김 상 태 교수 지도
석사학위 청구논문

Expression of floral MADS-box genes in *Nymphaea* based on transcriptome analyses

2018

성신여자대학교 대학원
생물학과
박 지 은

Expression of floral MADS-box genes in *Nymphaea* based on transcriptome analyses

김 상 태 교수 지도

이 논문을 석사학위논문으로 제출함

2018 년 5 월

성신여자대학교 대학원

생물학과

박 지 은

인 준 서

박지은의 석사학위 논문으로 인준함.

2018 년 5 월

심사위원장 최 상 철 인

심사위원 박 종 선 인

심사위원 김 상 태 인

성신여자대학교 대학원

ABSTRACT

As one of the basal angiosperms, Nymphaeaceae (Nymphaeales) is considered a sister group to all other angiosperms except *Amborella* and it has played an important role in the studies of angiosperm evolution. Nymphaeaceae is composed of 5 genera (*Nymphaea, Nuphar, Euryale, Barclaya,* and *Victoria*) and about 80 species and all of them are aquatic plants distributing throughout the world in the temperate and tropical regions. Unlike typical eudicots such as *Arabidopsis*, *Nymphaea* has flowers showing a series of morphological transition from petals to stamens. A taxon showing this special morphology is a perfect model to understand the functional role of floral genes in basal angiosperms. As a transcription factor, MADS-box genes are involved in the development process and floral organ identifications in plants. Based on the identifications of MADS-box genes and their functional studies, the ABCDE model has suggested explaining floral organ identities in *Arabidopsis.* Although modified model such as the fading border model is suggested to explain morphologies of basal angiosperm flowers, our understanding of floral genes in basal angiosperms is still not perfect. As one of the basal angiosperms, understanding genetic control of floral organ identities in *Nymphaea* will provide the key to understand floral diversifications and evolution of angiosperm flowers. In this study, I addressed expression levels of floral MADS-box genes based on transcriptomes from each floral organ in *Nymphaea* X 'Pygmaea Helvola' of which flowers show a clear morphological transition from petal to stamen. Fourteen type-II MADS-box

genes were detected and their orthologies against genes from *Arabidopsis* and *Amborella* are clarified by phylogenetic analysis. Expression of ABCDE genes in *Nymphaea* X 'Pygmaea Helvola' is different from those in *Arabidopsis*. A-class gene (*Ny.Py.He.AG1*) expressed in all floral organs except carpel. Two B-class genes (*Ny.Py.He.AP3* and *Ny.Py.He.PI*) showed broad expressions including sepals and expression of C-class genes (*Ny.Py.He.AG1* and *Ny.Py.He.AG2*) are extended to petals. Expression of D-class (*Ny.Py.He.AG3*), and E-class genes (*Ny.Py.He.SEP1*) showed the same expressions as those in *Arabidopsis*. In ABCDE genes, the expressional gradation of *Ny.Py.He.PI* is well matched with a morphological transition from petals to stamen. As non-ABCDE type-II MADS-box genes, *Ny.Py.He.AGL6* and *Ny.Py.He.AGL15/Ny.Py.He.SVP* also showed positive and negative expressional gradations from petals to stamen, respectively, indicating that these genes potentially involved in morphological transitions from petals to stamens in *Nymphaea*.

List

List of Figures

List of Tables

Ⅰ. Introduction

Flowering plants are one of the most flourishing organisms on earth and are a major component of the ecosystems (Zeng et al., 2014). The flower is a specific feature of angiosperms and distinguishes angiosperms from other plants. Understanding of genetic control of flower formation in early-diverging angiosperms will provide a key to understand diversification and evolution of angiosperms as well as utilization of angiosperms. Recent molecular phylogenetic studies have shown major groups of angiosperms and their relationship (e. g., Zeng et al., 2014; Fig. 1A). Most of the studies agreed that a grade of plant groups including Amborellaes, Nymphaeales, and Austrobailales is placed at that base of angiosperms (Mathews and Donoghue 1999; Qiu et al., 1999; Moore et al., 2007; Zheng et al., 2014; Fig. 1A). In this grade, *Amborella*, a single species in Amborellales, is a sister to other angiosperms. Nymphaeaceae, a monotypic family of Nymphaeales, it is located after the Amborellales at the base of the angiosperm phylogenetic tree (Fig. 1A). Early-diverging angiosperms including Nymphaeales are of particular interest in studies of floral evolution because of their great diversity in the arrangement and number of floral part (Endress et al., 2001; D. Soltis et al. 2002; Yoo et al., 2010a).

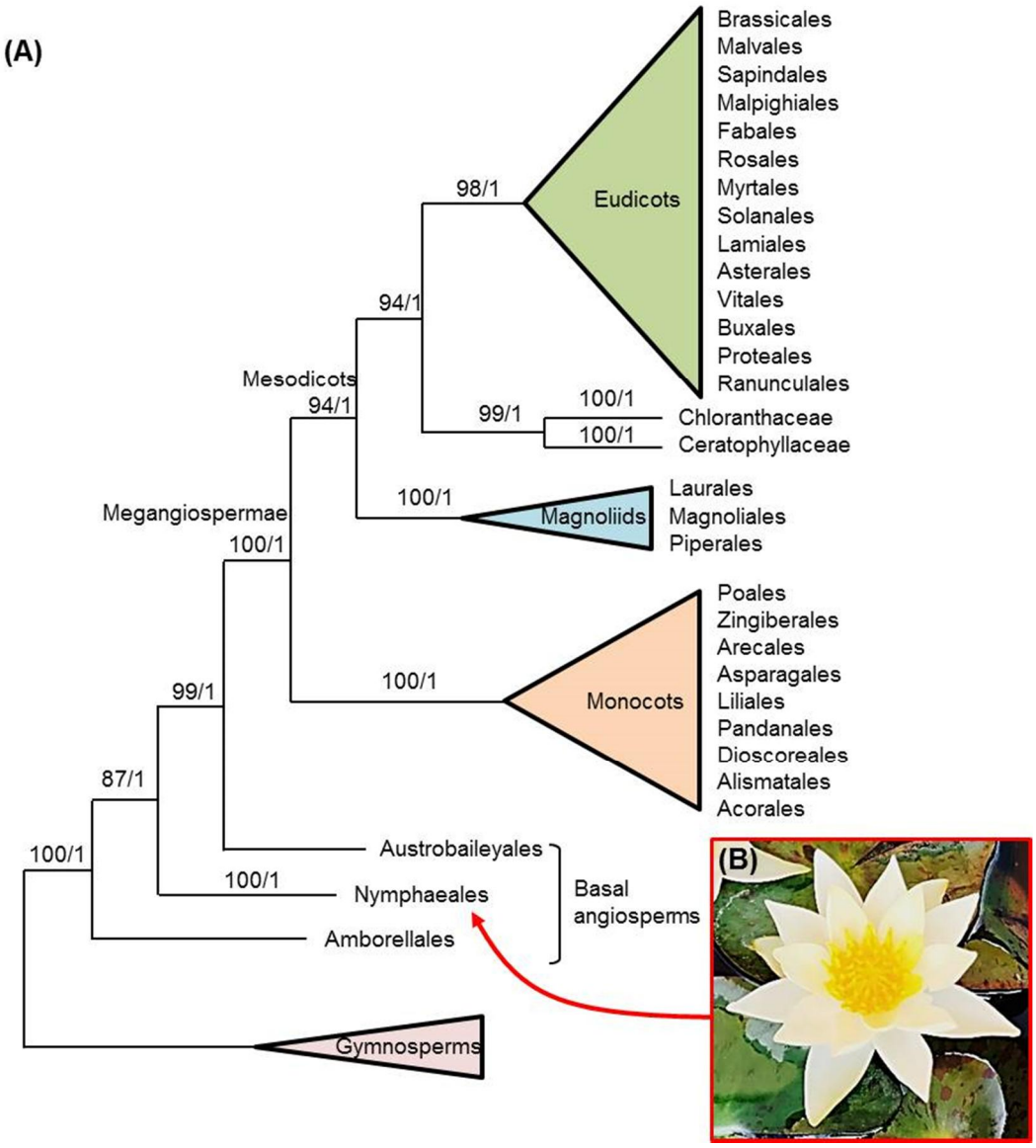

Figure 1. (A) Phylogenetic position of Nymphaeaceae in the angiosperm phylogenetic tree, which is summarized by Zeng et al. (2014). This tree is based on 59 single-copy nuclear genes for 61 representative angiosperms. Numbers above the node indicate bootstrap value and posterior probability. (B) Morphology of *Nymphaea* X 'Pygmaea Helvola'.

Since Nymphaeales have important systematic meaning as well as the beautiful appearance of most members, it attracts attention from many researchers (Borsch et al., 2008; Yoo et al., 2010a).

Nymphaeaceae (water lilies) are composed of five genera (*Nymphaea*, *Nuphar*, *Euryale*, *Barclaya*, and *Victoria*) including ca. 80 species and are distributed throughout the world in temperate and tropical regions (Schneider et al., 2003; Judd et al., 2016; Fig. 1B). They are annual or perennial herbs and are rooted in under-water soil, with leaves and flowers floating on or emergent from the surface (Judd et al., 2016). As a typical flower in Nymphaeaceae, the floral organs of *Nymphaea* were composed to four parts (sepals, petals, stamens, and carpels), like other angiosperms. However, a series of gradual morphological transition from petaloid tepals to petaloid staminodes to functional stamens is observed in most of the species in *Nymphaea* (Borsch et al., 2008; Yoo et al., 2010a). These characters are considered to a distinct feature of most basal angiosperms (Yoo et al., 2010a).

MADS-box genes play an essential role in the development process and organ formation as a transcription factor (Shore et al., 1995; Irish et al., 2003). It has about 180 bp of a conserved domain and the domain binds to specific DNA region to regulate transcripts (Shore et al.,

1995; Nam et al., 2003). Importantly, MADS-box genes play an essential role in the development and formation of plant floral organs (Irish et al., 2003; Nam et al., 2003). In the early MADS-box gene studies, *MINICHROMOSOME MAINTENANCE 1* (***M****CM1*), *AGAMOUS* (***A****G*), *DEFICIENS* (***D****EF*) and *SERUM RESPONSE FACTOR* (***S****RF*) have been found in *Saccharomyces cerevisiae*, *Arabidopsis thaliana, Antirrhinum majus,* and *Homo sapiens,* respectively (Norman et al., 1988; Passmore et al., 1989; Sommer et al., 1990; Yanofsky et al., 1990). The name of "MADS-box" was named by a combination of the first letters of the genes found from above, various organisms.

MADS-box genes are divided into two groups according to the composition of their domains (Alvarez-Buylla et al., 2000). Type Ⅰ MADS-box genes have only MADS-domain and type Ⅱ MADS-box genes has additional conserved K-domain (Pařenicová et al., 2003; Fig. 2A). Especially, type Ⅱ genes shared a conserved structural organization called MIKC, which is composed of **M**ADS-domain, **I**ntervening sequence, **K**eratin-like domain and **C**-terminal (Munster et al., 1997; Fig. 2A). The MIKC-type MADS-box genes are further divided into MIKC* (star) and MIKC^{C} (classic) depending on the length

and exon number of K-box domain (Henschel et al., 2002; Fig. 2A). The K-box domain of MIKC* has a number of exons than that of MIKC^{C} (Henschel et al., 2002; Fig. 2A).

Complete genome sequence of *Arabidopsis thaliana* has determined and 108 genes containing MADS-domain were detected in this genome (Arabidopsis Genome, 2000; Pařenicová et al., 2003). Based on phylogenetic analyses of these genes, Pařenicová et al. (2003) suggested six subgroups of MADS-box genes in *Arabidopsis*: MIKC, Mα, Mβ, Mγ, Mδ, and AGL33 (Fig. 2B). Among these Mα, Mβ, Mγ, and AGL33 correspond to the type Ⅰ MADS-box genes, while Mδ and MIKC correspond to type Ⅱ MADS-box genes (Pařenicová et al. 2003; Fig. 2B). Mδ genes are relevant to MIKC*-genes. Many of type Ⅱ MADS-box genes have reported being involved in the regulation of plant development and the formation of floral organs (Alvarez-Buylla et al. 2000).

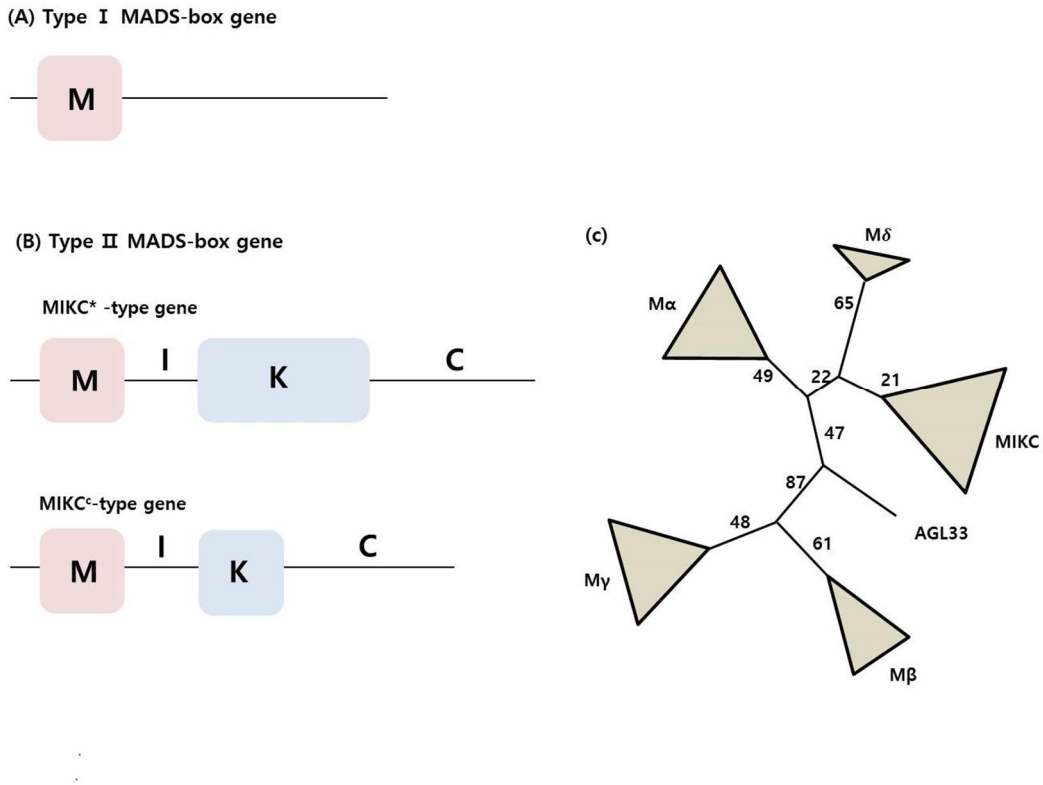

Figure 2. Domain structure and Phylogeny on MADS-box genes. Two types of MADS-box genes and their structure. (A) Non-MIKC type: type I genes plus animal type II genes. (B) MIKC type: type II plant genes. Boxes indicate conserved domains MADS (M) and Keratin-like (K) domains. I, intervening sequences; C, C-terminal sequences. (C) Phylogenetic analysis of MADS-box genes in *Arabidopsis* (Pařenicová et al., 2003).

'ABC model' was proposed for regulating the formation of the floral parts based on studies of genetic mutations in *Arabidopsis* (Coen and Meyerowitz, 1991; Fig. 3A). This model suggests that the formation of each part of flowers employ three gene functions combined (A, B and C). According to this model, A−genes control sepal identity, A− and B− genes control petal identity, B−and C−genes control stamen identity, and C−gene alone controls carpel identity (Coen and Meyerowitz, 1991). Subsequently, D and E functions are added to the ABC model and 'ABCDE model' has been proposed with D function controlling ovule identity and E function interacting with other functions to specify sepal, petal, stamen and carpel identity (Colombo et al., 1995; Pelaz et al., 2000; Honma and Goto, 2001). Therefore, A−, B−, C−, D−, and E− functions control each floral organ development in *Arabidopsis* as the combinatorial manor (Coen and Meyerowitz, 1991; Fig. 3A). In *Arabidopsis thaliana,* A−function genes are *APETALA1* (*AP1*), *APETALA* (*AP2*) and *LEUING* (*LEU*); B−function genes are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*); C−function gene is *AGAMOUS* (*AG*); D−function genes are *SHATTERPROOF1* (*SHP1=AGL1*), *SHATTERPROOF2* (*SHP2=AGL5*) and *SEEDSTICK* (*STK=AGL11*) (Favaro et al., 2003); E−function genes are *SEPALLATA1* (*SEP1=AGL2*), *SEP2* (=*AGL4*), *SEP3* (=*AGL9*) and

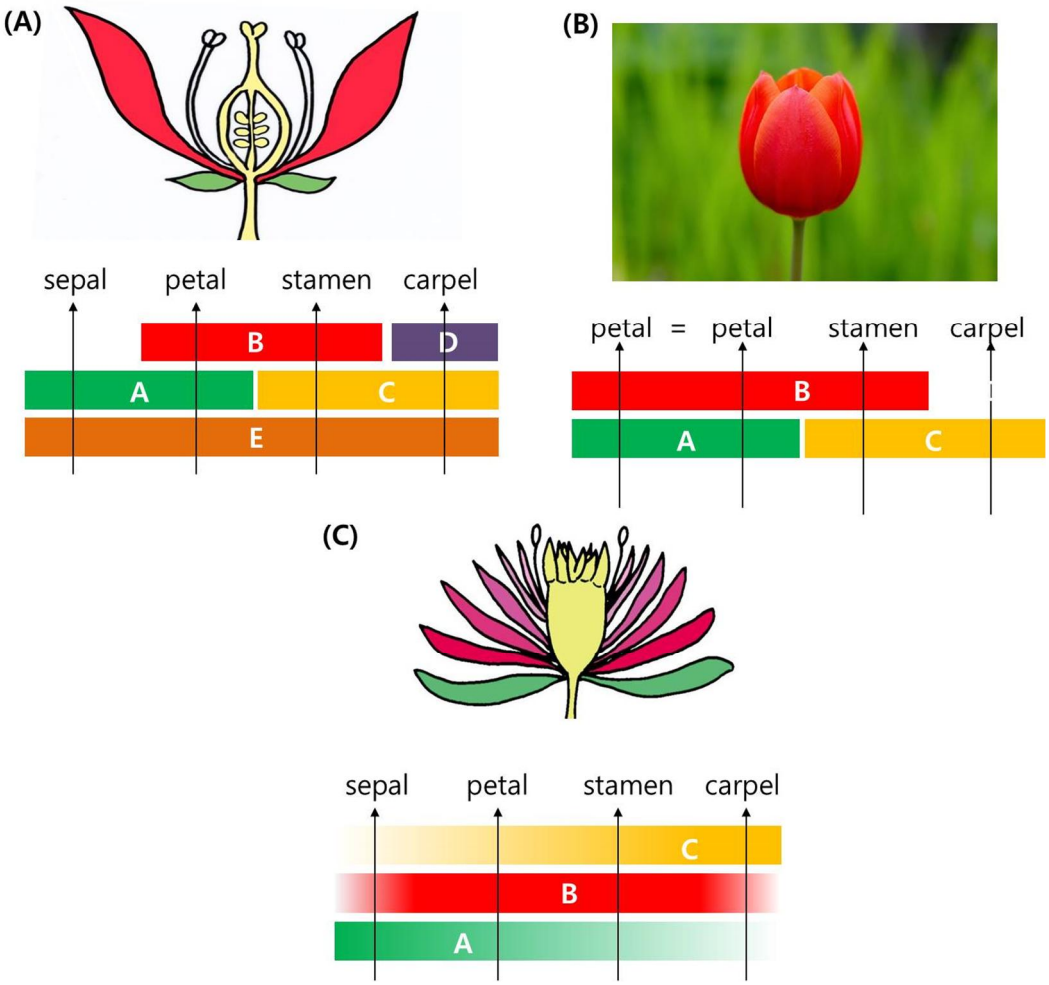

Figure 3. (A) The ABCDE model suggests explaining floral part identities in Arabidopsis (Coen and Meyerowitz, 1991; Colombo et al., 1995; Pelaz et al., 2000; Honma and Goto, 2001). (B) The sliding boundary model (van Tunen et al., 1993; Kramer et al., 2003) to explain identically shaped perianth such as flowers in a tulip. (C) The fading border model (Buzgo et al., 2005) related floral parts in *Nymphaea* (Chanderbali et al., 2016).

SEP4 (=*AGL3*). All of these genes are type II MADS-box genes except *AP2* and *LEU* (Theissen et al., 2000). Modification of the ABC-model has been suggested to explain various floral morphologies and their floral gene expression.

As a modification of ABC model, the "sliding-boundary model" is suggested to explain the multiple petaloid perianth whorls such as flowers in *Lilium*, *Tulipa* and *Ranunculus* (van Tunen et al., 1993; Kramer et al., 2003). According to this model, the expression of B-class genes is extended to the first whorl resulting identical morphology between the first and the second whorls (Fig. 3B). To explain the gradual transition in floral organ morphology in basal angiosperms such as *Amborella* and *Nymphaea*, the "fading borders model" has been suggested (Buzgo et al., 2005; Fig. 3C). In this model, MADS-box genes, which determine each floral organ, was widely expressed, but the expression level decreased at the edge of each zone (Buzgo et al., 2005). The B-class gene expression was broad in *Amborella* and this observation supported the fading border model (Kim et al., 2005).

To address quantification of gene expression, various methodologies have been developed such as 1) quantitative-RT PCR

(e.g., Kim et al., 2005a; Yoo et al., 2010a; Luo et al., 2011), 2) real-time PCR (e.g., Kim et al., 2005b), and 3) microarray (e.g., Yoo et al., 2010b). Recently, whole transcriptome sequencing (RNA-Seq) based on the next generation sequencing (NGS) has developed for a massive finding of expressed genes in samples and quantification of their expression (e. g., Vining et al., 2015; Wang et al., 2016). This method contains the following advantages compared to other methods: 1) it requires a small amount of RNA, 2) it produces high throughput result with relatively low cost, 3) quantification of expression level is objective compared to quantitative RT-PCR, 4) the experimental process is simple compared to real-time PCR and 5) all isoforms including alternatively spliced transcripts are detectable. Furthermore, reference genome information is not needed in NGS based RNA-Seq. Therefore it has become widely applied to model as well as non-model organisms to obtain mass sequence data for molecular marker development, gene discovery, quantification of transcripts, etc (Feng et al., 2012).

Some preliminary studies on expressions of floral genes in Nymphaeales have been done with methods of quantitative RT-PCRs and microarrays (Yoo et al., 2010a; Yoo et al., 2010b; Luo et al., 2011). In the study by Yoo et al., (2010a), thirty-six Floral MADS-box gene

homologs in Nymphaeales including *Cabomba* and *Nymphaea* were detected and their expressions in each floral organ were addressed. In *Nymphaea*, the gene expression patterns suggest that the innermost petals originated from petaloid staminodes. Microarray technique has been applied to the study of the floral transcriptome in *Nuphar* (Nymphaeaceae) (Yoo et al., 2010b). In this study, several thousand *Nuphar* genes with significantly upregulated floral expression, including homologs of the well-known ABCDE floral regulators, deployed in broadly overlapping transcriptional programs across floral organ categories. This showed that the expression patterns of *Nuphar* are the ancestral condition in angiosperms (Yoo et al., 2010b). In the study of Luo et al., (2011), orthologs of *AP2, AP3, PI, AG, SEP,* and *AGL6* have detected in *Nymphaea* and their expressions in each organ were addressed. Based on the results of floral MADS-box genes, Luo et al. (2011) have proposed an application of the fading borders model in *Nymphaea* (Luo et al., 2011).

In this study, I addressed expressions of floral genes in *Nymphaea* X 'Pygmaea Helvola' which shows a morphological transition from petal to stamen using transcriptome sequencing based on NGS and compared it with those from previous studies (Yoo et al., 2010a, b; Luo

et al., 2011). This process includes 1) total transcriptome assembly from each part of flower, 2) detection of MADS-box genes (especially focused on type-II MADS-box genes), 3) addressing relative expression analyses for detected MADS-box genes, and 4) phylogenetic analysis of detected MADS-box genes with those previously reported in *Arabidopsis* and basal angiosperms including Nymphaeaceae and *Amborella*.

Ⅱ. Materials and Methods

1. Materials

In this study, I have selected *Nymphaea* X 'Pygmaea Helvola' as a representative of *Nymphaea* because this species showed a series of morphological transition from perianth to stamen. The plant has cultivated several years at the greenhouse in the Sungshin Women's University to keep the strain for related future work. The voucher specimen is kept in the herbarium of Sungshin Women's University (Fig. 4; *S. Kim 2015-0398*; SWU0016625). Floral parts are collected at the stage of "just before the anthesis (unopened flowering bud)". Flowers are dissected and samples were collected from six morphologically distinguished parts: sepals (SE), outer-most petal whorl (PE1), inner-most petal whorl (PE2), the largest staminodes (SD), inner-most stamens (SN), and carpels (Fig. 5). In addition to floral parts, a young leaf sample (LE) is also included. Dissected floral parts are immediately dropped into liquid nitrogen and stored at −80 ℃ until RNA extractions.

Figure 4. The voucher specimen of a plant used in this study (*S. Kim 2015-0398*; SWU0016652).

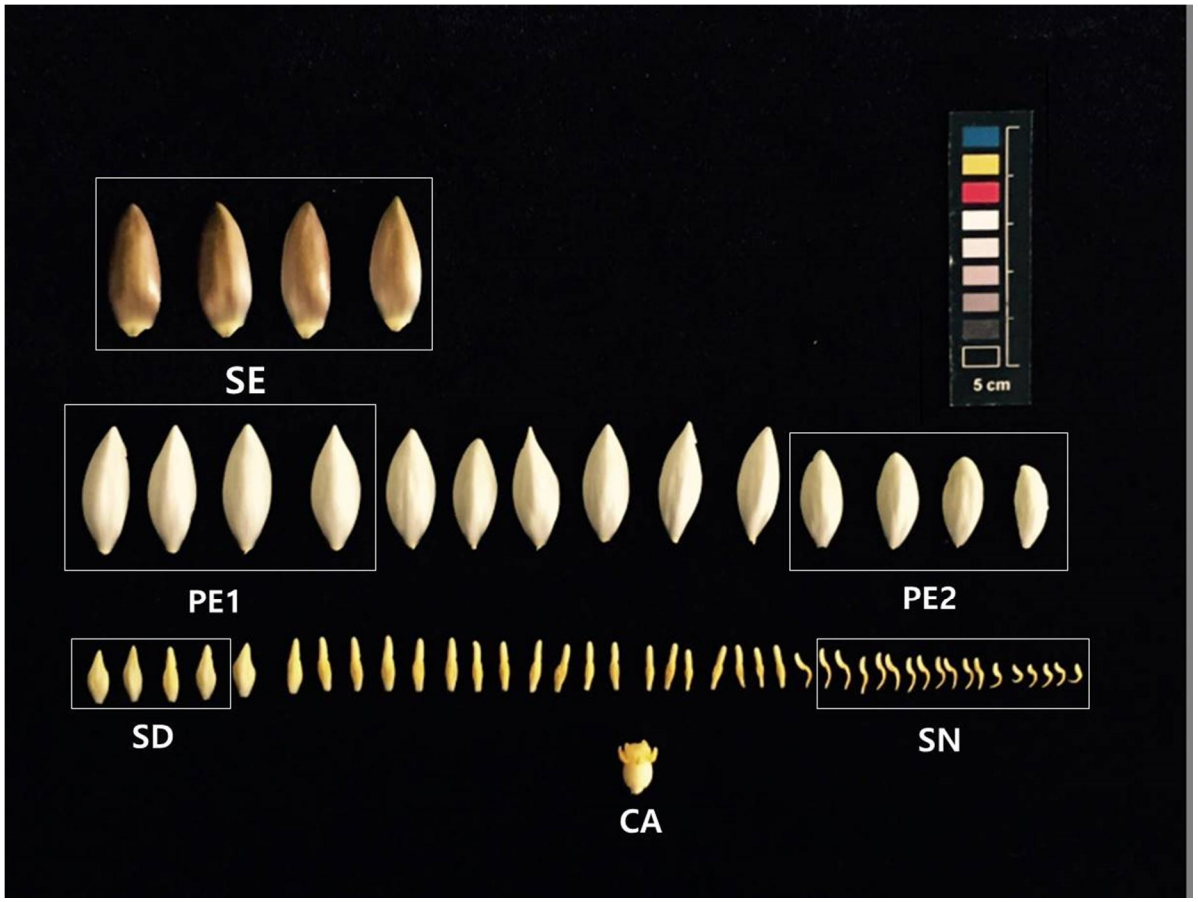

Figure 5. Dissected floral parts of *Nymphaea* X 'Pygmaea Helvola'; a series of morphological transition from four greenish yellow sepals (first row) to 14 white petals (second row) and from petaloid staminode to functional stamens (third row). SE, sepals; PE, petals; SD, staminodes; SN, stamens; CA, carpels.

2. Methods

RNA extraction — About 100 mg of plant samples were grounded using TissueLyser LT (Qiagen, Germany) for RNA extraction. Total RNA was extracted from each sample using the commercial kit (RNeasy Plant Mini; Qiagen, Germany) based on the manufacturer's instruction. The quality and quantity of extracted RNA were confirmed by electrophoresis using 1.5% agarose gel.

Library construction and sequencing —Libraries for paired-end sequencing in HiSeq 2500 (Illumina, San Diego, USA) were constructed using TruSeq rapid SBS kit (Illumina, San Diego, USA) (Fig. 6). Library construction and sequencing have performed by a commercial company (Macrogen Inc., Seoul, Korea) to yield 2 X 100-bp paired-end raw read.

Data analyses— Every raw data of each library was implemented filtering using trimmomatic (Bolger et al., 2014). This step is the process of removing technical sequences such as adapters and polymerase chain reaction (PCR) primers and performing quality filtering in raw data (Bolger et al., 2014). After the quality filtering, *de novo* assemblies have performed for each transcriptome and merged

transcriptome from all seven samples using SOAPdenovo-Trans (Xie et al., 2014).

Functional annotation and analyses of differential expression —In order to decide functional annotation of MADS-box genes, we used InterProScan (ver.5.18-57.0; Jones et al., 2014). Detected MADS-box gene sequences from each transcriptome and merged transcriptome were curated by the following process: 1) translate with a correct frame, 2) recognize initiation and stop codon compare to genes from *Arabidopsis* and *Amborella,* 3) cut out 5′ and 3′ ends outside of the coding region, 4) exclude short sequences less than 200 bp, 5) exclude potential chimeric sequences during the assembly, and 6) remove multiple identical sequences. To address relative expression level of each gene in different samples, 1) quality-filtered raw data from each transcriptome was mapped to sequences of MADS-box genes using Bowtie2 (ver. 1.2.2; Langmead et al., 2009) and 2) FPKM (Fragments Per Kilobase Million) values were calculated based on the abundance of mapped reads using Cufflinks (ver. 2.2.1; Trapnell et al., 2010). Default parameters were used for Bowtie2 and Cufflinks. Differential expressions in each MADS-box genes were clustered and visualized

using Heatmapper (http://www1.heatmapper.ca/; Babicki et al., 2016) with options of average linkage in the clustering method.

Confirmation of cDNA sequences — Detected MADS-box genes from transcriptome analysis were confirmed by Sanger sequencing using PCR of cDNA with specific primers. After the DNase treatment using DNA-*free*™ kit (Ambion, USA), cDNAs were synthesized from DNase treated RNAs with polyT primer (5′-CCG GAT CCT CTA GAG CGG CCG $C(T)_{17}$-3′ (Kramer et al., 1998) using SuperScript™ Ⅱ Reverse Transcriptase (Invitrogen, USA). MADS-box gene-specific primers were designed based on assembled sequences from transcriptome (Table 1). PCR reactions were performed using AmpONE™ Taq DNA Polymerase (GeneAll Inc., Korea) with a file of 1) 95 ℃ for 2 min, 2) 35 cycles of 95 ℃ for 20 sec, 55 ℃ for 40 sec, and 72 ℃ for 30 sec and 3) 72 ℃ for 5 min. The PCR products were checked by 1.3% agarose gel electrophoresis.

Table 1. Gene-specific primers used in this study

Primer name	Primer sequence (5'-3')	Target gene
Ny_mi_Pif	ATG GGG AGA GGT AAG ATT GAG	*Ny.Py.He.PI*
Ny_mi_Pir	CTA TTT GTT CTG CTG CAA GTT AG	
Ny_mi_Bsf	ATG GGA CGT GGG AAG ATT GAG	*Ny.Py.He.Bs*
Ny_mi_Bsr	TCA GAG CTG CAA GCC ATG G	
Ny_mi_AG1f	ATG GGG AGG GGG AAG ATT G	*Ny.Py.He.AG1*
Ny_mi_AG1r	TCA TCC AAG TTG CAA GGC AG	
Ny_mi_AGL2f	ATG GGA AGA GGG AGG GTG	*Ny.Py.He.AGL2*
Ny_mi_AGL2r	TCA TAC TAG CCA CGG AGG G	
Ny_mi_AGL6f	ATG GGG AGG GGG AGA GTC	*Ny.Py.He.AGL6*
Ny_mi_AGL6r	TTA CAG GAC CCA ACC TTG AAC	
Ny_mi_AGL22f	ATG GCG CGG GAG AAA ATC C	*Ny.Py.He.AGL22*
Ny_mi_AGL22r	TCA AAA GGG TAA CCC CAA CCT	

Phylogenetic analysis — To address orthology of each gene to *Arabidopsis* gene, detected MADS-box genes in *Nymphaea* X 'Pygmaea Helvola' were analyzed together with 108 MADS-box genes from *Arabidopsis* genome (Pařenicová et al. 2003). As a closely related species, MADS-box genes from *Amborella* genome (36 genes; Amborella genome project, 2013) and previously reported genes from Nymphaeaceae (29 genes; Yoo et al., 2010a) were also added in the

matrix. DNA sequences of combined data set were translated to amino acid sequence and aligned using CLUSTRALW (Thompson et al., 1997) with default option included in MEGA 7 (Kumar et al., 2016) as a module. Phylogenetic analysis was conducted using the maximum likelihood method (Kumar et al., 2016) in IQ-Tree (Nguyen et al., 2014) with the general time reversible (GTR) + gamma (Γ) + invariant (I) model, which was selected as the best base substitution model using the model test included in the IQ-Tree (Kalyaanamoorthy et al., 2017). To address the confidence of each node, 1,000 replications of bootstrap analysis were performed with the same options as in normal analysis (IQ-tree; Nguyen et al., 2014). Mδ genes were used as outgroup in the analysis of type-II MADS-box genes.

Based on their orthology, each gene was named by a combination of an abbreviation of the scientific name and their subgroup name. In case of detected isoforms in each gene, dash and numbers were followed in the name of the gene (Table 4).

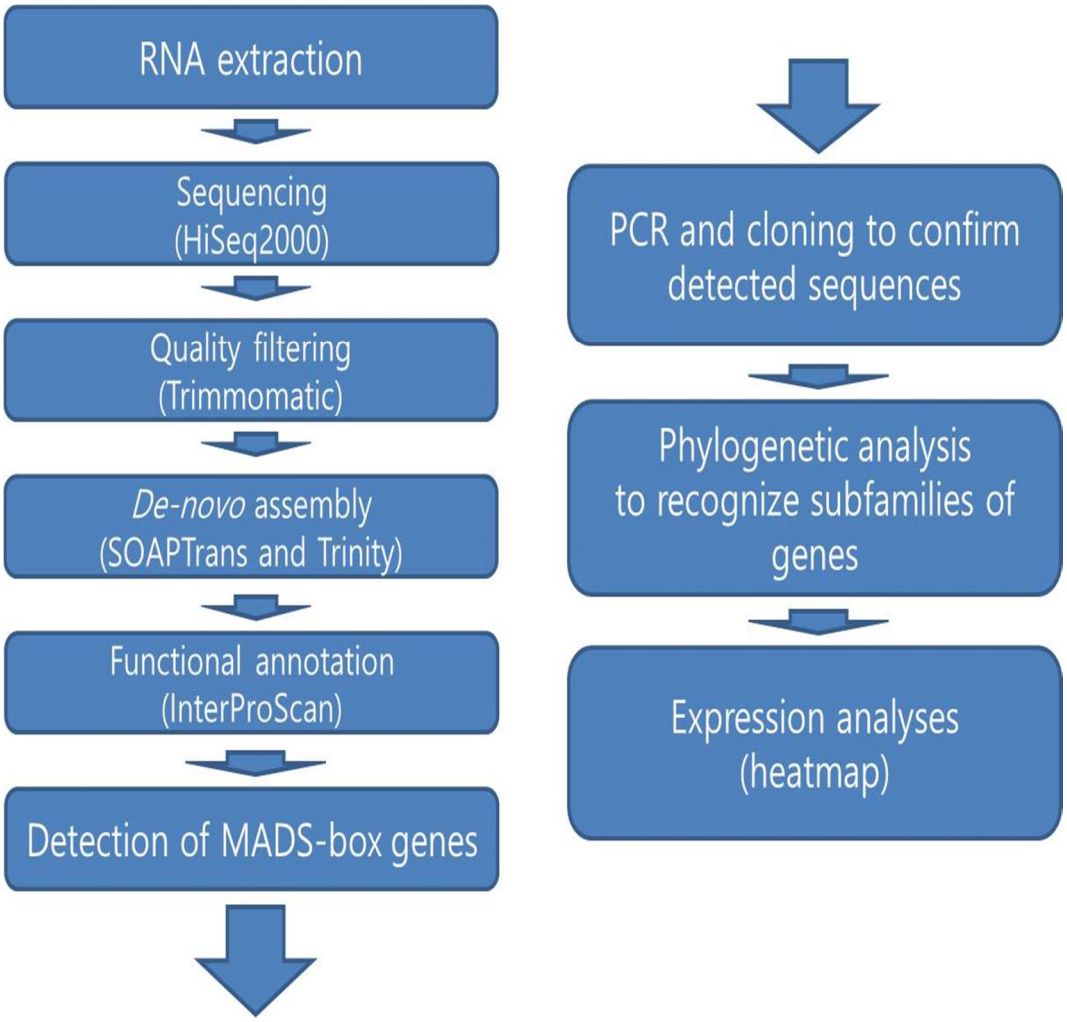

Figure 6. Schematic diagram of RNA-Seq analyses in this study.

Ⅲ. Results

1. Quality filtration and *de novo* assembly

Seven transcriptome libraries from leaves, sepals, petals 1, petals 2, staminodes, stamens, carpels (see Fig. 5) in *Nymphaea* X 'Pygmaea Helvola' were successfully constructed and sequenced. Total reads of raw data are 592 mega reads (592,855,630bp) and the size of each library is in the range of 64.6~141.7 mega reads (64,633,606~141,673,496bp) (Table 2). After the quality filtrations using trimmomatic (Bolger et al., 2014), 0.98% ~0.99% of reads are remained in each sample resulting in the range of 64.2~140.5 mega reads (64,201,738~140,531,664bp).

Sequences from each sample are assembled independently and combined sequences of all seven samples are also assembled using SOAPdenovo-Trans (Xie et al., 2014). In the independent assemblies of seven samples, about 186~ 395 K transcripts (81~176 Mbp) including isoforms are generated (Table 2). The highest N50 value is found in stamens (1,429 bp) and the lowest one is found in petals 2 (623 bp). In the combined data of seven samples, 1.02M transcripts (365 Mbp) are assembled and the N50 value is 746 bp.

Table 2. Statistics of transcriptome sequencings and assemblies in each library

	Leaves	Sepals	Petals 1	Petals 2	Staminodes	Stamens	Carpels	Combined
# of reads (raw data)	71,188,272	86,020,462	64,633,606	82,920,984	141,673,496	68,065,250	78,353,560	592,855,630
# of reads (after filteration)	69,861,642	85,418,086	64,201,738	82,326,314	140,531,664	67,643,268	77,744,698	509,982,712
# of transcripts	306,349	320,653	320,192	313,763	394,351	185,620	366,666	1,022,029
Total length (bp)	116,853,903	133,584,847	132,804,459	136,561,061	176,126,731	81,040,952	152,849,347	364,705,439
Maximum length (bp)	16,490	22,869	21,621	26,297	16,938	12,949	22,524	17,407
Average length (bp)	381.44	416.6	414.77	435.24	446.62	437	416.86	357
Minimum length (bp)	100	100	100	100	100	100	100	100
GC ratio (%)	43.92%	42.98%	42.66%	42.77%	42.46%	44.71%	42.56%	41.53%
N50 length (bp)	973	1,154	1,085	623	1,292	1,492	1,123	746

2. Detection of MADS-box genes and phylogenetic analysis

Proteins domains are detected from assembled transcripts from each data and also combined data using InterProScan (Jones et al., 2014). One hundred sixty-nine MADS-box genes in total with their 89 isoforms were detected from seven assemblies and an assembly from combined data (Table 3). The highest (30) and the lowest (15) MADS-box gene numbers are found in staminodes and sepals, respectively. The preliminary phylogenetic analysis of detected MADS-box genes identifies type-Ⅱ MADS-box genes based on their orthology to *Arabidopsis* MADS-box genes (data not shown). After the careful curation of these genes (see materials and methods), 51 type-Ⅱ MADS-box genes (including isoforms) were identified and many of them were confirmed by Sanger sequencing using cDNAs of these samples (Table 4 and 5). All of the genes were blasted in GenBank (https://blt.ncbi.nlm.nih.gov/Blast.cgi) to confirm the presence of their MADS-domain.

The phylogenetic analysis of 1) detected type - II MADS-box genes from *Nymphaea* X 'Pygmaea Helvola' in this study, 2) type-II MADS-box genes from genomes of *Arabidopsis* and *Amborella,*

and 3) previously reported MADS-box genes in *Nymphaea* (Yoo et al., 2010a) identifies subfamilies of newly detected genes based on previously suggested subfamily category of type-II MADS-box genes by Becker and Theissen (2003)(Fig. 7). The ML tree shows that all sequences detected in *Nymphaea* X 'Pygmaea Helvola' were placed in each clade which is relevant to each subfamily with high supporting values (>70 %). As a result, orthologous relationships between newly identified genes of *Nymphaea* X 'Pygmaea Helvola' in this study and *Arabidopsis* genes were clarified (Table 4). Based on their orthology, each gene was named by a combination of an abbreviation of the scientific name and their subgroup name. In case of detected isoforms in each gene, dash and numbers were followed in the name of the gene (Table 4).

Table 3. Number of MADS-box genes detected in this study

Library	# of MADS genes detected	# of MADS genes including isoforms
Combined	17	33
Leaves	18	25
Sepals	15	22
Petals 1	23	29
Petals 2	21	29
Staminodes	30	43
Stamens	22	33
Carpels	23	44
Total	169	258

Table 4. List of type-II MADS-box genes from *Nymphaea* transcriptome

Group	Subfamily name	Gene name	Sequence length	# of clones confirmed by Sanger sequencing
MIKC	SQUA	*Ny.Py.He.AP1-1*	792	
		Ny.Py.He.AP1-2	780	
		Ny.Py.He.AP1-3	780	
		Ny.Py.He.AP1-4	495	
		Ny.Py.He.AP1-5	780	
		Ny.Py.He.AP1-6	780	
		Ny.Py.He.AP1-7	744	
	DEF	*Ny.Py.He.AP3-1*	507	
		Ny.Py.He.AP3-2	507	
		Ny.Py.He.AP3-3	408	
	GLO	*Ny.Py.He.PI-1*	654	6
		Ny.Py.He.PI-2	510	
	GGM13	*Ny.Py.He.Bs-1*	699	1
		Ny.Py.He.Bs-2	699	
		Ny.Py.He.Bs-3	702	
	AG	*Ny.Py.He.AG1-1*	675	3
		Ny.Py.He.AG1-2	675	
		Ny.Py.He.AG1-3	675	
		Ny.Py.He.AG1-4	558	
		Ny.Py.He.AG2-1	702	
		Ny.Py.He.AG2-2	702	
		Ny.Py.He.AG2-3	702	
		Ny.Py.He.AG2-4	723	
		Ny.Py.He.AG2-5	702	
		Ny.Py.He.AG3-1	657	
	AGL2	*Ny.Py.He.SEP1-1*	738	6
		Ny.Py.He.SEP1-2	738	
		Ny.Py.He.SEP1-3	729	
		Ny.Py.He.SEP1-4	738	
	AGL6	*Ny.Py.He.AGL6-1*	735	7

		Ny.Py.He.AGL6-2	735	
		Ny.Py.He.AGL6-3	735	
		Ny.Py.He.AGL6-4	735	
		Ny.Py.He.AGL6-5	735	
	AGL12	*Ny.Py.He.AGL12-1*	603	
		Ny.Py.He.AGL12-2	498	
	AGL15	*Ny.Py.He.AGL15-1*	732	
		Ny.Py.He.AGL15-2	732	
		Ny.Py.He.AGL15-3	732	
	TM3	*Ny.Py.He.SOC1-1*	648	
		Ny.Py.He.SOC1-2	648	
		Ny.Py.He.SOC1-3	648	
		Ny.Py.He.SOC1-4	648	
		Ny.Py.He.SOC1-5	648	
	STMADS11	*Ny.Py.He.SVP-1*	738	1
		Ny.Py.He.SVP-2	747	
Mδ		*Ny.Py.He.AGL65-1*	1,053	
		Ny.Py.He.AGL65-2	1,053	
		Ny.Py.He.AGL65-3	1,053	
		Ny.Py.He.AGL65-4	1,053	
		Ny.Py.He.AGL65-5	1,056	

Table 5. MADS-box genes in *Nymphaea* compared with *Arabidopsis*

Phylogenetic group	Subfamily	Gene name	*Nymphaea* (transcriptome)	Confirmed by Sanger sequencing	Function in *Arabidopsis* (Smaczniak et al., 2012)
MIKC	SQUA	*Ny.Py.He.AP1*	O	X	A-class gene; meristem identity specification
	GLO	*Ny.Py.He.AP3*	O	O	B-class gene; petal and stamen specification
	DEF	*Ny.Py.He.PI*	O	X	
	AG	*Ny.Py.He.AG1*	O	O	C-class gene; carpel and stamen specification
	AG	*Ny.Py.He.AG2*	O	X	
	AG	*Ny.Py.He.AG3*	O	X	D-class gene; carpel and ovule development
	AGL2	*Ny.Py.He.AGL2*	O	O	E-class gene; sepal, petal, stamen and carpel specification
	AGL6	*Ny.Py.He.AGL6*	O	O	transition to flowering (activator); lateral organ development
	AGL12	*Ny.Py.He.AGL12*	O	X	root development cell-cycle regulation; transition to flowering (activator)
	AGL15	*Ny.Py.He.AGL15*	O	X	embryogenesis, transition to flowering (repressor) with AGL18; sepal and petal longevity; fruit maturation
	TM3	*Ny.Py.He.SOC1*	O	X	transition to flowering (activator); periodic lateral root formation
	STMADS11	*Ny.Py.He.SVP*	O	O	transition to flowering (repressor)
	GGM13	*Ny.Py.He.Bs*	O	O	seed pigmentation and endothelium development
Mδ		*Ny.Py.He.AGL65*	O	X	pollen maturation and tube growth

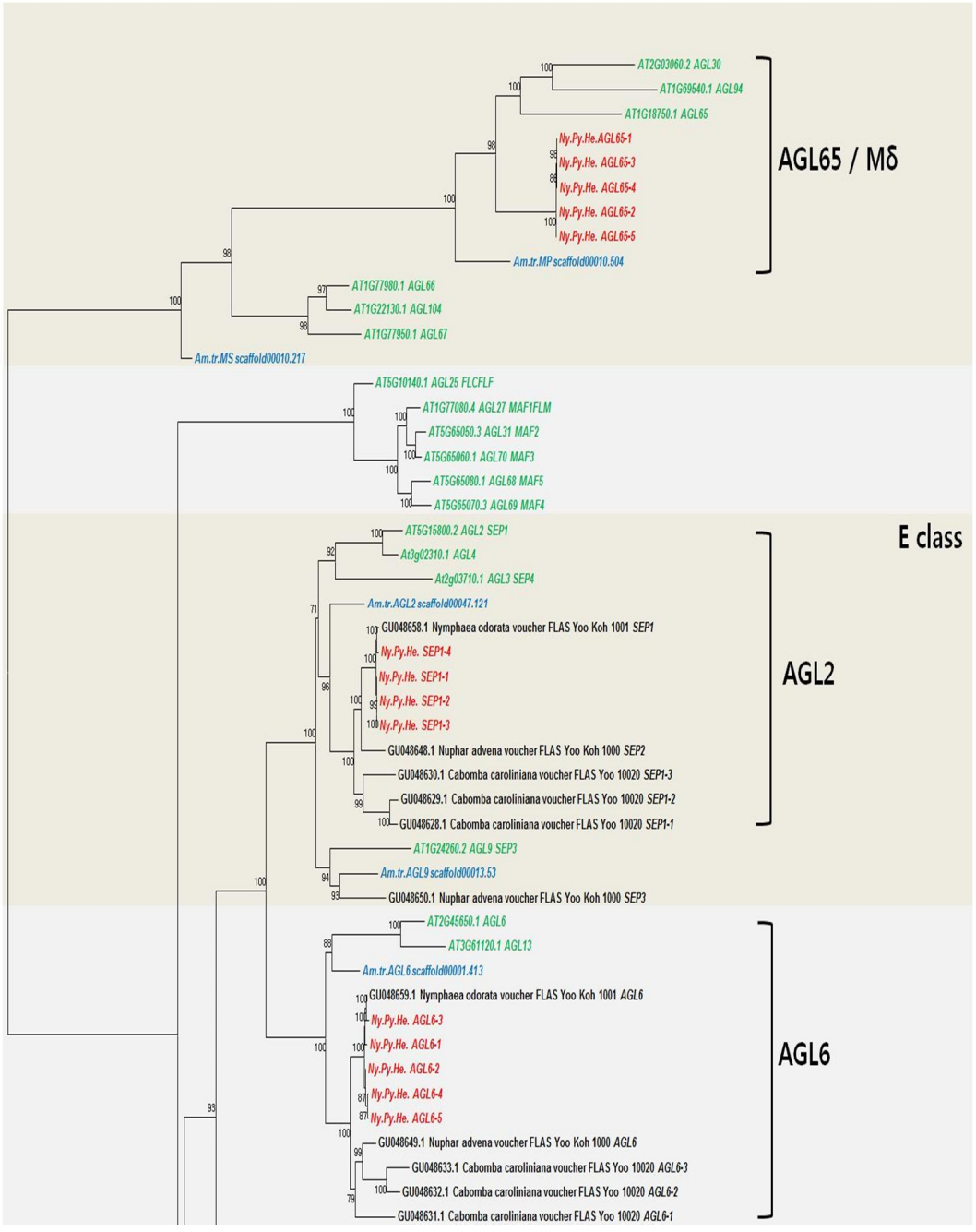

Figure 7. The maximum likelihood tree of type-II MADS-box genes from *Nymphaea*, *Arabidopsis,* and *Amborella.* Values above the node indicate bootstrap supports. Values below 50 % were deleted.

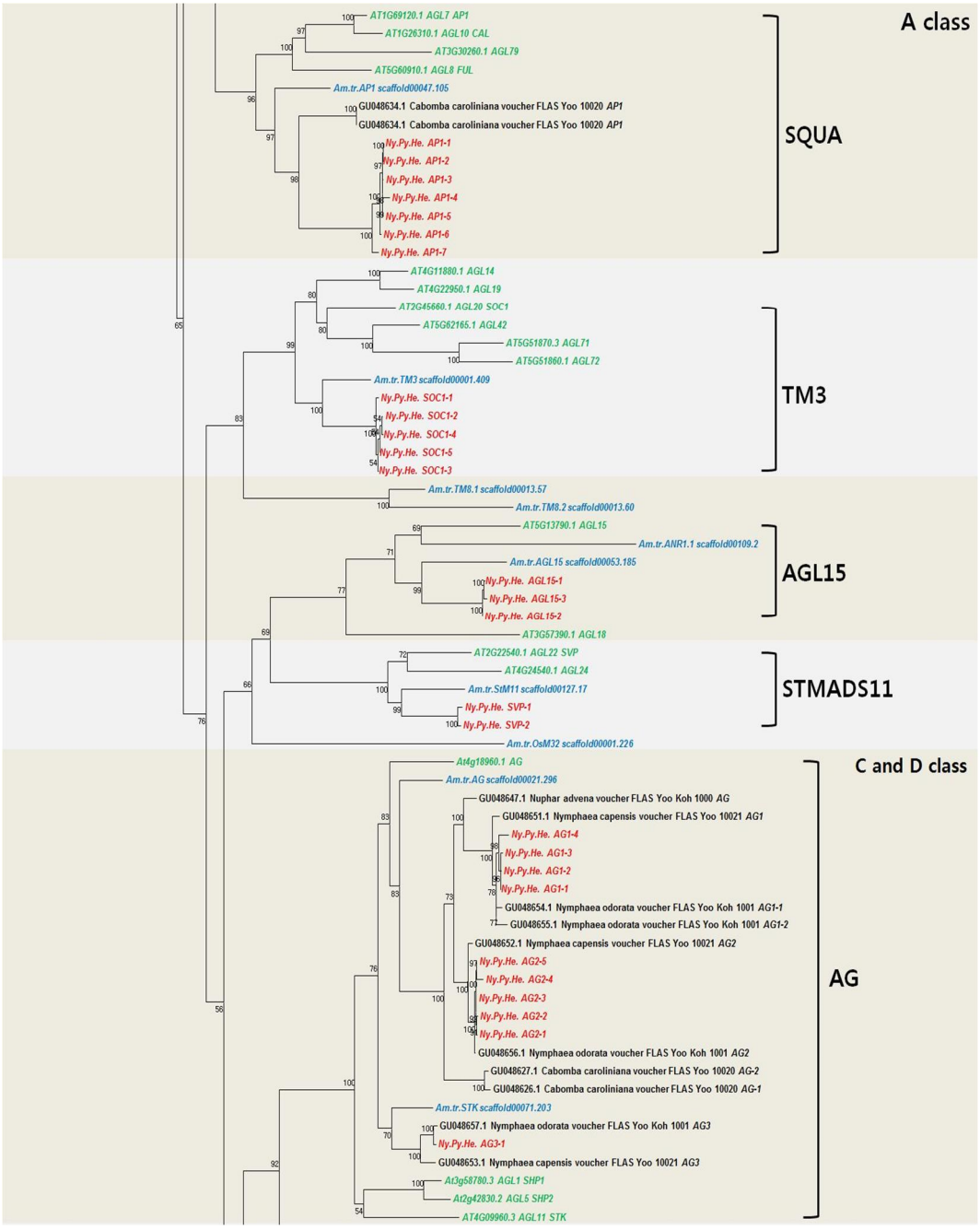

Figure 7. Continued

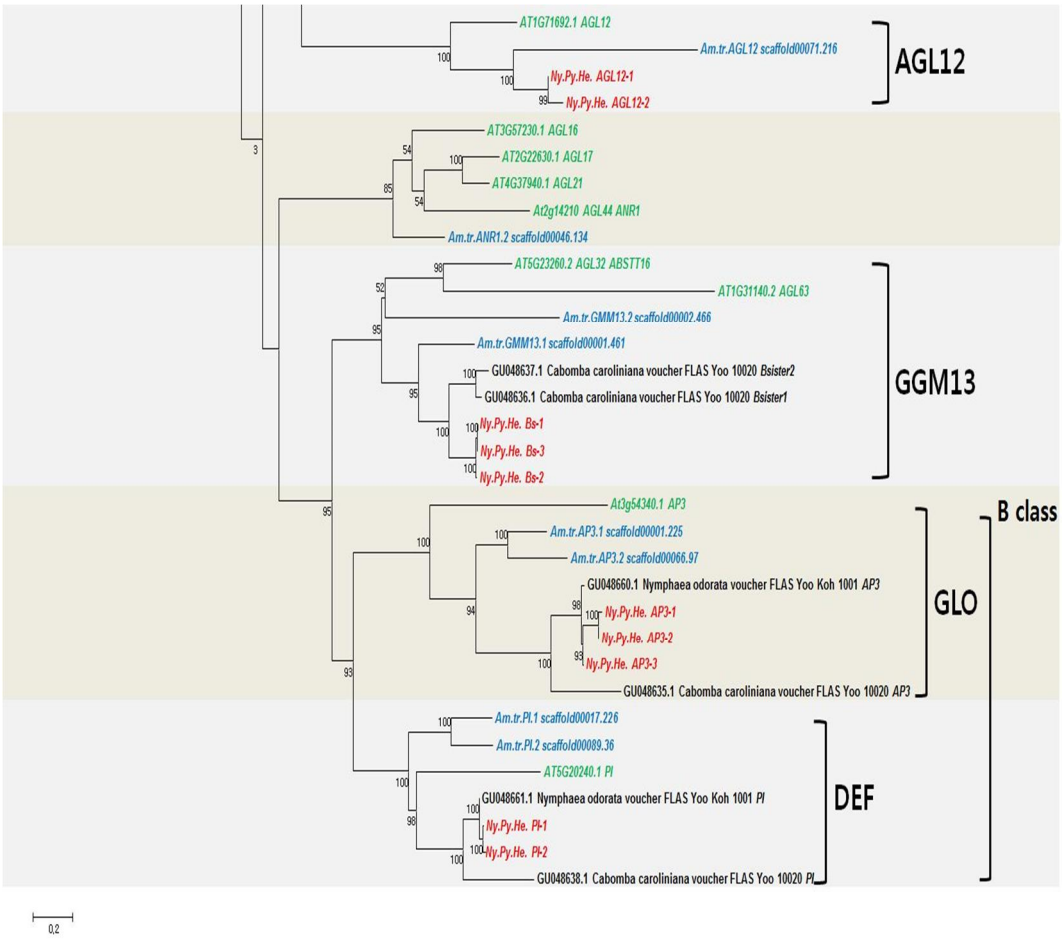

Figure 7. Continued

3. Analysis of gene expressions based on FPKM

To address relative expression level of each MADS-box gene in different tissues, quality-filtered reads are mapped on detected MADS-box genes using Bowtie2 (ver. 1.2.2; Langmead et al., 2009) and FPKM values were calculated with mapped result using Cufflinks (ver. 2.2.1; Trapnell et al., 2010) (Table 6). Based on FPKM value in each gene, relative expressions were visualized by heatmap using Heatmapper (http://www1.heatmapper.ca/; Babicki et al., 2016) based on 1) original FPKM values (Fig. 8), 2) log FPKM values (Fig. 9), and 3) normalized FPKM values (Fig. 10). Normalized values of each gene for different tissues were compared as histograms side by side (Fig. 11).

As A-class genes, *AP1* homolog (*Ny.Py.He.AP1*) and their isoforms were expressed all organs except carpels, with the strongest expression in leaves. The maximum expression was found in leaves. In the two B-class genes, *AP3* homolog (*Ny.Py.He.AP3*) and their isoforms were expressed all organ except leaves, with the strongest expression in petals 1. Two isoforms of *PI* homologs (*Ny.Py.He.PI*) were expressed in sepals, petals 1, petals 2, staminodes, and stamens. Gradual increase of expression was found from outer perianth to stamens. *Ny.Py.He.PI* and *Ny.Py.He.AP3* showed similar expression patterns except for leaf and

carpel. The B-class genes exhibit a broader expression, unlike the expression pattern in *Arabidopsis*. These results are the same as those in *Amborella*, which is a sister to all other angiosperms. As C-class genes, three different *AG* homologs (*Ny.Py.He.AG1~3*) and their isoforms were detected and these are orthologs to previously detected three *AG* homologs in *Nymphaea* (Yoo et al., 2010a). *Ny.Py.He.AG1* was expressed in all floral organs except sepals. All expression levels were similar, but the minimum expression was found in petals 1 (Fig. 11). The highest expression of *Ny.Py.He.AG2* was found in staminodes and moderate expressions are found in petals 2, stamens, and carpels. Interestingly only little expression is found in the outer-most petals (PE1). *Ny.Py.He.AG3* showed carpel specific expression. Strong expression of *SEP1* homolog (*Ny.Py.He.SEP1*) is found in all the organs except for the leaves. It is confirmed that *Ny.Py.He.SEP1* is a floral specific gene in *Nymphaea* (Fig. 11) and this is the same as the E-class gene expression in *Arabidopsis*.

As non-ABCDE genes, strong expressions of *AGL6* homolog (*Ny.Py.He.AGL6*) is found in sepals and outer-most petals but the expression reduced in inner petals and more reduced in staminodes (Fig. 11). This pattern follows morphological gradation from petal to stamen,

indicating that *Ny.Py.He.AGL6* potentially a positive factor of petal-like morphology. Interestingly, only one isoform of *Ny.Py.He.AGL6* is moderately expressed on leaves. AGL6 subfamily is a sister to SEP subfamily in the previous studies on the phylogeny of MADS-box genes (Kim et al., 2005a). AGL15 homologs (*Ny.Py.He.AGL15*) and *SVP* homologs (*Ny.Py.He.SVP*) are expressed in most of the organs except sepal (Fig. 11). Especially, expression level is increased inwardly from petal to carpel indicating these genes are potentially positive factors to forming stamen-like morphology but negative factors to forming petal-like morphology. Genes of B-sister subgroup (*Ny.Py.He.Bs*) and *AGL12* homologs (*Ny.Ph.He.AGL12*) are showed carpel-specific expression (Fig. 11). Strong expression of *AGL65* homologs (*Ny.Py.He.AGL65*) and *SOC1* homologs (*Ny.Py.He.SOC1*) are found in leaves and expressions of these genes in flowers are minor (Fig. 11)

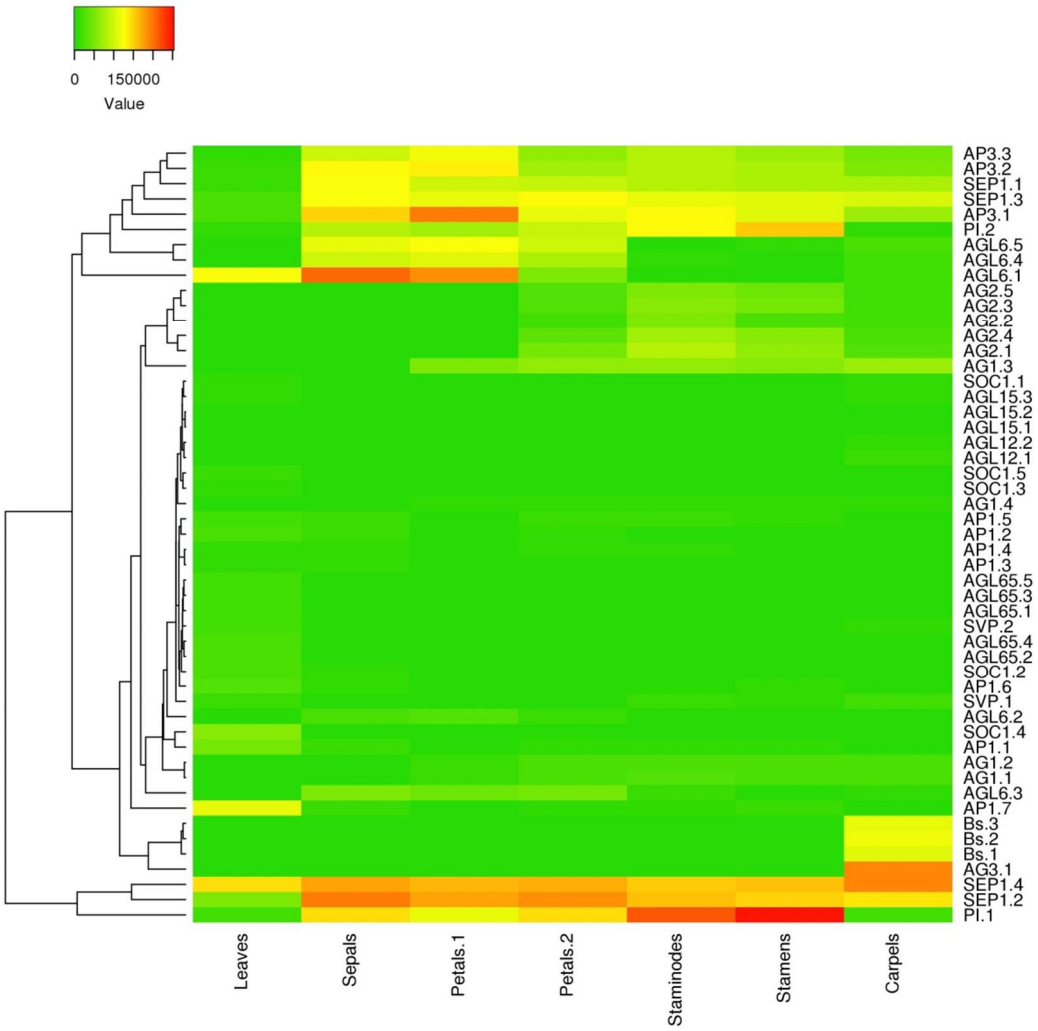

Figure 8. The expression level of type-II MADS-box genes based on their FPKM value (original) and their relationships.

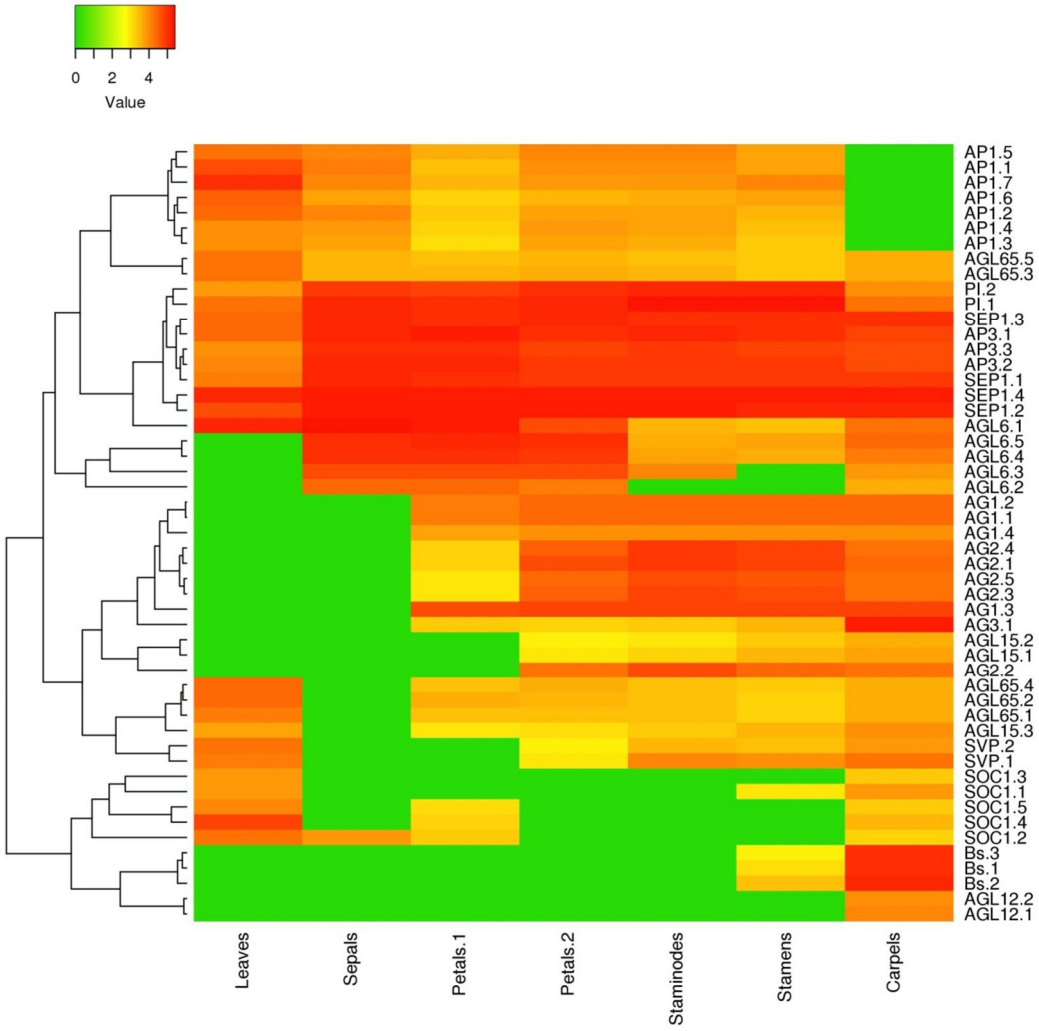

Figure 9. The expression level of type-II MADS-box genes based on their FPKM value (log scale) and their relationships.

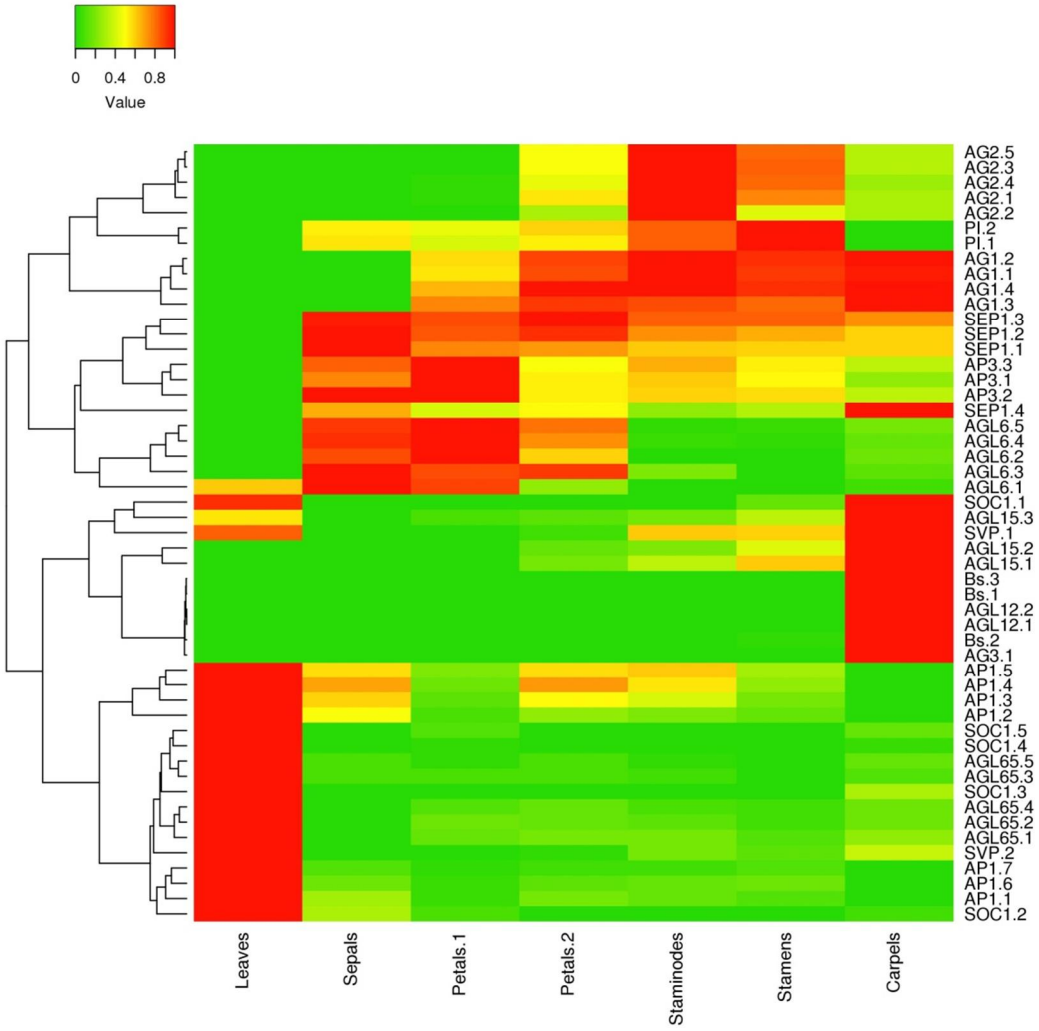

Figure 10. The expression level of type-II MADS-box genes based on their FPKM value (normalized) and their relationships.

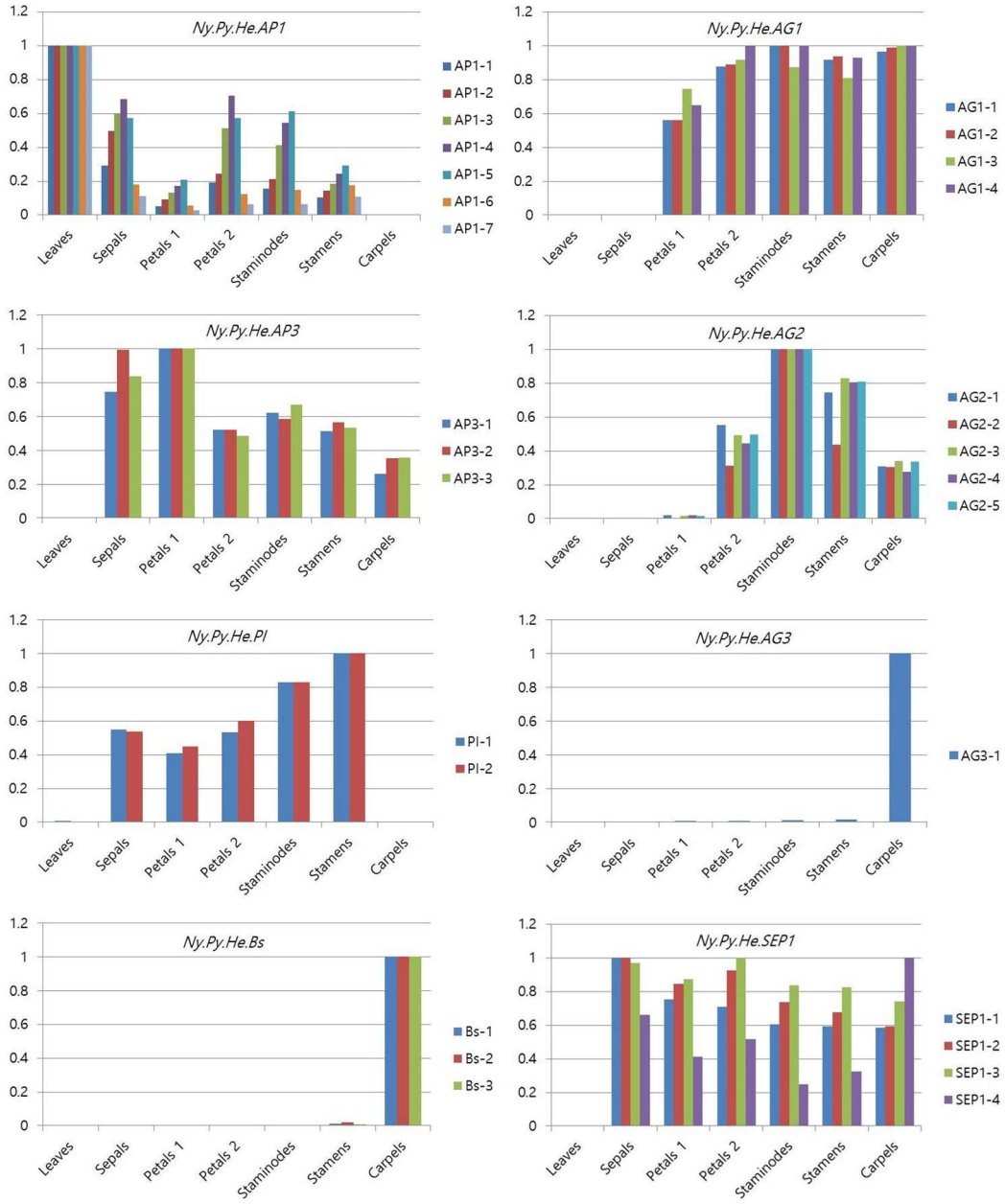

Figure 11. Expression of type Ⅱ MADS-box genes in *Nymphaea* X 'Pygmaea Helvola' based on normalized FPKM

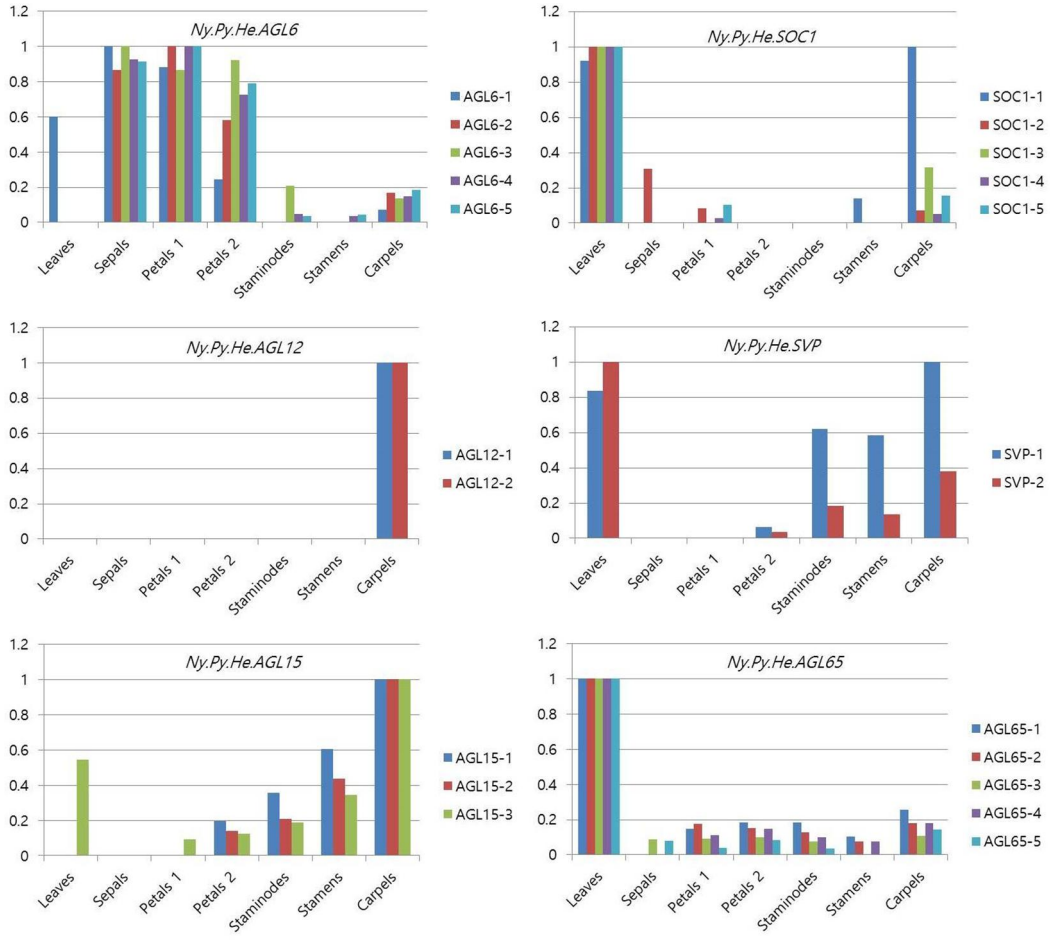

Figure 11. Continued.

Table 6. Relative expressions (FPKM) of MADS-box genes in each floral parts and leaves

MADS-homologues	Leaves	Sepals	Petal 1	Petal 2	Staminodes	Stamens	Carpels
Ny.Py.He.AP1-1	49908.1	14597.9	2592.92	9544.4	7858.08	5125.52	0
Ny.Py.He.AP1-2	22662.8	11247.9	2110.83	5543.62	4777.71	3253.45	0
Ny.Py.He.AP1-3	9822.29	5852.95	1284.57	5039.66	4057.89	1820.56	0
Ny.Py.He.AP1-4	9684.94	6629.03	1670.03	6819.17	5254.42	2370.01	0
Ny.Py.He.AP1-5	18083	10375.5	3773.96	10331.3	11088.4	5291.41	0
Ny.Py.He.AP1-6	29779	5317.91	1665.05	3642.62	4464.18	5202.97	0
Ny.Py.He.AP1-7	115874	12751.6	3259.03	7398.71	7483.45	12680.9	0
Ny.Py.He.AP3-1	21736.3	151410	196183	112794	129665	111092	66909.7
Ny.Py.He.AP3-2	10364.9	131493	132234	73859.9	81469.6	79112.8	53152.6
Ny.Py.He.AP3-3	9879.4	100460	117986	62166.4	82303.7	67263.1	48339.8
Ny.Py.He.PI-1	17888.7	146117	113360	142603	212892	252950	16446
Ny.Py.He.PI-2	7639.92	85255.2	72301.8	94117	127109	152102	8421.42
Ny.Py.He.Bs-1	0	0	0	0	0	1113.05	109432
Ny.Py.He.Bs-2	0	0	0	0	0	2445.69	120792
Ny.Py.He.Bs-3	0	0	0	0	0	689.53	115798
Ny.Py.He.AG1-1	0	0	14349.7	22444.2	25638.4	23460.5	24741
Ny.Py.He.AG1-2	0	0	13419	21288.1	23957.8	22461.8	23682.6
Ny.Py.He.AG1-3	0	0	52319.2	64258.1	61272.6	56875.8	70193.6
Ny.Py.He.AG1-4	0	0	5832.69	8986.54	8947.53	8340.37	8960.33
Ny.Py.He.AG2-1	0	0	1691.88	46242.5	83872.5	62437.8	25946.8
Ny.Py.He.AG2-2	0	0	0	17145.8	54991.1	23991.5	16765.6

Ny.Py.He.AG2-3	0	0	1041.91	29316	59331.5	49062.6	20125.4
Ny.Py.He.AG2-4	0	0	1509.74	33465.9	75128	60422.2	20631
Ny.Py.He.AG2-5	0	0	911.877	26116.8	52763.2	42764.8	17838.5
Ny.Py.He.AG3-1	0	0	1746.3	1497.33	2093.76	3257.26	191777
Ny.Py.He.SEP1-1	14057.1	125715	97928.4	93157	81734.1	80170.5	79213.3
Ny.Py.He.SEP1-2	52871.5	195745	173392	185283	157947	149376	137237
Ny.Py.He.SEP1-3	22403.6	125720	115500	129239	111687	110334	101324
Ny.Py.He.SEP1-4	144393	175917	163980	168952	156197	159893	192060
Ny.Py.He.AGL6-1	125243	207030	182988	52106.2	2870.59	2396.76	17447.9
Ny.Py.He.AGL6-2	0	22484.8	26028.5	15113.5	0	0	4361.97
Ny.Py.He.AGL6-3	0	52606	45409.1	48341.4	10968.3	0	7060.81
Ny.Py.He.AGL6-4	0	100545	108734	78950.4	5358.81	4107.06	16160.7
Ny.Py.He.AGL6-5	0	111376	122211	96390.2	4403.56	5565.24	22577.3
Ny.Py.He.AGL12-1	0	0	0	0	0	0	12645.7
Ny.Py.He.AGL12-2	0	0	0	0	0	0	9214.98
Ny.Py.He.AGL15-1	0	0	0	938.116	1711.06	2909.82	4818.17
Ny.Py.He.AGL15-2	0	0	0	663.46	971.602	2048.49	4683.19
Ny.Py.He.AGL15-3	5120.85	0	874.116	1149.29	1783.8	3249.61	9401.96
Ny.Py.He.SOC1-1	6078.84	0	0	0	0	929.69	6588.1
Ny.Py.He.SOC1-2	20838.5	6392.41	1780.77	0	0	0	1459.62
Ny.Py.He.SOC1-3	6283.77	0	0	0	0	0	1995.95
Ny.Py.He.SOC1-4	58844.5	0	1526.22	0	0	0	3090.97
Ny.Py.He.SOC1-5	11666.5	0	1194.99	0	0	0	1797.55
Ny.Py.He.SVP-1	13775.9	0	0	1053.33	10166.3	9622.7	16425.7

Ny.Py.He.SVP-2	19048.7	0	0	662.2	3519.08	2573.51	7214.78
Ny.Py.He.AGL65-1	15452.2	0	2275.14	2835.2	2840.01	1599.77	3939.64
Ny.Py.He.AGL65-2	21529.4	0	3774.08	3314.35	2725.67	1651.34	3870.54
Ny.Py.He.AGL65-3	16980.9	3550.04	3566.65	3689.18	3360.52	2232.19	3818.8
Ny.Py.He.AGL65-4	24673.7	0	2786.47	3689.18	2441.97	1922.4	4434.25
Ny.Py.He.AGL65-5	20214.6	3452.33	2672.25	3467.37	2614.66	1967.6	4611.16

Ⅳ. Discussion

ABCDE model and fading border model in *Nymphaea*

Classic ABCDE model (Coen and Meyerowitz, 1991; Colombo et al., 1995; Pelaz et al., 2000; Honma and Gota, 2011) explains discrete morphology of each floral part based on the digitalized signal which is expressional on or off of the gene. However, the fading border model (Buzgo et al., 2005) explains the morphological transition based on the combination of gradational expression(s) of related genes. In the morphological transition of *Nymphaea* X 'Pygmaea Helvola', the expression patterns of the MADS-box genes are expected to be matched with the 'fading border model' whatever the involved genes are.

In this study, fourteen type-II MADS-box genes were detected and their orthologies against genes from *Arabidopsis* and *Amborella* are clarified by phylogenetic analysis. Expression of ABCDE genes in *Nymphaea* X 'Pygmaea Helvola' is different from those in *Arabidopsis*. In *Nymphaea* X 'Pygmaea Helvola', A-class gene (*Ny.Py.He.AG1*) expressed in all floral organs except carpel (Fig. 12). Kim et al. (2005) have suggested that *AP1* homologs may have different functions in basal angiosperms than in eudicots. On the basis of the broader patterns of

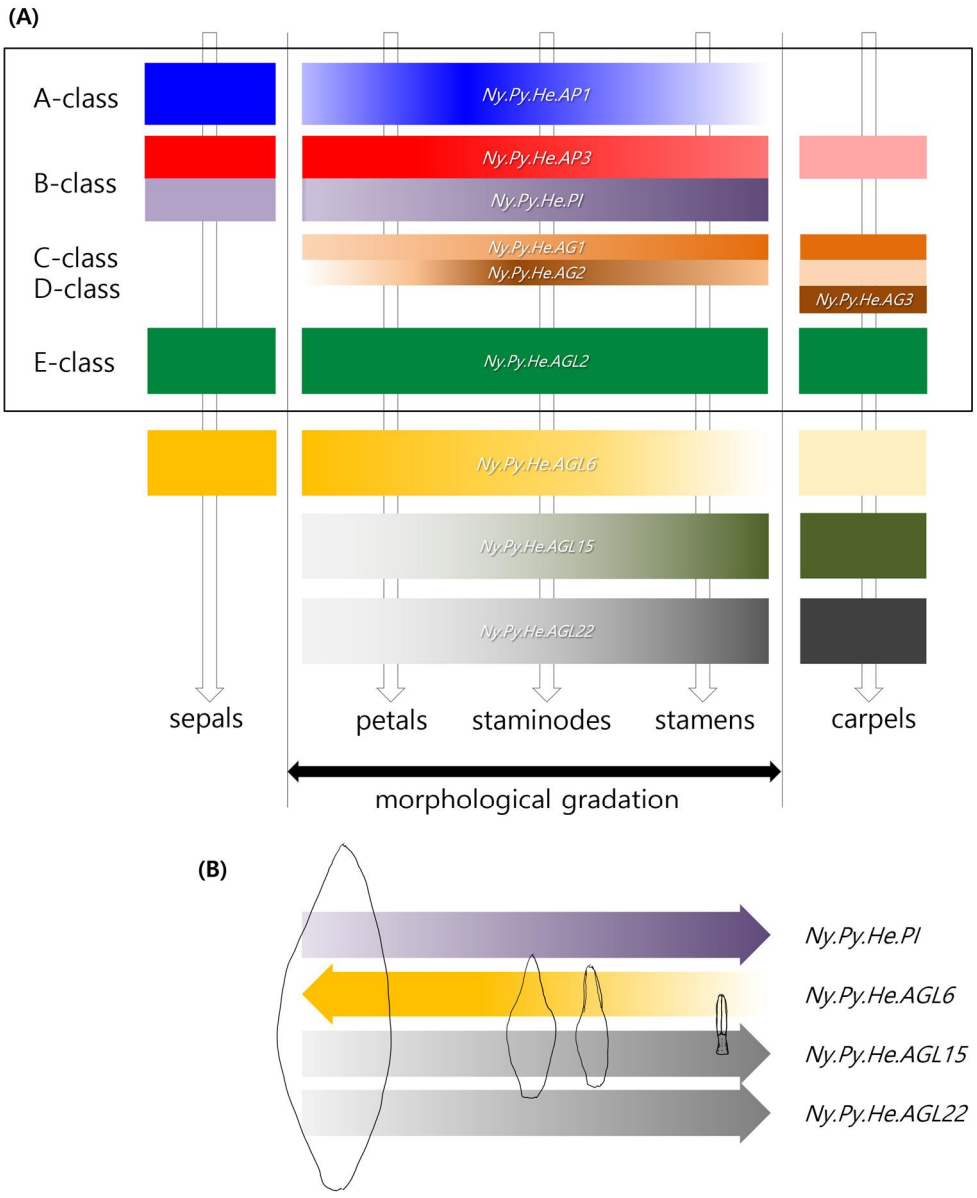

Figure 12. (A) summary of gene expressions involved in classic ABCDE model and additional genes showing expression gradations relevant to morphological gradation from petals to stamens. A box indicates genes involved in classic ABCDE model. (B) Increase/decrease directions of expressions of genes showing expression gradations.

the gene expression in *Nymphaea,* Yoo et al. (2010a) supported this idea and my data also confirmed it. Broad expression of B-class genes in *Amborella* was an example to explain the fading border model (Buzgo et al., 2005). In this study, two B-class genes (*Ny.Py.He.AP3* and *Ny.Py.He.PI*) showed broad expressions including sepals (Fig. 12). The result was well matched with those in *Amborella*. Furthermore, the expressional gradation of *Ny.Py.He.PI* is well matched with a morphological transition from petals to stamen. Expressions of C-class gene in both *Arabidopsis* and *Amborella* are restricted to carpels (Coen and Mayerowitz, 1991; Kim et al., 2005). However, the expression of the C-class genes (*Ny.Py.He.AG1* and *Ny.Py.He.AG2*) is extended to petals in *Nymphaea* (Fig. 12). In the ABCDE models, D-class genes in *Arabidopsis* (*STK*, *SHP1* and *2*) are related to specifying ovule or female gametophyte identities (Pelaz et al., 2000). Expression of D-class gene(*Ny.Py.He.AG3*), and E-class genes (*Ny.Py.He.SEP1*) showed the same expressions as those in *Arabidopsis.* The *GGM13*-like genes are a sister to both of B-class genes and called as B sister (B_S) (Becker et al. 2002; Becker and Theisen, 2003). In contrast to B genes, B_S genes are mainly expressed in female reproductive organs, especially in ovules (Becker and Theisen, 2003). *GGM13* homolog in *Nymphaea*

(*Ny.Py.He.Bs*) is also detected in this study and it has a similar expression with *Ny.Py.He.AG3* (Fig. 11).

As non-ABCDE type-II MADS-box genes, *Ny.Py.He.AGL6* and *Ny.Py.He.AGL15/Ny.Py.He.SVP* showed positive and negative expressional gradations from petals to stamen, respectively, indicating that these genes potentially involved in morphological transitions from petals to stamens in *Nymphaea*. In *Nymphaea* X 'Pyamaea Helvola', the expression pattern of *Ny.Py.He.AGL6* is also remarkable. This is strongly expressed in sepals, petals 1, and petals 2. In *Arabidopsis*, *AGL6* is not a floral organ identity gene and was expressed in perianth and ovules (Schauer et al. 2009; Yoo et al.2010a). In other *Nymphaea*, *AGL6*-homologs are expressed in sepals, petals, staminodes, and ovules of the early developmental stage (Yoo et al. 2010a).
These gene expression patterns support that *AGL6* genes play an essential role in perianth and ovule development in *Nymphaea*.
Also, some *AGL6* homologs from basal angiosperms are expressed only in perianth of *Liriodendron* (Kim et al. 2005) and *Persea* (Chanderbali et al. 2006) indicating that *AGL6* homologs may be especially likely candidates for genes involved in perianth development in basal angiosperms (Yoo et al. 2010a).

In summary, expressions of A−, B−, and C−class genes of *Nymphaea* X 'Pygmaea Helvola' are broader than those in eudicots (*Arabidopsis*). However, expressions of D−, E−class and Bs genes are the same as those reported from eudicots. Therefore, the overall point of view, 'fading border model' is applicable to explain the morphology of floral parts in *Nymphaea* X 'Pygmaea Helvola'.

Confirmation of transcript sequences and isoforms

The result of this study is from RNA−Seq which produce many isoforms in each gene. Because of the possible assembly error to make chimeric sequences the confirmation of identified genes by Sanger sequencing using cDNA and gene−specific primers is needed. Although only six genes (*Ny.Py.He.PI*, *Ny.Py.He.AG1*, *Ny.Py.He.AGL2*, *Ny.Py.He.AGL6*, *Ny.Py.He.AGL22,* and *Ny.Py.He.Bs*) were confirmed by Sanger sequencing in this study (Table 4 and 5), cDNA sequencing of all other type−II genes is needed eventually.

Although I could not include interpretation of differential expressions of isoforms in each gene in this study, the confirmation of detected isotypes in each gene may key to interpret their expression

level. To clarify the result in the future, confirmation of genome sequences to reduce the number of isotypes in each gene is needed. Also, the detailed investigation of expressions for the alternatively spliced form is needed.

References

Alvarez-Buylla, E. R., S. Pelaz, S. J. Liljegren, S. E. Gold, C. Burgeff, G. S. Ditta, L. R. De Pouplana, L. Martínez-Castilla, and M. F. Yanofsky (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences* 97: 5328-5333.

Babicki, S., D. Arndt, A. Marcu, Y. Liang, J. R. Grant, A. Maciejewski, and D. S. Wishart (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* 44: W147-W153.

Becker, A. and G. Theißen (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular phylogenetics and evolution* 29: 464-489.

Bolger, A. M., M. Lohse, and B. Usadel (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.

Borsch, T., C. Löhne, and J. Wiersema (2008). Phylogeny and evolutionary patterns in Nymphaeales: integrating genes, genomes, and morphology. *Taxon* 57: 1052-1052.

Buzgo, M., P. S. Soltis, S. Kim, and D. E. Soltis (2005). The making of the flower. *Biologist* 52.

Buzgo, M., P. S. Soltis, and D. E. Soltis (2004). Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *International Journal of Plant Sciences* 165: 925-947.

Coen, E. S. and E. M. Meyerowitz (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353: 31.

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

Endress, P. K. (2001). The flowers in extant basal angiosperms and inferences on ancestral flowers. *International Journal of Plant Sciences* 162: 1111-1140.

Favaro, R., A. Pinyopich, R. Battaglia, M. Kooiker, L. Borghi, G. Ditta, M. F. Yanofsky, M. M. Kater, and L. Colombo (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant Cell* 15: 2603-2611.

Feng, C., M. Chen, C.-j. Xu, L. Bai, X.-r. Yin, X. Li, A. C. Allan, I. B. Ferguson, and K.-s. Chen (2012). Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC genomics* 13: 19.

Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, and S. Sykes (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* 108: 1513-1518.

Henschel, K., R. Kofuji, M. Hasebe, H. Saedler, T. Münster, and G. Theißen (2002). Two ancient classes of MIKC-type MADS-box genes are

present in the moss *Physcomitrella patens*. *Molecular Biology and Evolution 19*: 801-814.

Honma, T. and K. Goto (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409: 525.

Initiative, A. G. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796.

Irish, V. F. (2003). The evolution of floral homeotic gene function. *Bioessays* 25: 637-646.

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, and S. Hunter (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236-1240.

Judd, W., C. S. Campbell, E. Kellogg, and P. Stevens (1999). Plant systematics. A phylogenetic approach. *Sinauer Associates*, Sunderland, Mass., USA 464: 3-4.

Kalyaanamoorthy, S., B. Q. Minh, T. K. Wong, A. von Haeseler, and L. S. Jermiin (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* 14: 587.

Kim, S., J. Koh, H. Ma, Y. Hu, P. K. Endress, B. A. Hauser, M. Buzgo, P. S. Soltis, and D. E. Soltis (2005a). Sequence and expression studies of A-, B-, and E-class MADS-box homologues in *Eupomatia* (Eupomatiaceae): support for the bracteate origin of the calyptra. *International Journal of Plant Sciences* 166: 185-198.

Kim, S., J. Koh, M. J. Yoo, H. Kong, Y. Hu, H. Ma, P. S. Soltis, and D. E. Soltis (2005b). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *Plant J* 43: 724-744.

Kramer, E. M., V. S. Di Stilio, and P. M. Schlüter (2003). Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. *International Journal of Plant Sciences* 164: 1-11.

Kramer, E. M., R. L. Dorit, and V. F. Irish (1998). Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* 149: 765-783.

Kumar, S., G. Stecher, and K. Tamura (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33: 1870-1874.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10: R25.

Loehne, C., T. Borsch, and J. H. Wiersema (2007). Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers. *Botanical Journal of the Linnean Society* 154: 141-163.

Luo, H., S. Chen, J. Jiang, Y. Chen, F. Chen, N. Teng, D. Yin and, C. Huang (2011). The expression of floral organ identity genes in contrasting water lily cultivars. *Plant cell reports* 30: 1909-1918.

Mathews, S. and M. J. Donoghue (1999). The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947-950.

Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences* 104: 19363-19368.

Münster, T., J. Pahnke, A. Di Rosa, J. T. Kim, W. Martin, H. Saedler, and G. Theissen (1997). Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proceedings of the National Academy of Sciences* 94: 2415-2420.

Nam, J., C. W. depamphilis, H. Ma, and M. Nei (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Molecular biology and evolution* 20: 1435-1447.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 32: 268-274.

Norman, C., M. Runswick, R. Pollock, and R. Treisman (1988). Isolation and properties of cDNA clones encoding *SRF*, a transcription factor that binds to the *c-fos* serum response element. *Cell* 55: 989-1003.

Pařenicová, L., S. de Folter, M. Kieffer, D. S. Horner, C. Favalli, J. Busscher, H. E. Cook, R. M. Ingram, M. M. Kater, and B. Davies (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *The Plant Cell* 15: 1538-1551.

Passmore, S., R. Elble, and B.-K. Tye (1989). A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. *Genes & development* 3: 921-935.

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.

Qiu, Y.-L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, E. A. Zimmer, Z. Chen, V. Savolainen, and M. W. Chase (1999). The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404.

Schneider, E. L., S. C. Tucker, and P. S. Williamson (2003). Floral development in the Nymphaeales. *International Journal of Plant Sciences* 164: S279-S292.

Shore, P. and A. D. Sharrocks (1995). The MADS-box family of transcription factors. *The FEBS Journal* 229: 1-13.

Smaczniak, C., R. G. Immink, G. C. Angenent, and K. Kaufmann (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139: 3081-3098.

Soltis, D. E., H. Ma, M. W. Frohlich, P. S. Soltis, V. A. Albert, D. G. Oppenheimer, N. S. Altman, C. dePamphilis, and J. Leebens-Mack (2007). The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression. *Trends Plant Sci* 12: 358-367.

Soltis, D. E., P. S. Soltis, V. A. Albert, D. G. Oppenheimer, H. Ma, M. W. Frohlich, G. Theißen, and F. G. P. R. Group (2002). Missing links: the

genetic architecture of flower and floral diversification. *Trends in plant science* 7: 22-31.

Sommer, H., J.-P. Beltran, P. Huijser, H. Pape, W. Lönnig, H. Saedler, and Z. Schwarz-Sommer (1990). *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *The EMBO Journal* 9: 605-613.

Theissen, G., A. Becker, A. Di Rosa, A. Kanno, J. T. Kim, T. Münster, K.-U. Winter, and H. Saedler (2000). A short history of MADS-box genes in plants. *Plant Molecular Evolution*, Springer: 115-149.

Theissen, G. and H. Saedler (2001). Plant biology: floral quartets. *Nature* 409: 469.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* 25: 4876-4882.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. Van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28: 511.

van Tunen, A. J., W. Eikelboom, and G. C. Angenent (1993). Floral organogenesis in *Tulipa*. *Flowering Newsletter*: 33-38.

Vining, K. J., E. Romanel, R. C. Jones, A. Klocko, M. Alves-Ferreira, C. A. Hefer, V. Amarasinghe, P. Dharmawardhana, S. Naithani, and M. Ranik (2015). The floral transcriptome of *Eucalyptus grandis*. *New Phytologist* 206: 1406–1422.

Wang, L., Z. Lu, W. Li, J. Xu, K. Luo, W. Lu, L. Zhang, and B. Jin (2016). Global comparative analysis of expressed genes in ovules and leaves of *Ginkgo biloba* L. *Tree genetics & genomes* 12: 29.

Wang, Z., M. Gerstein, and M. Snyder (2009). RNA–Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57–63.

Xie, Y., G. Wu, J. Tang, R. Luo, J. Patterson, S. Liu, W. Huang, G. He, S. Gu, S. Li, X. Zhou, T. W. Lam, Y. Li, X. Xu, G. K. Wong and J. Wang (2014). "SOAPdenovo–Trans: *de novo* transcriptome assembly with short RNA–Seq reads." *Bioinformatics* 30: 1660–1666.

Yanofsky, M. F., H. Ma, J. L. Bowman, G. N. Drews, K. A. Feldmann, and E. M. Meyerowitz (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346: 35.

Yoo, M. J., A. S. Chanderbali, N. S. Altman, P. S. Soltis, and D. E. Soltis (2010b). Evolutionary trends in the floral transcriptome: insights from one of the basalmost angiosperms, the water lily *Nuphar advena* (Nymphaeaceae). *The Plant Journal* 64: 687–698.

Yoo, M. J., Pamela S. Soltis, and Douglas E. Soltis (2010a). Expression of Floral MADS-Box Genes in Two Divergent Water Lilies: Nymphaeales andNelumbo. *International Journal of Plant Sciences* 171: 121–146.

Zeng, L., Q. Zhang, R. Sun, H. Kong, N. Zhang and H. Ma (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature communications* 5: 4956.

국문초록

기저피자식물들 중 한 분류군인 수련과는 수련목에 속하며, 이는 전체 피자식물 중 암보렐라를 제외한 전체 피자식물의 자매 그룹으로 피자식물 진화연구에서 중요한 역할을 해 오고 있다. 수련과는 수생 식물로써 5 속(*Nymphaea, Nuphar, Euryale, Barclaya, Victoria*) 약 80 종을 갖고 있고, 전 세계의 온대와 열대 지역에 분포한다. 애기장대와 같은 전형적인 진정쌍자엽식물들과는 달리 수련과 식물들의 꽃은 꽃잎에서 수술로 이어지는 일련의 형태적 전이가 발견된다. 이러한 수련속 식물의 특수한 형태는 기저피자식물들에 있어서 꽃발생유전자들이 형태형성에 기하는 역할을 연구하기 위한 매우 좋은 모델이 된다. 전사인자인 MADS-box 유전자들은 식물에 있어서 발생과정과 꽃의 기관형성에 관여한다. MADS-box 유전자들의 결정과 이들의 기능에 대한 연구로서 애기장대에서 꽃의 각 기관형성을 결정하는 ABCDE 모델이 제시된 바 있다. 기저피자식물 꽃들의 형태를 설명하기 위해 fading border 모델이 제시된 바 있지만 기저피자식물들의 꽃발생 유전자들에 대한 이해는 아직 완전하지 못하다. 이에 기저피자식물군의 한 분류군으로서 수련속 식물에 대한 꽃의 기관형성에 대한 유전적 조절에 대한 이해는 피자식물의 꽃의 진화와 다양화를 이해하는 중요한 열쇠가 된다. 본 연구에서는 꽃잎으로부터 수술로의 형태적 전이를 보여주는 *Nymphaea* X 'Pygmaea Helvola'에 대하여 각각의 꽃 기관으로부터

얻은 전사체를 바탕으로 꽃발생에 관여하는 MADS-box 유전자들의 발현양상을 조사하였다. 그 결과 *Arabidopsis* 와 *Amborella* 에서의 type-II 유전자들과 상동인 14 개의 유전자들을 검출하였고, 계통분석에 의해 이들의 상동성을 확인하였다. *Nymphaea* X 'Pygmaea Helvola'에서의 ABCDE 유전자들의 발현은 *Arabidopsis* 와는 매우 상이하였다. A-class 유전자(*Ny.Py.He.AG1*)는 심피를 제외한 모든 꽃 기관에서 발현하였다. 두 개의 B-class 유전자들(*Ny.Py.He.AP3* 와 *Ny.Py.He.PI*)은 그 발현이 꽃받침에까지, 그리고 C-class 유전자들 (*Ny.Py.He.AG1* 와 *Ny.Py.He.AG2*)은 꽃잎에까지 확장되었다. D-class (*Ny.Py.He.AG3*)와 E-class (*Ny.Py.He.SEP1*) 유전자들의 발현은 *Arabidopsis* 에서의 발현과 일치하였다. ABCDE 유전자들 중 *Ny.Py.He.PI* 는 꽃잎에서 수술로 점차 변하는 형태적 변이를 반영하여 발현량이 변화 하였다. ABCDE 유전자가 아닌 type-II MADS-box 유전자들 중 *Ny.Py.He.AGL6* 와 *Ny.Py.He.AGL15 / Ny.Py.He.SVP* 는 각각 꽃잎으로부터 수술로 각각 증가 또는 감소하는 발현 구배를 보였는데, 이는 이들 유전자가 수련속 식물에서 발견되는 꽃잎에서 수술로의 형태적 전이에 관여하는 유전자일 가능성을 제시하고 있다.