAF3333762.1_8741
			Ca.li.CM022090.1_KAF3331989.1_10683
			Ca.li.CM022090.1_KAF3332121.1_10815
			Ca.li.CM022090.1_KAF3332123.1_10817
			Ca.li.CM022090.1_KAF3332124.1_10818
			Ca.li.CM022091.1_KAF3331013.1_11655
			Ca.li.CM022092.1_KAF3330238.1_12857
			Ca.li.CM022093.1_KAF3329249.1_13615
			Ca.li.CM022094.1_KAF3327899.1_13793
			Ca.li.CM022099.1_KAF3324789.1_17354
			Ca.li.CM022100.1_KAF3323921.1_17796
			Ca.li.CM022100.1_KAF3323922.1_17797
			Ca.li.CM022100.1_KAF3324293.1_18168

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.li.CM022100.1_KAF3324294.1_18169
			Ca.li.CM022102.1_KAF3322960.1_19357
			Ca.li.CM022102.1_KAF3322975.1_19372
			Ca.li.CM022103.1_KAF3322558.1_20107
			Ca.li.CM022106.1_KAF3320894.1_21436
			Ca.li.CM022106.1_KAF3320926.1_21468
			Ca.li.SWLB01000076.1_KAF3320272.1_22128
			Ca.li.SWLB01000118.1_KAF3320180.1_22218
			Ca.li.SWLB01000138.1_KAF3320131.1_22268
			Ca.li.SWLB01000214.1_KAF3319710.1_21774
		C. paxii	Ca.pa.g1152.t1
			Ca.pa.g12908.t1
			Ca.pa.g15899.t1
			Ca.pa.g16136.t1
			Ca.pa.g16137.t1
			Ca.pa.g17457.t1
			Ca.pa.g17803.t1
			Ca.pa.g17813.t1
			Ca.pa.g18493.t1
			Ca.pa.g18495.t1
			Ca.pa.g18613.t1
			Ca.pa.g2313.t1
			Ca.pa.g288.t1
			Ca.pa.g3929.t1
			Ca.pa.g4518.t1
			Ca.pa.g4520.t1
			Ca.pa.g7165.t1
			Ca.pa.g8862.t1
		C. scorparia	Ca.sc.Csco01g02820.1
			Ca.sc.Csco01g02840.1
			Ca.sc.Csco01g02850.1
			Ca.sc.Csco01g07000.1
			Ca.sc.Csco01g11230.1
			Ca.sc.Csco01g11370.1
			Ca.sc.Csco02g03530.1
			Ca.sc.Csco02g06970.1
			Ca.sc.Csco02g06980.1
			Ca.sc.Csco02g06990.1
			Ca.sc.Csco02g09620.1
			Ca.sc.Csco03g06760.1
			Ca.sc.Csco03g07130.1
			Ca.sc.Csco03g07140.1
			Ca.sc.Csco03g07150.1
			Ca.sc.Csco03g07160.1
			Ca.sc.Csco03g07180.1
			Ca.sc.Csco03g08560.1
			Ca.sc.Csco05g05620.1
			Ca.sc.Csco06g04810.1
			Ca.sc.Csco09g06410.1
			Ca.sc.Csco09g06430.1
			Ca.sc.Csco10g04300.1
			Ca.sc.Csco10g04460.1
			Ca.sc.Csco11g07360.1
			Ca.sc.Csco11g07370.1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.sc.Csco12g07980.1
			Ca.sc.Csco13g02030.1
			Ca.sc.Csco13g02060.1
			Ca.sc.Csco13g02090.1
			Ca.sc.Csco13g05110.1
			Ca.sc.Csco13g05120.1
			Ca.sc.Csco17g06000.1
			Ca.sc.Csco18g03520.1
			Ca.sc.Csco18g06250.1
			Ca.sc.Csco19g06150.1
			Ca.sc.Csco20g04540.1
			Ca.sc.Csco21g06130.1
			Ca.sc.Csco22g05440.1
			Ca.sc.Csco22g07530.1
			Ca.sc.Csco23g00660.1
			Ca.sc.Csco23g05730.1
			Ca.sc.Csco24g00360.1
			Ca.sc.Csco24g01600.1
			Ca.sc.Csco24g04210.1
			Ca.sc.Csco25g01210.1
			Ca.sc.Csco25g01220.1
			Ca.sc.Csco25g01230.1
			Ca.sc.Csco25g05040.1
			Ca.sc.Csco26g04530.1
			Ca.sc.Csco26g05050.1
			Ca.sc.Csco26g05820.1
			Ca.sc.Csco27g00840.1
			Ca.sc.Csco27g03490.1
			Ca.sc.Csco28g02050.1
			Ca.sc.Csco29g03240.1
			Ca.sc.Csco29g03250.1
			Ca.sc.Csco31g00130.1
		C. breviculmis	Ca.br.g12527.t1
			Ca.br.g13256.t1
			Ca.br.g13267.t1
			Ca.br.g13268.t1
			Ca.br.g13421.t1
			Ca.br.g15142.t1
			Ca.br.g4080.t1
			Ca.br.g4942.t1
			Ca.br.g4945.t1
			Ca.br.g4948.t1
			Ca.br.g5069.t1
			Ca.br.g5158.t1
			Ca.br.g5620.t1
			Ca.br.g5882.t1
			Ca.br.g5883.t1
			Ca.br.g7698.t1
			Ca.br.g7867.t1
			Ca.br.g9673.t1
			Ca.br.g9674.t1
		C. capricornis	Ca.ca.g10623.t1
			Ca.ca.g11272.t1
			Ca.ca.g13298.t1

Appendix Table 1- (Continued)

Type	Subfamily name	Species	Gene name
			Ca.ca.g13880.t1
			Ca.ca.g14659.t1
			Ca.ca.g17944.t1
			Ca.ca.g17945.t1
			Ca.ca.g4792.t1
			Ca.ca.g4904.t1
			Ca.ca.g4992.t1
			Ca.ca.g5140.t1
			Ca.ca.g516.t1
			Ca.ca.g6230.t1
			Ca.ca.g6438.t1
			Ca.ca.g6817.t1
			Ca.ca.g6818.t1
			Ca.ca.g7868.t1
			Ca.ca.g8542.t1
			Ca.ca.g8543.t1
			Ca.ca.g8555.t1
		C. dickinsii	Ca.di.g1091.t1
			Ca.di.g12001.t1
			Ca.di.g13094.t1
			Ca.di.g14121.t1
			Ca.di.g15313.t1
			Ca.di.g16758.t1
			Ca.di.g16759.t1
			Ca.di.g17771.t1
			Ca.di.g17953.t1
			Ca.di.g18341.t1
			Ca.di.g3838.t1
			Ca.di.g3953.t1
			Ca.di.g4704.t1
			Ca.di.g4705.t1
			Ca.di.g4706.t1
			Ca.di.g4718.t1
			Ca.di.g5211.t1
			Ca.di.g5513.t1
			Ca.di.g5666.t1
			Ca.di.g5822.t1
			Ca.di.g752.t1
			Ca.di.g753.t1
			Ca.di.g8012.t1
			Ca.di.g8013.t1
			Ca.di.g8569.t1
			Ca.di.g936.t1

AKNOWLEDGEMENTS

감사의 글

학부 연구생을 시작으로 연구실에 들어와 석사학위를 마치기까지 많은 분께 도움을 받았습니다. 우선 식물분자계통학이란 분야를 처음 접한 저에게 다양한 지식과 열정을 공유해주시고 연구실에서 생활하면서 많은 경험을 하도록 이끌어주신 저의 지도 교수님 김상태 교수님 감사드립니다. 멋모르는 학부 연구생이 나노포어와 같은 최신 기술을 익히고 주도적으로 실험하여 결과를 내도록 북돋아 주시고, 외부 분들과 만날 때도 학생 혼자 나노포어 실험해서 결과를 잘 만들고 있다고 자랑도 아끼지 않으시고, 교수님 덕분에 힘을 많이 얻었습니다. 또 석사 신분으로 미국 큰 학회에 가서 구두 발표를 하는 기회를 만들어주셔서 감사합니다. 저에게 있어서 값진 경험으로 남을 것 같습니다.

저의 학위를 위해 바쁘신 와중에도 시간을 내어 심사해 주시고 지도해주신 성신여자대학교 최상철 교수님, 이재원 교수님 진심으로 감사드립니다. 시간을 따로 내어 파이썬 등 스터디도 진행해 주시고, 학위 주제 관련하여 처음 시도해 보는 다양한 생물정보학적 분석을 수행하는 데 있어 많은 관심과 도움을 주신 저의 또 다른 지도교수님이라 할 수 있는 최상철 교수님 감사드립니다. 익숙하지 않은 연구 주제임에도 불구하고 관심 가지고 심사해주시고 힘든 상황에서 낙담하지 않도록 응원해 주신 이재원 교수님 감사드립니다. 그리고 학부 시절부터 다양한 생명공학 분야를 공부할 수 있게 해주신 성신여자대학교 바이오생명공학과 교수님들께도 감사의 말씀을 드립니다.

사초속이라는 어려운 식물을 연구하는 데 많은 정보와 식물을 제공해 주

신 조양훈 선생님 감사드립니다. 미국 학회에서 처음 만나서 우리 연구실의 연구 주제에 큰 관심을 가지고 직접 수행하는데 어려웠던 gene prediction 작업을 열심히 수행해 준 Bishnu 감사합니다. 여러 번 시도했음에도 불구하고 매번 실패했던 사초속 식물의 염색체 수 세는 실험에 도움을 준 강원대학교 김혜린 언니 감사합니다. 이외에 학부 연구생 시절부터 학위를 마치기까지 도움을 주시고 다양한 경험을 하게 해주신 모든 분에게 감사의 인사를 전하고 싶습니다.

학부 3학년 여름방학부터 석사 학위를 마치기까지 저와 함께 시간을 보낸 모든 연구실 식구도 감사합니다. 덕분에 약 4년이라는 긴 시간 동안 적응도 잘하고 편하게 생활할 수 있었습니다. 학부생 시절부터 따뜻한 관심과 애정을 주신 김성민 박사님, 처음 들어간 연구실에 잘 적응할 수 있도록 도와주신 보라 언니, 윤경 언니, 수현 언니, 대학원 시절 항상 함께하며 서로 돕고 밥도 먹고 동고동락한 현지 언니, 세은 언니, 정우 언니, 지은(이), 민경, 연주 언니, 맡은 바에 최선을 다해주고 꽤 긴 시간 함께한 학부생들인 지은(서), 명보, 승진, 지예, 보라, 예슬, 수경, 그리고 예은 언니, 윤주 언니, 보현, 희지 언니, 규리 언니까지 모두 정말 감사합니다.

연구실 생활하면서 체력적으로도, 심적으로도 힘들어 남몰래 눈물을 훔치기도 했는데, 중간중간 약속을 잡고 만난 친구들 덕분에 힘든 기억도 잠시나마 잊고 행복할 수 있었습니다. 학교, 집, 학교, 집만 왔다 갔다 하는 생활 사이에 힐링과 활력, 그리고 즐거움을 선물해 준 모든 친구들 사랑하고 감사합니다.

마지막으로 무한한 사랑을 주시고 언제나 저의 든든한 버팀목이 되어주는 우리 가족에게도 감사 인사를 전합니다. 제가 우리 가족의 일원이라 너무 행복합니다. 우리 가족 모두 세상에서 제일 존경하고 사랑합니다.