



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

농학석사 학위논문

RNA-seq을 이용한 소나무 (*Pinus densiflora*)의 구과 발달 과정에서 전사체 분석 및 개화 조절 유전자 탐색

Transcriptome analysis and identification of flowering regulatory genes using RNA-seq in the strobilus development of *Pinus densiflora*

2023년 2월

서울대학교 대학원

농림생물자원학부 산림환경학 전공

이 다 영

RNA-seq을 이용한 소나무 (*Pinus densiflora*)의 구과 발달 과정에서
전사체 분석 및 개화 조절 유전자 탐색

지도교수 강 규 석

이 논문을 농학석사 학위논문으로 제출함
2022년 12월

서울대학교 대학원
농림생물자원학부 산림환경학 전공
이 다 영

이다영의 농학석사 학위논문을 인준함
2023년 1월

위 원 장 _____ 김 현 석 _____ (인)

부위원장 _____ 강 규 석 _____ (인)

위 원 _____ 심 상 래 _____ (인)

초 록

종자식물에게 개화 현상은 생존과 번식을 위해 필수적인 과정이며, 영양생장에서 생식생장으로의 전환 과정에는 다양한 유전적 변화가 발생한다. 다양한 개화 조절 유전자에 대한 연구가 이루어진 속씨식물과 비교했을 때, 겉씨식물인 침엽수의 경우에는 해당 연구가 부족한 상황이다. 침엽수에서 개화 조절 유전자의 발현을 증가시킨다면 구과 생산량이 증가할 것이며, 조기 개화를 유도한다면 육종 속도를 가속화함으로써 이른 차대검정이 가능할 것이다. 본 연구에서는 소나무(*Pinus densiflora*)의 생식기관별로 RNA를 추출한 뒤, RNA-seq 데이터를 이용하여 소나무의 표준전사체를 생성하였다. 그리고 기관별로 전사체를 비교 분석하여 소나무 암구과 및 수구과 발달의 생물학적 및 생리적 특성과 개화 조절 유전자를 탐색하였다. 본 연구결과에서 7종류(*Pinus densiflora* *APETALA2-LIKE*(PdAP2L), *DEFICIENTS AGAMOUS-LIKE*(PdDAL), *MINICHROMOSOME MAINTENANCE1 AGAMOUS DEFICIENS SERUM RESPONSE FACTOR 1*(PdMADS1), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1*(PdSOC1), *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3*(PdSPL3), *UNUSUAL FLORAL ORGANS*(PdUFO), *WUSCHE*(PdWUS))의 후보 개화 조절 유전자를 발견하였으며, 해당 유전자들은 정단과 암구과뿐만 아니라 다양한 영양기관에서도 발현하였다. 추후 후보 유전자들의 정확한 서열을 추출하여 형질전환을 통해 그 기능을 검증할 수 있을 것이다. 최종적으로 소나무의 다양한 개화 조절 유전자들을 공개 데이터베이스에 등록 가능할 것이라 기대한다.

주요어 : RNA-seq, 소나무, 구과 발달, 전사체, 개화 조절 유전자
학 번 : 2021-28445

목 차

| | |
|---|----|
| 제1장 서론 | 1 |
| 제1절 연구의 배경 | 1 |
| 제2절 연구의 목적 | 4 |
| 제2장 연구사 | 5 |
| 제1절 소나무 구과 발달 특성 연구 | 5 |
| 제2절 식물 전사체 발현 및 시퀀싱 연구 | 7 |
| 제3절 종자식물의 개화유전자 연구 | 9 |
| 제3장 재료 및 방법 | 12 |
| 제1절 실험 재료 및 샘플 채취 | 12 |
| 제2절 RNA 추출, cDNA 라이브러리 제작 및 Illumina 시퀀싱 .. | 14 |
| 제3절 소나무의 표준전사체 생성 | 16 |
| 제4절 발현량 정량화 및 DEG 분석 | 18 |
| 제5절 소나무 기관 비교별 GO enrichment 분석 | 19 |
| 제6절 상동성 기반의 개화 조절 유전자 탐색 | 20 |
| 제7절 qRT-PCR과 NGS 데이터의 상관분석 | 21 |
| 제4장 연구결과 | 24 |
| 제1절 소나무 전사체의 시퀀싱과 표준전사체 생성 | 24 |
| 제2절 소나무 기관들의 유전자 차등 발현 | 27 |
| 제3절 소나무 기관별 분자세포생물학적 특성 비교 | 35 |
| 제4절 소나무의 후보 개화 조절 유전자 탐색 및 기관별 비교 .. | 51 |
| 제5절 DEG 발현량에 대한 qRT-PCR 검증 | 58 |
| 제5장 고찰 | 59 |

| | |
|--------------------------------------|----|
| 제1절 소나무 구과 발달 과정의 생물학적 및 생리적 특성..... | 59 |
| 제2절 소나무 개화 조절 유전자의 발현 특성 | 66 |
| | |
| 제6장 결 론 | 71 |
| | |
| 참고 문헌..... | 72 |
| | |
| Appendix | 87 |
| | |
| Abstract..... | 93 |

표 목차

| | |
|---|----|
| [Table 1] Information on the sampling of <i>Pinus densiflora</i> . | 12 |
| [Table 2] Oligonucleotide primers used for validation by qRT-PCR | 23 |
| [Table 3] Statistics of raw sequencing data..... | 25 |
| [Table 4] Statistics of transcripts and genes from <i>de novo</i> assembly in <i>Pinus densiflora</i> | 26 |
| [Table 5] BLASTX best-hit results of candidate flowering regulatory genes of <i>Pinus densiflora</i> | 53 |

그림 목차

| | |
|---|----|
| [Figure 1] Structure of vegetative and reproductive organs of <i>Pinus densiflora</i> | 13 |
| [Figure 2] Process of the preparation of cDNA library | 15 |
| [Figure 3] Process of cluster generation..... | 15 |
| [Figure 4] Example of FASTQ file format..... | 16 |
| [Figure 5] PCA plots (A–B) and Pearson correlation matrix (C) of all samples..... | 28 |
| [Figure 6] Number of DEGs between reproductive organ samples of <i>Pinus densiflora</i> | 29 |
| [Figure 7] Heatmap of DEGs and samples used in pairwise comparisons | 30 |
| [Figure 8] MA plot and Volcano plot of each pairwise comparison | 31 |
| [Figure 9] Heatmap of DEGs on each comparison | 32 |
| [Figure 10] The number of DEGs in each comparison..... | 34 |
| [Figure 11] Top 30 GO terms in Gene Ontology analysis for A vs F1 comparison, A enriched..... | 37 |
| [Figure 12] Top 30 GO terms in Gene Ontology analysis for A vs F1 comparison, F1 enriched | 38 |
| [Figure 13] Top 30 GO terms in Gene Ontology analysis for F1 vs F3 comparison, F1 enriched..... | 41 |
| [Figure 14] Top 30 GO terms in Gene Ontology analysis for F1 vs F3 comparison, F3 enriched..... | 42 |
| [Figure 15] Top 30 GO terms in Gene Ontology analysis for | |

| | |
|--|----|
| F1 vs M1 comparison, F1 enriched | 45 |
| [Figure 16] Top 30 GO terms in Gene Ontology analysis for F1 vs M1 comparison, M1 enriched | 46 |
| [Figure 17] Top 30 GO terms in Gene Ontology analysis for M1 vs M2 comparison, M1 enriched | 49 |
| [Figure 18] Top 30 GO terms in Gene Ontology analysis for M1 vs M2 comparison, M2 enriched | 50 |
| [Figure 19] Expression of candidate flowering regulatory genes between reproductive organs of <i>Pinus densiflora</i> | 55 |
| [Figure 20] PCA plot with expression values of the candidate flowering regulatory genes of <i>Pinus densiflora</i> | 57 |
| [Figure 21] Validation of RNA–sequencing results by qRT– PCR | 58 |

제1장 서 론

제1절 연구의 배경

종자식물에게 있어 영양생장에서 생식생장으로의 전환은 생존을 위해 필수적인 단계이다. 속씨식물의 경우에는 영양생장에서 생식생장으로 전환되는 과정에 대한 유전적인 연구가 많이 진행되었지만, 겉씨식물인 다년생 침엽수에서는 상대적으로 이 과정이 거의 밝혀지지 않았다(Ma et al, 2021). 애기장대(*Arabidopsis thaliana*)와 같은 모델식물에서는 *LEAFY(LFY)*와 *MINICHROMOSOME MAINTENANCE1 AGAMOUS DEFICIENS SERUM RESPONSE FACTOR(MADS) -box* 등 꽃 발달에 관여하는 대표적인 개화 조절 유전자들의 기능과 특성에 대해 많은 연구가 이루어졌다(Ng and Yanofsky, 2001; Moyroud et al., 2010). 하지만 겉씨식물과 특히 침엽수인 소나무속(*Pinus*)에 있어서 개화 조절 유전자의 기능이나 그 발현 양상에 대한 연구는 부족한 상황이다.

임목은 일반적으로 긴 육종 기간을 가지며 선발 육종, 교잡 육종 등 전통적인 육종 방법에 의해 새로운 품종을 개발하는 것은 수십년이 필요하다. 이러한 긴 육종 기간은 다년생 임목의 경우 생식생장을 시작하기 전 긴 영양생장 기간을 가지기 때문이다. 이처럼 느리게 성장하는 임목은 현대 사회에서 빠르게 변화하는 요구에 충분히 대응하기가 힘들며, 결과적으로 숲 전체의 소실과 파괴를 야기할 수 있다(Fenning and Gershenzon, 2002).

지금까지 국내에서는 소나무(*Pinus desiflora* Sieb. & Zucc.)의 생식생장에 대한 연구가 주로 채종원의 클론 집단을 기반으로 한 구과와

종자의 물리적 특성과 생산량 변이, 종자의 발아 특성 등을 중심으로 이루어졌다. 이후 동위효소 또는 Simple Sequence Repeat(SSR)와 같은 유전적 마커가 개발됨에 따라 교배 양식이나 유전다양성 분석 등이 가능하게 되었다(Kang and Lindgren, 1998; Lee et al., 2003; Chung et al., 2019). 현대 기술이 더욱 발전함에 따라 형질 전환 또는 유전자 변형 등 생명공학기술을 통해 생식 구조가 변형된 임목 개체를 개발하는 것에 대한 필요성이 대두되고 있다(Fenning and Gershenzon, 2002).

한편 임목은 매년 상당한 양의 에너지와 양분을 생식생장 과정에 사용한다. 침엽수에서 구과 생산과 영양 생장은 반비례 관계로 나타났으며 라디아타소나무(*Pinus radiata*)의 경우 수구과 생산이 약 13%의 영양 생장을 감소시킨다고 보고된다(El-Kassaby and Barclay, 1991; Cremer, 1992; Strauss et al., 1995). 따라서 구과를 생산하지 않는 개체를 개발한다면 생식 생장에 사용되는 에너지와 양분을 영양생장에 사용함으로써 임분의 목재 생산성 증가에 도움을 줄 수 있을 것이다.

반대로, 유전자 편집기술을 통해 개화에 관여하는 유전자들의 발현을 조절하여 조기 개화를 유도한다면 육종 속도를 가속화할 수 있다. 우수한 유전자형을 선발하여 조성된 클론 집단에서 조기 개화를 유도한다면 이른 차대검정이 가능할 것이며, 대부분의 자연 임분에서 발생할 수 있는 결실 불균형을 해결할 수 있을 것이다(Bonnet-Masimbert, 1987).

침엽수는 긴 생식생장 기간과 크고 복잡한 유전체를 가지며, 신뢰할 수 있는 표준유전체(reference genome)가 부족하기 때문에 침엽수의 유전자 기능을 이해하는 것은 대단히 어려운 편이다. 또한 겉씨식물에서 돌연변이가 발생하는 경우가 부족하기 때문에 많은 유전자들의 기능은 밝혀지지 않은 상황이다(Moyroud et al, 2010). 소나무속에서 전장

유전체(whole genome)가 분석된 수종은 매우 적고, 소나무속의 유전체 크기는 상당히 큰 편에 속한다. 따라서 침엽수의 개화와 같은 복잡한 현상을 유전적으로 이해하기 위한 연구는 제한적으로 수행되었다.

소나무속의 구과 발달에 대해서 전사체 분석을 통해 유전적인 관점에서 해석한 연구는 국외에 일부 존재하지만(Niu et al., 2016; Feng et al., 2018), 소나무의 경우에는 최근 뉴질랜드에서 진행된 하나의 선행 연구를 제외하고는 연구된 사례가 없다(Fritsche et al., 2022). 해당 선행 연구에서는 *MADS-box* type II 유전자 등 소나무의 초기 구과 발달에 직접적으로 관여하는 유전자들에 초점을 맞추었다. 하지만 식물의 개화 현상은 특정한 개화 유전자들만 발현이 되어 나타나는 것이 아니라 다양한 식물 호르몬과 대사 작용 등 수많은 세포 활동과 환경적 영향이 함께 포함된 결과로 나타난다. 따라서 *LFY*, *MADS-box* 유전자 등 모델식물에서 다수 연구된 개화 조절 유전자뿐만 아니라 소나무 구과 발달 과정에서 일어나는 전반적인 유전생리적 현상에 대해서도 이해할 필요가 있다.

제2절 연구의 목적

본 연구에서는 소나무의 구과 발달 과정에서 총 6개 기관인 정단, 암구과, 수구과, 잎, 줄기, 뿌리의 전사체 발현을 비교하였다. 특히 암구과와 수구과에서는 발달 단계에 따라 유전적 양상이 달라질 수 있으므로 시기별로 비교하여 발현되는 유전자들의 기능을 탐색하였다. 따라서 본 연구의 목적은 RNA-seq 데이터를 사용해서 소나무의 표준전사체(reference transcriptome)를 생성한 뒤 기관별 전사체 발현의 특성을 분석하고자 한다. 그리고 상동성(homology)을 기반으로 하여 소나무 구과 발달에 관여하는 유전자들의 기능을 탐색하고 기관별 발현을 비교하고자 한다. 마지막으로 생물정보학적 데이터 결과와 실제 생물학적 시료에서의 결과를 비교하여 본 연구의 유전자 발현 결과를 검증하고자 한다.

소나무 성목을 대상으로 기관별 전사체 분석을 진행함으로써 소나무의 구과 발달 과정에서 발생하는 전체적인 생물학적 및 생리적 현상을 이해할 수 있을 것이다. 게다가 직접적으로 소나무 개화에 관여하는 유전자들의 기능을 분석하여 소나무의 후보 개화 조절 유전자를 탐색할 수 있을 것이다. 본 연구를 통해 조기개화 개체를 선발하기 위한 마커 개발이나 구과 생산량이 우수한 소나무의 품종 개발을 위해 필요한 정보를 얻을 수 있을 것이라 생각한다.

제2장 연구사

제1절 소나무 구과 발달 특성 연구

소나무(*P. densiflora*)는 건조하고 척박한 환경에 잘 적응해 넓은 범위에서 생존이 가능하며, 국내에서 단일 수종으로서 가장 넓은 면적을 차지하고 있다. 그리고 소나무는 우리나라에서 조림에 사용된 중요한 자생 수종으로 우수한 개체의 클론을 보존하기 위해 넓은 면적의 채종원이 조성되어 있다(Kang, 2000). 꽃을 피우는 데 1~2년이 걸리는 속씨식물과 달리, 겉씨식물인 침엽수는 일반적으로 완전한 구과를 생산하기까지 3년이 걸린다(Singh, 1978). 첫째 해 여름에 꽃눈이 형성되고 둘째 해 봄에 수분 현상이 일어나며, 셋째 해 가을에 종자가 성숙함으로써 2년에 걸쳐 구과 종자가 성숙한다. 이렇듯 침엽수는 한번의 생식 활동에 긴 시간을 사용하며 환경 조건에 민감하게 반응하면서 생식 구조를 형성한다(Houle and Fillion, 1993).

침엽수의 생활사는 크게 세 단계로 구분할 수 있는데, 각 단계마다 환경적 신호에 반응하는 정도가 변화한다. 첫 번째는 생식능력이 없는 단계이며 오직 영양성장만 가능하고 외부에서 생식 발달을 유도하는 조건이 형성되더라도 생식 구조를 형성하지 못한다. 두 번째는 생식 구조 형성이 가능한 단계인데, 가뭄 스트레스나 지베렐린의 외생 처리와 같은 강한 외적 신호가 존재할 때 생식생장이 가능하다. 마지막은 성숙 단계이며 환경 조건과 상관없이 규칙적으로 생식생장이 일어난다(Poethig, 1990; Greenwood, 1995; Bond, 2000; Williams, 2009). 각 단계별 전환은 수많은 요인에 의해 조절되는데, 특히 분열조직에서 식물 내부와 외부의 신호에 반응하며 복잡한 변화가

일어난다(Hackett, 1987, 1993; Greenwood, 1995). 분열조직내 유전자 발현의 변화로 인해 물리적 형태와 생리적 현상이 조절되는데, 이는 식물이 각 생활사 단계에 적응하기 위한 전략이라고 할 수 있다(Day and Greenwood, 2011; Wendling et al., 2014).

개화 현상은 임목 육종을 위한 필수적인 과정이며 생리학과 유전학 분야에서 침엽수의 개화 생리에 대한 다수의 연구가 이루어졌다(Bonnet-Masimbert, 1987). 임목의 개화 조절은 전통적인 재배법(cultural practice)인 수분 스트레스, 뿌리 전정, 환상 박피, 질소 비료, 온도 상승 등에 제한되어 있었다가(Jackson and Sweet, 1972), 1960년대에 이르러 측백나무과(Cupressaceae)와 낙우송과(Taxodiaceae)에서 지베렐린산(gibberellic acid)과 같은 성장조절 물질의 처리를 통해서도 개화 조절이 가능하다는 것이 처음 연구되었다(Kato et al., 1958; Pharis and Morf, 1967). 그 이후 소나무과(Pinaceae)에서도 지베렐린이 로지폴소나무(*Pinus contorta*)와 개솔송나무(*Pseudotsuga menziesii*)의 개화를 유도하는데 효과가 높다는 것이 밝혀졌다(Pharis, 1975a, 1967, 1975b).

제2절 식물 전사체 발현 및 시퀀싱 연구

생물학과 생명공학 등 여러 분야에서는 세포의 생리와 활성을 연구하기 위해 빠른 유전체 분석 기술이 필요했으며, 특히 전사체(transcriptome)에 대한 연구는 DNA 마이크로어레이(DNA microarray) 기술의 등장과 인간 유전체의 완전 해독 이후부터 널리 사용되기 시작하였다. DNA 마이크로어레이 기술이나 Real-Time(RT) PCR 기술로 mRNA 발현을 측정하기 시작했는데, 마이크로어레이 기술은 정교함이나 민감성이 부족하였고 RT-PCR 기술은 전장유전체(genome-wide)의 유전자 발현 연구에 사용할 수 없었다(Mardis, 2008). 반면에 빠르고 비용도 적게 드는 Next-Generation Sequencing(NGS) 기술은 고속대용량처리(high-throughput) 방식으로 유전자 발현뿐만 아니라 유전자의 구조와 기능을 분석할 수 있고, non-coding RNA 등 단백질로 번역되지 않는 비암호화 RNA의 정보를 제공한다. 이러한 NGS는 합성에 의한 시퀀싱, 즉 Sequencing-By-Synthesis(SBS, pyrosequencing) 기술을 기반으로 하며 짧은 서열을 대량으로 동시에 시퀀싱하는 방법인 RNA-seq이 대표적이다(Denoed et al., 2008). RNA-seq은 생물학적 시료에서 RNA 서열의 존재와 양을 밝히는 실험적, 정보학적 방법을 말하며 단일 가닥인 RNA 분자의 염기서열을 확인할 수 있다. 마이크로어레이와 같은 이전 방법들과의 차별성은 기존의 서열 정보가 없어도 전사체 발현의 정량화가 가능하며 다양한 분석 결과를 제공한다는 것이다(Mutz et al., 2013; Korpelainen, et al., 2015).

유전체 서열의 정보를 얻기 위한 시퀀싱 기술은 세대적인 변화를 겪었다. 1세대 염기서열 분석법은 일반적으로 Sanger 시퀀싱을 의미한다. Sanger 시퀀싱은 효소의 반응을 이용하는 방법으로

1970년대 Friedrich Sanger에 의해 처음 만들어졌다. 임의적인 짧은 프라이머를 사용하여 DNA 중합효소가 상보적 가닥을 연장하는 것으로 진행된다. 이때 각각 다른 색을 띄는 형광물질로 라벨링된 네 가지의 ddNTP 중 하나가 부착하면 DNA 신장이 중단되며, 이를 통해 해당 DNA 서열을 확인할 수 있다(Sanger et al., 1977; Korpelainen, et al., 2015). NGS라고 불리는 차세대 염기서열 분석법도 염기서열의 합성 화학반응에 의해 시퀀싱이 진행되지만, 동시에 대량으로 시퀀싱이 가능해서 한 번에 수백만 개의 시퀀싱 반응이 가능하다. 서로 다른 색을 띄는 형광물질이 결합된 네 개의 뉴클레오타이드가 서열에 따라 결합하면서 빛을 생성함으로써 서열 확인이 가능하다(Marsh, 2007). 최근 개발된 3세대 염기서열 분석법은 마찬가지로 동시에 대량으로 시퀀싱이 가능하고 합성 화학반응을 이용하지만, PCR 증폭 과정을 생략하고 긴 DNA나 RNA 단일 분자를 시퀀싱하는 방법이다. 한 번에 분석 가능한 시퀀싱 수는 더 적지만 반응에 사용가능한 서열의 길이는 더 길다는 장점이 있어 최신 연구에 활발히 사용되고 있다(Petersen et al., 2019).

제3절 종자식물의 개화유전자 연구

모델식물에서는 다양한 돌연변이를 통해 단백질과 유전자의 기능을 추측할 수 있는데, *FLORICAURA(FLO)*와 *LEAFY(LFY)*를 금어초속(*Antirrhinum*)과 애기장대에서 각각 분리한 것을 시작으로, 다양한 속씨식물에서 꽃 형성 초기 단계를 조절하는 특정한 전사인자들을 발견할 수 있었다. 이후 속씨식물과의 상동성과 계통적 분석을 통해 *LFY/FLO* homolog들이 겉씨식물과 비종자식물을 포함하여 많은 종에서 확인되었다(Moyroud et al., 2010). 식물 특이적 전사인자인 *LFY*는 꽃분열조직에서 전체적인 생식구조 형성을 유도하는 최상위 조절자로서 꽃기관 원시세포(floral organ primordia)의 유전자 발현을 조절한다(Blazques et al., 2006; Liu et al., 2009; Moyroud et al., 2009).

*LFY*는 식물의 꽃눈에서 균일하게 발현되기 때문에 특정 부분에서 특정한 유전자의 발현을 유도하기 위해서는 지역적 공동조절자(regional coregulators)가 필요하다. 애기장대에서 *SEPALLATA3(SEP3)*는 *LFY*의 공동활성인자(coactivator)로 작용하면서 직접적인 꽃기관 형성 유전자인 B(*APETALA3, AP3*)와 C(*AGAMOUS, AG*) 유전자의 발현을 유도한다(Liu et al., 2009a). 여기서 *AG*는 여러 추가적인 공동활성인자와 공동억제인자(corepressor)로 인해 꽃 중심 부위에서만 제한적으로 발현된다(Liu et al., 2009b; Irish, 2010). 또한 *AP3*를 활성화하기 위해서 *LFY*는 *UNUSUAL FLORAL ORGANS(UFO)*로 알려진 F-box 단백질에 부착해야 한다(Chae et al., 2008). 즉 속씨식물의 서로 다른 꽃 기관 형성은 여러 유전자들의 상호작용에 의해 조절되는데, 이를 ABC 모델이라고 하며 이후 ABCDE 모델로 확장되었다(Weigel and Meyerowitz, 1993; Theißen et al., 2016).

애기장대에서 A-class 유전자는 *APETALA1(API)*과 *AP2*, B-class 유전자는 *AP3*와 *PISTILLATA(P1)*, C-class 유전자는 *AG*, D-class 유전자는 *SEEDSTICK(STK)*, *SHATTERPROOF1(SHP1)*과 *SHP2*, E-class 유전자는 *SEPALLATA1,2,3,4(SEP1,2,3,4)*와 *AGAMOUS-LIKE 6(AGL6)*이 존재한다. E와 A 유전자의 중복적인 발현은 꽃받침을 유도하고, E와 B, A가 함께 발현하여 꽃잎을 유도한다. E와 B, C는 수술을 형성하고, E와 C는 암술을 형성한다. 그리고 E와 C, D가 동시에 발현하여 자방이 형성된다(Bowman et al., 1991; Favaro et al., 2003; Theiben et al., 2016).

MADS-box라는 이름은 효모의 *MINICHROMOSOME MAINTENANCE1(MCM1)*, 애기장대의 *AG*, 금어초속의 *DEFICIENS(DEF)*, 그리고 인간의 *SERUM RESPONSE FACTOR(SRF)*에서 유래하였다(Riechmann and Meyerowitz, 1997). 이 유전자들은 MADS-box라고 불리는 DNA 결합 도메인을 가지는 전사인자를 암호화하며, 이러한 전사인자들은 유사한 DNA 서열(consensus DNA sequence motif 또는 *cis-regulatory element*)을 인식한다. 유전자 연구가 가장 많이 이루어진 모델식물인 애기장대는 40개 이상의 *MADS-box* 유전자를 가지고 있으며, 해당 유전자들은 꽃 형성, 개화 시기 조절, 영양 생장 제한 등 다양한 기능을 가지고 있다(Ng and Yanofsky, 2001). *MADS-box* 유전자의 계통적 분석 연구에 따르면, 유전자 기능과 관련하여 이들을 특정한 subfamiliy(clades)로 구분할 수 있다고 한다(Theissen et al., 1996; Purugganan, 1997). 식물, 동물, 균류의 가장 최근의 공통 조상은 적어도 2개의 *MADS-box* 유전자 계통을 가졌으며, 분화 이후의 *MADS-box*는 특정한 두 계통(type I, type II)이 남아서 이어진 것으로 추정된다. 식물의 type II 계통에서는 K 도메인과 같은 특정한

구조가 존재하는데, 이들은 식물 계통이 동물-균류에서 분리된 이후 변화하여 생성된 것으로 보인다(Alvarez-Buylla et al., 2000).

겉씨식물에서도 B, C/D와 E 유전자 등 기관 형성을 조절하는 호메오 유전자(homeotic gene)의 ortholog들이 확인되었다(Mouradov et al., 1998; Gramzow and Theissen, 2010; Uddenberg et al., 2013; Chen et al., 2017). B와 C 유전자의 ortholog는 겉씨식물의 수구과와 암구과에서 발현하는데, 이를 통해 해당 유전자들이 겉씨식물의 생식구조 형성에 관여한다고 추정된다(Theissen and Becker, 2004). 게다가 다양한 *MADS-box* homolog가 겉씨식물의 분열조직에 생식적 특성을 부여한다는 것이 발견되었으나, 유전적 발현 양상은 속씨식물과 다르며 그 역할에 대해서는 아직 완전히 연구되지 않았다(Benlloch et al., 2007).

제3장 재료 및 방법

제1절 실험 재료 및 샘플 채취

경기도 수원시 서울대학교 칠보산 학술림에 식재된 성목 소나무(*P. densiflora*)를 실험 재료로 사용하였다. 2022년 3월에서 5월까지의 기간동안 소나무 3 개체를 대상으로 하였으며, 생식구조 형성이 시작되는 초기 정단부 생장에서부터 화분이 날릴 때까지를 포함하는 기간 동안 샘플 채취를 진행하였다(Table 1).

Table 1. Information on the sampling of *Pinus densiflora*

| Sample abbreviation | Sample name | Sampling date | Number of biological replications |
|----------------------------|-----------------------|----------------------|--|
| A | Apex | 2022.03.21 | 3 |
| F1 | Female strobili early | 2022.05.02 | 3 |
| F2 | Female strobili mid | 2022.05.09 | 3 |
| F3 | Female strobili late | 2022.05.16 | 3 |
| L | Needle | 2022.05.23 | 3 |
| M1 | Male strobili early | 2022.04.25 | 3 |
| M2 | Male strobili mid | 2022.05.02 | 3 |
| M3 | Male strobili late | 2022.05.09 | 3 |
| R | Root | 2022.05.16 | 3 |
| S | Stem | 2022.05.16 | 3 |

3월 말 초기 정단부 샘플을 채취한 뒤, 구과 생장이 시작되는 4월 말부터 일주일에 한 번씩 샘플을 채취하였다. 암구과는 수구과에 비해

발달이 늦기 때문에 수구과보다 일주일 늦게 채취를 시작하였으며, 암구과와 수구과 모두 총 3번의 발달 단계로 구분하였다(Figure 1). 그리고 암구과 발달이 끝날 때 추가로 생식구조가 없는 뿌리, 줄기, 잎을 채취하였다. 모든 샘플은 채취 후 액체질소에 급속 냉동하였으며, RNA 추출 전까지 -80°C 냉동고에 보관하였다.

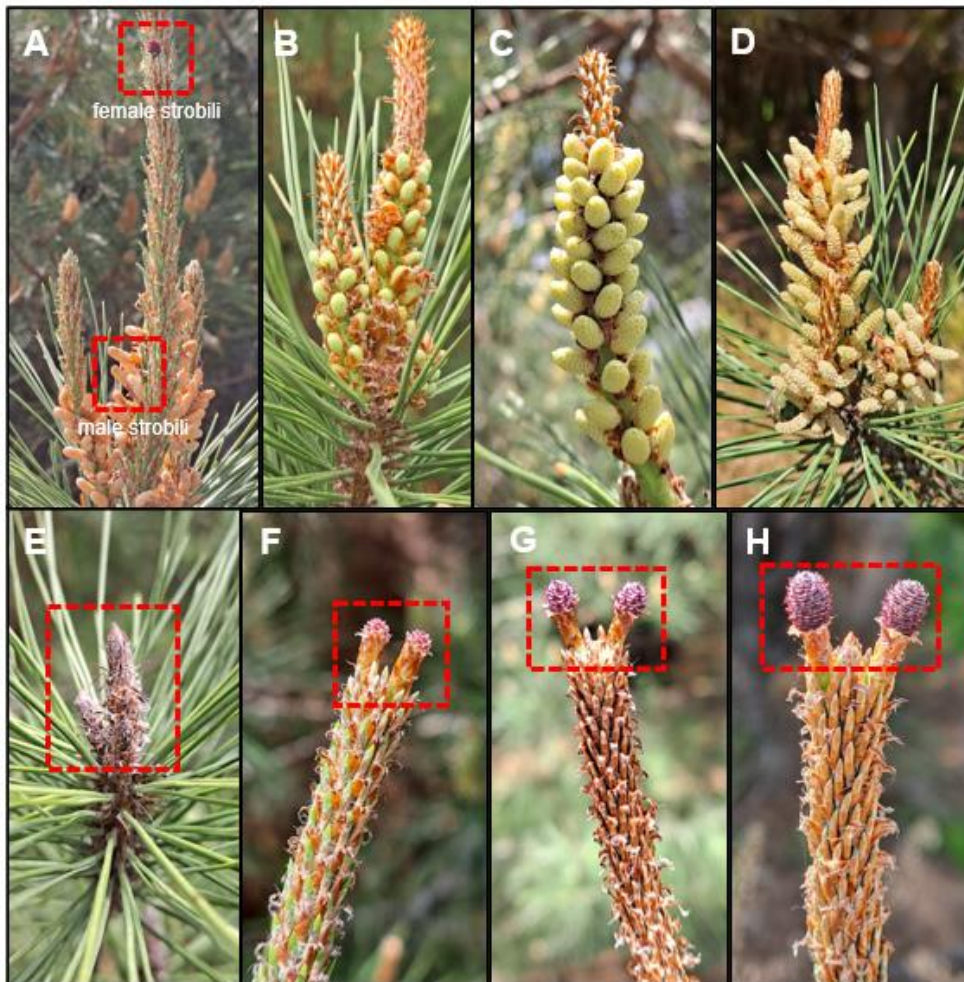


Figure 1. Structure of vegetative and reproductive organs of *Pinus densiflora*. A: position of female (upper) and male (lower) strobili on apex shoot, B-D: development of male strobili, E: apex shoot, F-H: development of female strobili

제2절 RNA 추출, cDNA 라이브러리 제작 및 Illumina 시퀀싱

RNA 추출을 위한 모든 시료는 TissueLyzer(QIAGEN)를 사용하여 각각 30회/초에서 1분 30초 2회로 설정하여 극저온 상태에서 분쇄하였다. 그리고 IQeasy™ plus Plant RNA Extraction Mini Kit (iNtRON)를 사용하여 제조사의 실험법에 따라 RNA를 추출하였다. 총 RNA의 양(quantity)과 순도(integrity)는 Agilent 4200 TapeStation을 사용하여 평가하였으며, 총 RNA 양이 500 ng 이상이고 RNA Integrity Number(RIN)가 8.0 이상인 것을 기준으로 최종 RNA 샘플을 선정하였다.

총 500 ng의 RNA를 사용하여 각 샘플별 cDNA 라이브러리를 제작하였다(**Figure 2**). 먼저 RNA를 단편화한 후 double-stranded cDNA(dsDNA)를 합성하였다. DNA 중합효소 등을 이용하여 단편화된 dsDNA의 말단 수선(end repair)을 진행하였으며, 5'-말단에 인산기와 3'-말단에 아데닌(Adenine, A) 염기를 추가하여 3'-overhang 형태로 만들어 어댑터(adapter) 서열과 상보적인 결합이 가능하게 하였다. 어댑터에는 프라이머 서열과 index 서열 등 이후 시퀀싱 작업에서 필요한 서열들이 포함되어 있는데, 이러한 어댑터 서열에 존재하는 3'-말단의 티민(Thymine, T) 잔기와 단편화된 DNA 말단의 아데닌 잔기가 결합하여 DNA와 어댑터를 연결하였다. 시퀀싱에 필요한 충분한 양의 라이브러리를 얻기 위해 PCR 증폭 과정을 진행했으며, 라이브러리 제작 과정에서 생산된 부산물은 자성 비드를 사용하여 제거하였다(LabGenomics).

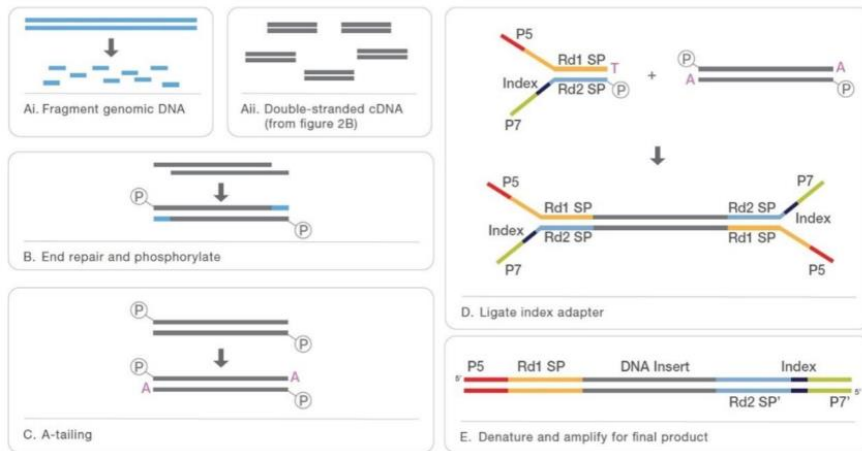


Figure 2. Process of the preparation of cDNA library (LabGenomics)

PCR로 증폭된 DNA를 단일 가닥으로 만들어 플로우 셀(flow cell) 표면에 고정시키고, 어댑터 서열과 상보적인 프라이머가 결합하여 대규모 복제 과정이 이루어지게 하였다(**Figure 3**). 클러스터로 정렬된 DNA에 서로 다른 색을 띠는 형광으로 라벨링된 염기(A, T, G, C)가 결합하면서 발생하는 빛을 측정하였다. 100 bp(base pair) 서열을 양방향에서 시퀀싱하여(paired-end sequencing) 총 200 cycle로 진행하였으며, Illumina platform인 HiSeq2000(Illumina, San Diego, CA)을 사용하였다(LabGenomics).

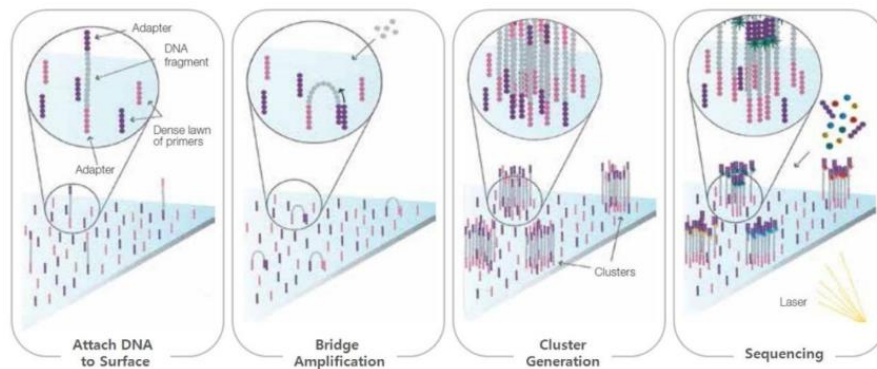


Figure 3. Process of cluster generation (LabGenomics)

제3절 소나무의 표준전사체 생성

시퀀싱으로 생성된 raw read의 품질을 확인하기 위해 FastQC(v0.11.9)를 사용하였다(Andrews, 2010) (Figure 4). 어댑터를 포함하거나 주어진 threshold quality보다 낮은 quality score를 가진 read는 Trimmomatic(v0.36)을 사용하여 제거하였다 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) (Bolger et al., 2014).

```
@A01377:126:HFJVJDRX2:1:2101:1597:1000 1:N:0:TCTAGGAG+AGGTCACT
TACGATCTCACGCCGACCTAACATTGGATTCTGTGGGTGTTTTGGACAGCAAGGCTGAAGATGTCCCAACAGCTCTCAATCCCCTGCGGAGGGGTCGGG
+
FFFFFFFF:FFFFF,FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:1850:1000 1:N:0:TCTAGGAG+AGGTCACT
ACAGTTTCTTCTCTGTGGAGGGGTACCTGCTCTTCTCGGGAAGAGGTTCTCCACATAATCGTTCTCAAATGCATCAGTAGGAACCTCATATATTC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:3043:1000 1:N:0:TCTAGGAG+AGGTCCCT
GCCGTCTAAATCCCCTCTCCATCATCGCCGACGCTCTGAGTCCGTCAGGGAATCCAACCTAACAGACCCGAATTCTCTTCAAGCTCTCCAAAGTT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:3097:1000 1:N:0:TCTAGGAG+AGGTCACT
GCCGTCTAAATCCCCTCTCCATCATCGCCGACGCTCTGAGTCCGTCAGGGAATCCAACCTAACAGACCCGAATTCTCTTCAAGCTCTCCAAAGTT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:4345:1000 1:N:0:TCTAGGAG+AGGTCACT
GAGAATAGGGAACCTCAAGTTAGAAGAGGGAAGAGTCTAGCAACGTGTTCCAAGGGAATACCATCTGCCGTGTTCTTCTCAACTCTGGATGGTTATATT
+
:FFF,FFFFFF,FFFFF:FFFFFFF:FFFFFFFFFFFFF::FFFFFFFFFFFFF:FFFFFFF:FFF:FFFFFF,FFFFFFFFFFFFF:FFF:F,FF:FFF
@A01377:126:HFJVJDRX2:1:2101:4417:1000 1:N:0:TCTAGGAG+AGGTCACT
GGAGGAAGCAACTGATAAAGAGAGATGAGAGAGGTGATAAATCCGAAGCCCGGTAATTCTGTGCGTGACCTTCTTAGTGCCAACTGAAGTAGATCGG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:4689:1000 1:N:0:TCTAGGAG+AGGTCACT
TCGAATGGAAAGCCCGCCAAACACACACGAGCGCGCAGGTTGCACAGGCCCGCCACAGGGCGCACGCTTCTGTTCTCTGACCACCCGAGGGCAAACC
+
FFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:4797:1000 1:N:0:TCTAGGAG+AGGTCACT
GCCTCTCTGTTCTCTCTCAACATACTCCGCCACTGGATTTCAATACGCTGCTGCTGTTGGAAAAATCTG6GAGAACTTCTTGCACTACTATTGC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A01377:126:HFJVJDRX2:1:2101:5177:1000 1:N:0:TCTAGGAG+AGGTCACT
TCCTCATCTGGGTGTTCAACTATGGGCTCTGTGCTAGCAGGTTAGATTACAGATGCTGGAGCTTCATTGTGCATCAGGCAGCTCTCAATCACAGTCCCC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```

Figure 4. Example of FASTQ file format (A-1_1_paired.fq)

Trinity(v2.11.0)를 사용하여 드노보 어셈블리(*de novo* assembly) 방법으로 표준전사체를 생성하였다(Grabherr et al., 2011; Haas et al., 2013). 모든 유전체 서열(contig)의 길이에 대한 평균과 중앙값, N50은 Trinity에서 제공하는 ‘TrinityStats.pl’ 스크립트를 활용하여 계산하였다.

TransDecoder(<https://github.com/TransDecoder/TransDecoder/wiki>)를 사용하여 조립된 모든 서열들 중 100 아미노산 이상의 Open Reading Frame(ORF)를 만족하는 서열을 추출하였다. 동형 유전자 (gene isoforms)를 제거하기 위해 CD-HIT(v4.8.1)를 사용하여 전체 서열들을 클러스터링 한 뒤, 99% 유사한 서열들 중 가장 긴 서열을 추출하여 표준전사체로 활용하였다(Li and Godzik, 2006). 추출한 뉴클레오타이드 서열들을 하나의 파일인 “NRCDs(Non-Redundant CoDing Sequence)_Trinity.fasta”로 지정하여 최종 표준전사체로서 이후 분석에 사용하였다.

제4절 발현량 정량화 및 DEG 분석

차등 발현 유전자(Differentially Expressed Genes, DEGs) 분석은 Trinity에서 제공하는 분석법과 스크립트를 참고하여 진행하였다(Grabherr et al., 2011; Haas et al., 2013). 전사체 발현 정도를 결정하기 위해 RNA-Seq by Expectation-Maximization(RSEM) (<http://deweylab.github.io/RSEM/>) 방법을 사용하여 샘플별 모든 전사체를 표준전사체에 mapping 하였다(Li and Dewey, 2011).

생성된 발현값을 정규화하기 위해 Trimmed Mean of M-values(TMM) 정규화 방법을 사용하였다. 또한 생물학적 반복을 통한 통계적 유의성을 확인하기 위해 주성분분석(Principal Components Analysis, PCA)과 샘플간 상관관계 분석을 진행하였다. R 패키지인 edgeR을 사용하여 모든 샘플들에 대해 DEG 분석을 실시하였다. 4배 이상의 Fold Change(FC)를 가지고 p -value가 0.001보다 작으면 DEG로 판단하였으며, 샘플별 DEG 발현값에 대한 heatmap을 생성하여 시각화하였다.

제5절 소나무 기관 비교별 GO enrichment 분석

총 4개의 기관별 비교(A vs F1, F1 vs F3, F1 vs M1, M1 vs M2)에서 각각의 DEG들을 추출하였다. TransDecoder(<https://github.com/TransDecoder/TransDecoder/wiki>)를 사용하여 단백질 서열을 예측하고 Trinotate pipeline을 사용하여 여러 종류의 functional annotation을 포함하는 SQLite 데이터베이스를 생성하였으며, 그 과정은 다음과 같다(Bryant et al., 2017).

먼저 BLAST+(Camacho et al., 2009)를 사용하여 RefSeq plant(O’Leary et al., 2016) 데이터베이스와의 서열 상동성을 비교하였고, HMMER(Eddy, 2011)을 이용하여 PFAM(Finn et al., 2016)으로 단백질 도메인을 확인하였다. 또한 SignalP(Petersen et al., 2011)로 단백질 신호 펩타이드(signal peptide)를 확인하였고, TmHMM(Krogh et al., 2001)으로 막관통성(transmembrane) 도메인을 예측하였으며, eggNOG(Huerta-Cepas et al., 2016), GO(Gene Ontology Consortium, 2015), KEGG(Kanehisa et al., 2016)를 포함한 여러 데이터베이스와 비교하였다. 모든 결과는 Trinotate SQLite 데이터베이스에 추가하였다. Trinity 소프트웨어 패키지가 제공하는 스크립트를 활용하여 각 비교별로 Gene Ontology(GO) enrichment 분석을 진행하였다.

제6절 상동성 기반의 개화 조절 유전자 탐색

GO enrichment 분석 결과를 기반으로 하여 개화에 직접적으로 관여하는 GO term에 포함된 전사체들을 모두 추출하였다. 해당 전사체들의 기능을 밝히기 위해 Basic Local Alignment Search Tool for protein database using a translated nucleotide query (BLASTX, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>)를 사용하여 NCBI Non-Redundant (NR) protein sequences 데이터베이스와의 상동성을 비교하고 각 전사체 서열과 가장 유사한 다른 수종들의 유전자를 검색하였다. 상동성 점수(score)가 가장 높은 전사체 서열을 선정하여 각 개화 조절 유전자를 대표하는 후보 유전자(candidate gene)로 지정하였으며, TMM 발현값을 이용하여 기관별 개화 조절 유전자들의 발현을 비교하였다.

제7절 qRT-PCR과 NGS 데이터의 상관분석

줄기(S)를 비교군으로 설정하여 초기 암구과(F1)와 초기 수구과(M1) 샘플 모두에서 up-regulated DEG에 해당하는 서열 6개와 down-regulated DEG에 해당하는 서열 6개를 각각 랜덤으로 선발하여 총 12개의 서열을 추출하였다. 모든 기관에서 항상 높게 발현하는 하우스키핑 유전자(housekeeping gene)인 18S rRNA를 기준 유전자(reference gene)로 선택하였으며(Niu et al., 2016), 테다소나무(*Pinus taeda*)의 18S rRNA 유전자 서열(Accession: AH001728.2)과 가장 상동성이 높은 서열을 앞서 생성한 소나무 표준전사체에서 추출하였다. Primer3(<https://bioinfo.ut.ee/primer3/>)에서 모든 조건은 그대로 유지하고 증폭길이 70-150 bp와 melting temperature(T_m) 60°C만 변경하여 총 13개의 프라이머를 제작하였다(Table 2).

3개의 기관(S, F1, M1)에 대해 소나무 3 개체의 RNA 샘플을 사용하여 총 9개의 cDNA를 합성하였으며, 이때 ReverTra Ace® qPCR RT Master Mix(TOYOBO)를 사용하였다. 3차 증류수로 희석한 RNA template 8 ul(총 100 ng)와 5x RT Master Mix 2 ul를 포함하여 샘플 당 총 반응용액은 10 ul로 조성하였다. PCR 반응은 37°C에서 15분, 50°C에서 5분, 98°C에서 5분 반응시킨 후 4°C에서 임시 보관하였다. PCR 장치는 TaKaRa PCR Thermal Cycler Dice® TP600을 사용하였다.

앞서 제작한 프라이머들에 대한 각 샘플들의 상대적인 발현량을 확인하기 위해 quantitative Real-Time PCR(qRT-PCR)을 진행하였다. 총 반응용액 20 ul 당 ExcelTaq™ 2X Fast Q-PCR Master Mix(SYBR, no ROX)(SMOBIO) 10 ul, cDNA template 2 ul(총 100 fg), 포워드/리

버스 프라이머 각각 4 ul(2 uM)를 포함하도록 조성하였다. 95°C에서 30초 열변성한 후, 95°C 15초와 60°C 30초를 40회 반복하여 형광을 측정하고 65~95°C 구간에서 0.5°C씩 5초 간격으로 melting-curve를 분석하였다. qRT-PCR은 CFX Connect™ Real-Time System과 Bio-Rad CFX Maestro를 사용하여 진행하였다. RNA-seq 발현값과의 비교를 위해 회귀 분석을 실시하였으며, 증폭 데이터는 Delta Delta Ct 방법으로 다음과 같이 계산하였다(Equations 1-1, 1-2 and 1-3).

$$\Delta Ct(\text{sample}) = Ct(\text{experimental gene}) - Ct(\text{reference gene}) \dots\dots \text{Equation 1-1}$$

$$\Delta\Delta Ct = \Delta Ct(\text{calibrator tissue}) - \Delta Ct(\text{other tissue}) \dots\dots\dots \text{Equation 1-2}$$

$$\text{Fold change from calibrator tissue} = 2^{-\Delta\Delta Ct} = \frac{1}{2^{\Delta\Delta Ct}} \dots\dots\dots \text{Equation 1-3}$$

Table 2. Oligonucleotide primers used for validation by qRT-PCR

| Target gene | Forward primer | T_m (°C) | Reverse primer | T_m (°C) |
|------------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| TRINITY_DN2904_c0_g2_i1 | GCTTGACAAGGCTGCAACTC | 60.9 | GACGCATCTTGTCCTTCGC | 61.0 |
| TRINITY_DN1177_c0_g2_i2 | AGGTACCAGATCGAGCAGAGT | 61.8 | ACAGTCAGCCTTTTCCTGGA | 61.6 |
| TRINITY_DN16942_c0_g1_i14 | GCTCATCCACTGACACCTCC | 63.7 | TATCCATCCACAGCCAGCAG | 61.9 |
| TRINITY_DN4827_c0_g1_i12 | CGCAGACATTTTCACCGTGG | 61.7 | GTGTGACTTCGCTTCCTCCA | 62.5 |
| TRINITY_DN844_c0_g2_i9 | GGTACGGATCCTGTGCATT | 61.9 | AAGCCCTGGTCAAGTCGAAG | 62.8 |
| TRINITY_DN17799_c0_g1_i3 | CATCTTCAATGGGGTTTGGGT | 61.2 | TATGGGATCTGTCTGCTGC | 61.5 |
| TRINITY_DN16126_c0_g1_i1 | GATGTGGCCCTGTAACGGAA | 62.8 | CTGTGTATCTCTGCGCTCGT | 60.4 |
| TRINITY_DN10128_c0_g1_i1 | GGTTTGAAGACGTCGGGGAT | 63.3 | GCATCAAACCTGTGGACGCTC | 60.9 |
| TRINITY_DN3863_c0_g1_i21 | AATCCTGGGCCCTACTCAT | 64.2 | GACTTTTCAGGAGGGGTGGG | 65.6 |
| TRINITY_DN3931_c0_g1_i1 | AGCTCGCTAACCCATGCAAT | 60.3 | AGTGTTGTTCTTGGTGGGCT | 61.6 |
| TRINITY_DN262285_c0_g1_i1 | ACTTTGTGCGCAGCCCATACT | 60.3 | CACCGATCAGAGCTTGCAGA | 61.8 |
| TRINITY_DN2835_c0_g1_i10 | TGACTACTGTGGCGGATTGG | 62.2 | TTGATCTCATCGAGCGAGGC | 61.7 |
| TRINITY_DN986_c1_g1_i10 (18S rRNA) | TCAGAGTGTTTGGGTTCGG | 63.7 | GGACCTGGTAAGTTTCCCCG | 64.9 |

* T_m: melting temperature

제4장 연구결과

제1절 소나무 전사체의 시퀀싱과 표준전사체 생성

소나무의 구과 발달 과정에서 채취한 모든 샘플들의 RNA를 사용하여 Illumina 플랫폼으로 시퀀싱하였다. 총 30개 샘플의 시퀀싱 결과 총 1,878,697,858개의 100 bp read가 생성되었다(Table 3).

평균적으로 97.01%의 read가 Q20 이상, 91.79%가 Q30 이상의 품질을 보였으며 평균 GC 함량은 45.61%였다. 인식하지 못한 염기의 비율을 나타내는 N 함량은 모든 샘플에서 0으로 나타났다. 낮은 품질과 어댑터를 포함하는 read를 제거하는 전처리 과정을 통해 총 1,770,005,422개(94.21%)의 clean read를 얻었다.

Table 3. Statistics of raw sequencing data (Quality scale: +33)

| | Sample | Yield | Read | N(%) | GC(%) | Q20(%) | Q30(%) |
|----|---------------|------------------------|----------------------|-------------|--------------|---------------|---------------|
| 1 | A-rep1 | 6,361,938,490 | 62,989,490 | 0 | 45.03 | 97.22 | 92.16 |
| 2 | A-rep2 | 5,826,017,138 | 57,683,338 | 0 | 44.96 | 96.91 | 91.52 |
| 3 | A-rep3 | 8,358,700,006 | 82,759,406 | 0 | 45.63 | 97.09 | 91.92 |
| 4 | F1-rep1 | 5,706,921,776 | 56,504,176 | 0 | 45.51 | 97.00 | 91.82 |
| 5 | F1-rep2 | 7,215,509,690 | 71,440,690 | 0 | 45.37 | 96.96 | 91.67 |
| 6 | F1-rep3 | 6,659,688,712 | 65,937,512 | 0 | 45.65 | 97.03 | 91.88 |
| 7 | F2-rep1 | 6,324,507,082 | 62,618,882 | 0 | 45.89 | 97.20 | 92.19 |
| 8 | F2-rep2 | 7,927,177,506 | 78,486,906 | 0 | 45.60 | 97.06 | 91.95 |
| 9 | F2-rep3 | 6,326,002,690 | 62,633,690 | 0 | 45.79 | 97.06 | 91.94 |
| 10 | F3-rep1 | 5,953,705,176 | 58,947,576 | 0 | 45.71 | 96.93 | 91.75 |
| 11 | F3-rep2 | 6,292,284,850 | 62,299,850 | 0 | 45.40 | 96.94 | 91.56 |
| 12 | F3-rep3 | 6,492,444,832 | 64,281,632 | 0 | 45.94 | 97.01 | 91.80 |
| 13 | L-rep1 | 6,462,792,040 | 63,988,040 | 0 | 47.61 | 96.97 | 91.70 |
| 14 | L-rep2 | 6,182,348,976 | 61,211,376 | 0 | 45.96 | 97.00 | 91.74 |
| 15 | L-rep3 | 7,098,430,894 | 70,281,494 | 0 | 47.15 | 97.07 | 91.92 |
| 16 | M1-rep1 | 7,113,335,060 | 70,429,060 | 0 | 45.81 | 97.30 | 92.37 |
| 17 | M1-rep2 | 5,954,972,726 | 58,960,126 | 0 | 45.62 | 97.05 | 91.89 |
| 18 | M1-rep3 | 5,825,425,480 | 57,677,480 | 0 | 45.45 | 97.29 | 92.37 |
| 19 | M2-rep1 | 5,845,992,514 | 57,881,114 | 0 | 45.57 | 96.49 | 90.67 |
| 20 | M2-rep2 | 5,749,552,462 | 56,926,262 | 0 | 45.15 | 97.16 | 92.03 |
| 21 | M2-rep3 | 6,740,956,140 | 66,742,140 | 0 | 45.02 | 96.60 | 90.89 |
| 22 | M3-rep1 | 5,759,522,576 | 57,024,976 | 0 | 45.55 | 97.21 | 92.18 |
| 23 | M3-rep2 | 5,899,182,548 | 58,407,748 | 0 | 45.63 | 97.03 | 91.91 |
| 24 | M3-rep3 | 6,596,112,040 | 65,308,040 | 0 | 45.36 | 96.99 | 91.77 |
| 25 | R-rep1 | 5,258,468,040 | 52,064,040 | 0 | 45.45 | 97.01 | 91.79 |
| 26 | R-rep2 | 5,177,615,520 | 51,263,520 | 0 | 45.49 | 97.20 | 92.08 |
| 27 | R-rep3 | 5,929,032,492 | 58,703,292 | 0 | 45.37 | 96.87 | 91.41 |
| 28 | S-rep1 | 5,874,811,854 | 58,166,454 | 0 | 45.43 | 96.77 | 91.24 |
| 29 | S-rep2 | 5,575,590,466 | 55,203,866 | 0 | 45.11 | 97.12 | 91.87 |
| 30 | S-rep3 | 7,259,443,882 | 71,875,682 | 0 | 45.16 | 96.90 | 91.58 |
| | Total | 189,748,483,658 | 1,878,697,858 | 0.0 | 45.61 | 97.01 | 91.79 |

* Q30(%): Yield of bases with Q30 or higher

드노보 어셈블리를 통해 소나무의 표준전사체 서열에 해당하는 하나의 파일을 생성하였다. Trinity 통계 결과에 따르면, 총 621,759개의 ‘transcript’와 총 390,513개의 ‘gene’이 생성되었고 평균 GC 함량은 41.75%로 나타났다(Table 4). 이때 유사한 ‘transcript’ 서열들 중 가장 긴 서열을 대표적으로 ‘gene’이라고 지정하였다. 99% 유사한 동형 유전자들을 제거하여 소나무 전사체를 대표하는 최종 파일을 생성하였다.

Table 4. Statistics of transcripts and genes from *de novo* assembly in *Pinus densiflora*

| | Transcript | Gene |
|-----------------------|-------------------|-------------|
| Total number | 621,759 | 390,513 |
| Contig N10 | 4,333 | 3,591 |
| Contig N20 | 3,164 | 2,342 |
| Contig N30 | 2,426 | 1,563 |
| Contig N40 | 1,883 | 1,020 |
| Contig N50 | 1,408 | 680 |
| Median contig length | 381 | 309 |
| Average contig | 764.97 | 534.08 |
| Total assembled bases | 475,628,258 | 208,565,793 |
| Percent GC (%) | 41.75 | |

* Gene: the longest of the 99% similar isoforms

제2절 소나무 기관들의 유전자 차등 발현

DEG 분석에 앞서, 계층적 군집분석(hierarchical clustering) 방법을 사용하여 소나무의 기관별, 시기별 샘플 유사성을 확인하였다(Figure 5). 암구과(F1~3)를 제외하고 모든 생물학적 반복 샘플들(rep1~3)은 각각 높은 유사성을 나타내었다. 암구과는 반복 샘플간 상대적으로 낮은 상관 계수를 보였으며, 암구과의 각 발달 단계별로 완전하게 클러스터링이 되지 않았음을 알 수 있다. 한편 소나무의 서로 다른 6개 기관들(A, F, L, M, S, R)은 확실하게 구분이 되어 분포하였다. 암구과는 세 번의 발달 시기가 합쳐져서 가장 유사하게 묶였으며, 상대적으로 정단과 가장 가깝게 나타났다.

수구과의 경우 초기 발달 단계(M1, M2)는 서로 유사하게 묶였지만 후기 발달 단계(M3)는 다른 수구과 단계들과 분리되어 나타났다. 그리고 줄기(S)와 잎(L) 샘플들은 상대적으로 서로 가깝게 분포하였으며, 수구과 집단보다는 정단/암구과 집단과 가깝게 묶였다. 마지막으로 뿌리(R) 샘플들은 다른 기관 샘플들과 뚜렷하게 분리되어 나타났다.

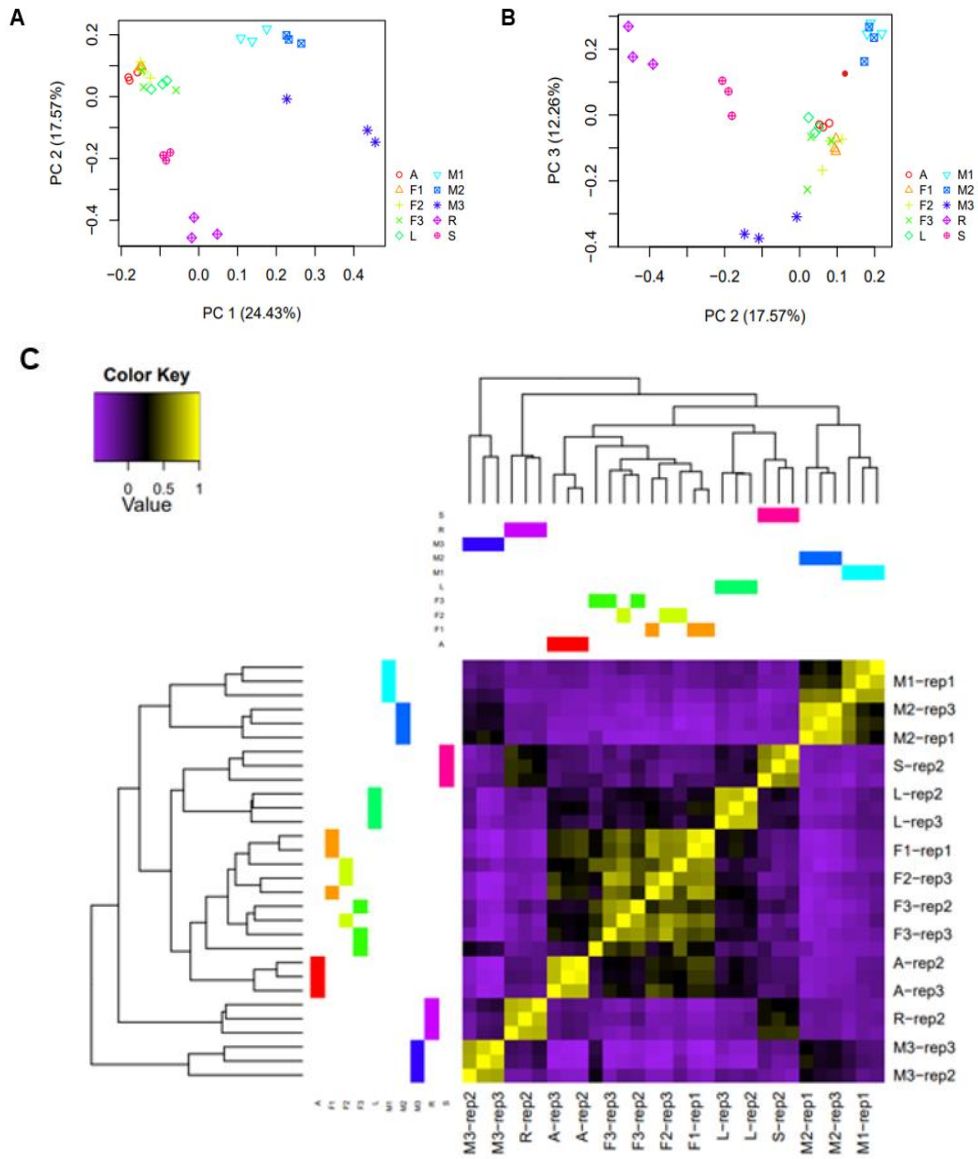


Figure 5. PCA plots (A-B) and Pearson correlation matrix (C) of all samples

총 10개 기관의 샘플들을 대상으로 한 차등발현 분석 결과, 전체 45개의 쌍별 비교(pairwise comparison)에서 총 25,307개의 DEG가 확인되었다(Figure 6). 뿌리(R)와 다른 기관들의 비교에서 가장 많은 DEG가 나타났으며, 정단(A)과 암구과(F1~3) 샘플들의 비교에서 가장 적은 발현 차이를 보였다. 암구과 발달 단계별 비교에서는 F2 vs F3에서 70개 DEG로 가장 적었고, F1 vs F3에서 597개 DEG로 가장 많았다. 수구과 발달 단계별 비교에서는 M1 vs M2에서 1,343개 DEG로 가장 적었고, M1 vs M3에서 4,802개 DEG로 가장 많았다.

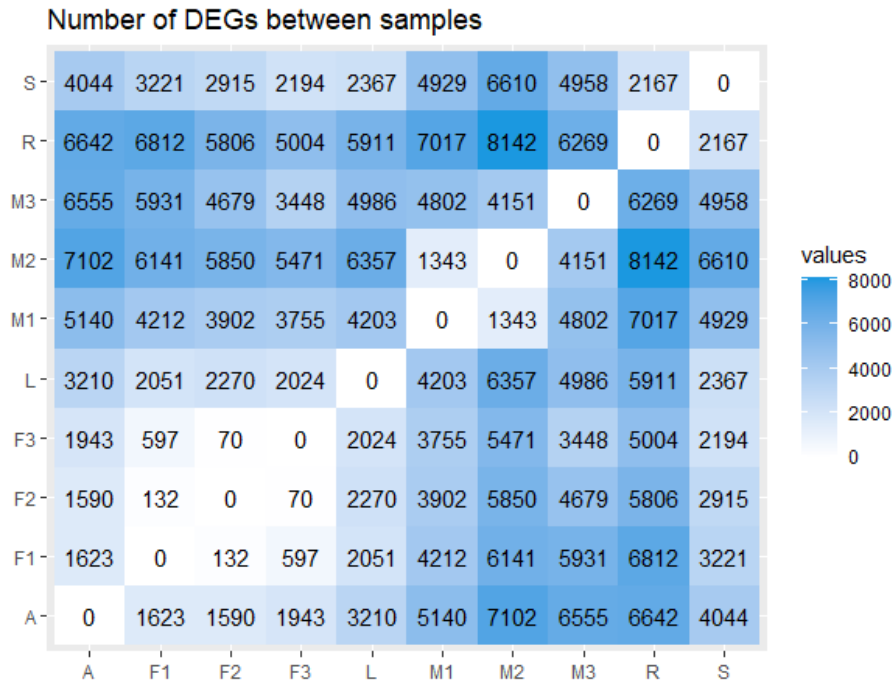


Figure 6. Number of DEGs between reproductive organ samples of *Pinus densiflora*

정단과 초기 암구과 비교(A vs F1), 초기와 후기 암구과 비교(F1 vs F3), 초기 암구과와 수구과 비교(F1 vs M1), 그리고 초기와 중기

수구과 비교(M1 vs M2)에서 얻은 DEG들을 사용하여 총 4개의 비교에 대한 전사체 발현을 자세하게 분석하였다(Figures 7 and 9). 각 비교에서의 MA plot과 Volcano plot을 통해 False Discovery Rate(FDR)<0.05인 DEG들을 시각화하였다(Figure 8).

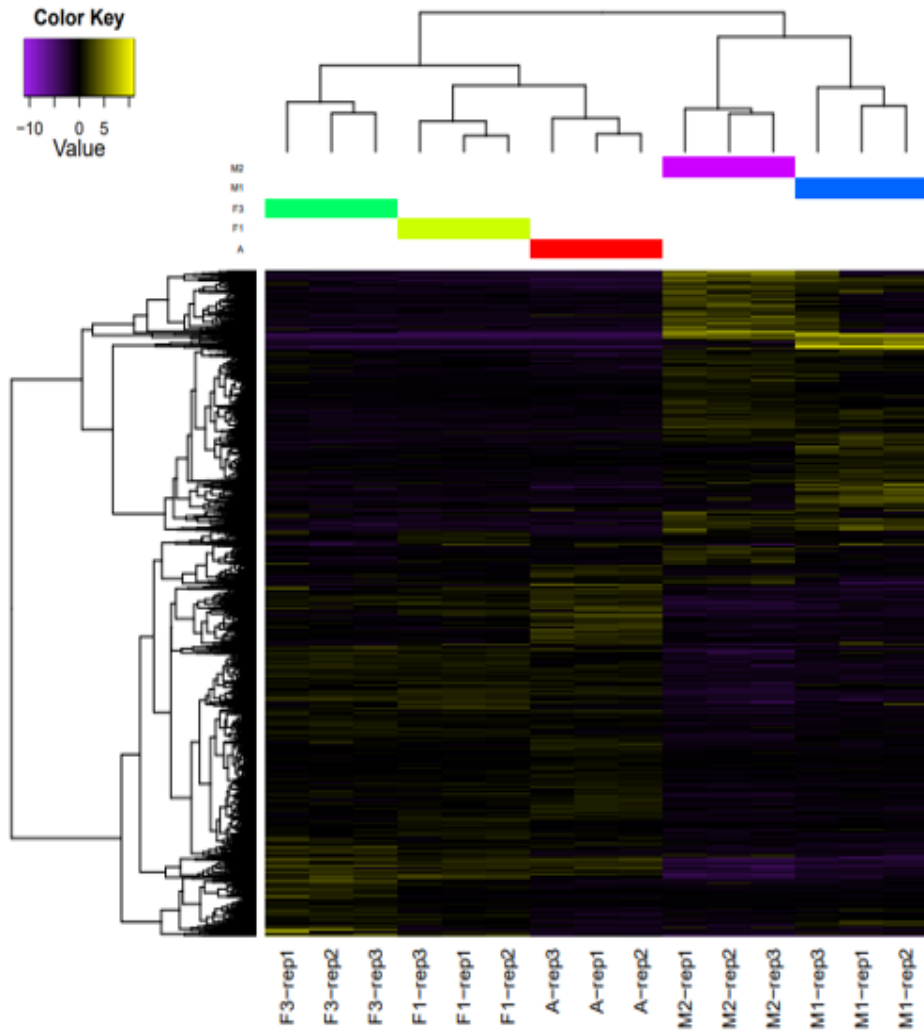


Figure 7. Heatmap of DEGs and samples used in pairwise comparisons

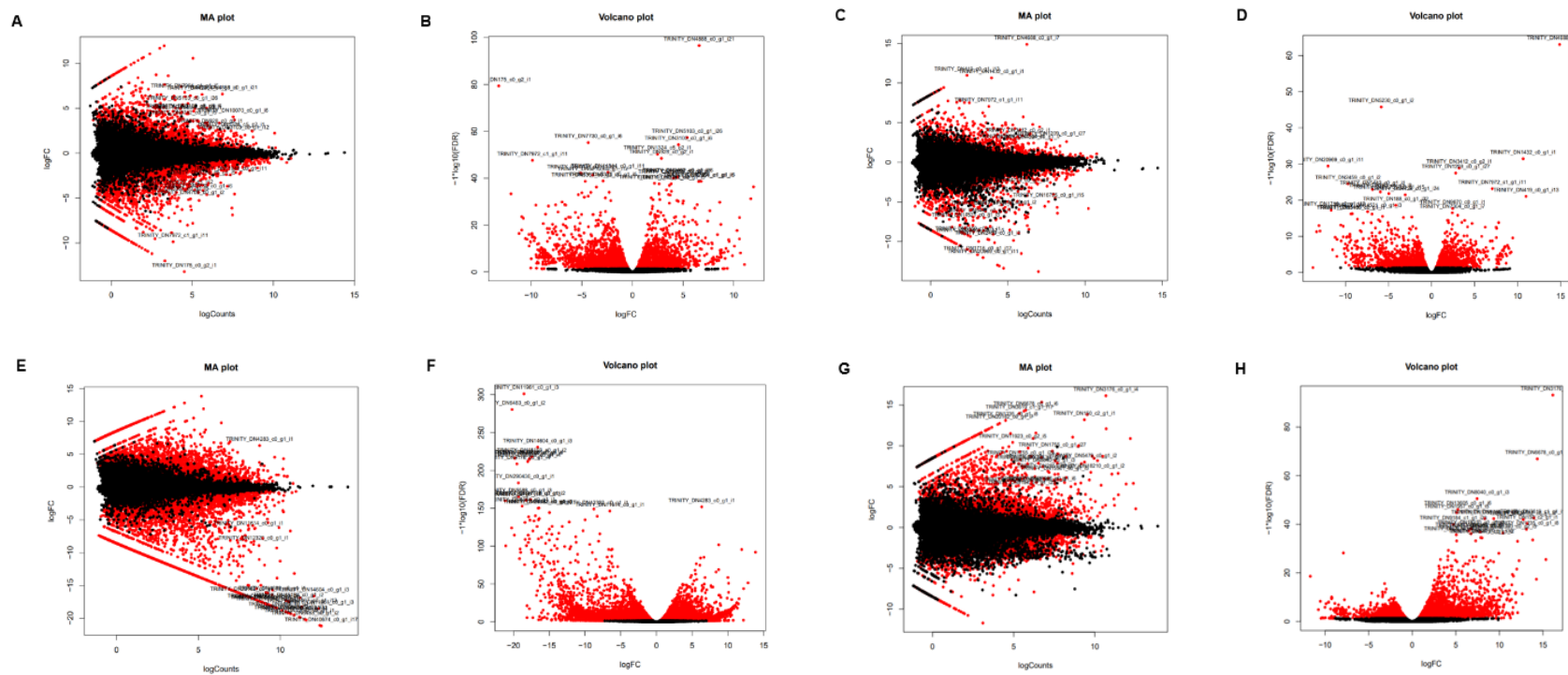


Figure 8. MA plot and Volcano plot of each pairwise comparison. A-B: A vs F1, C-D: F1 vs F3, E-F: F1 vs M1, G-H: M1 vs M2

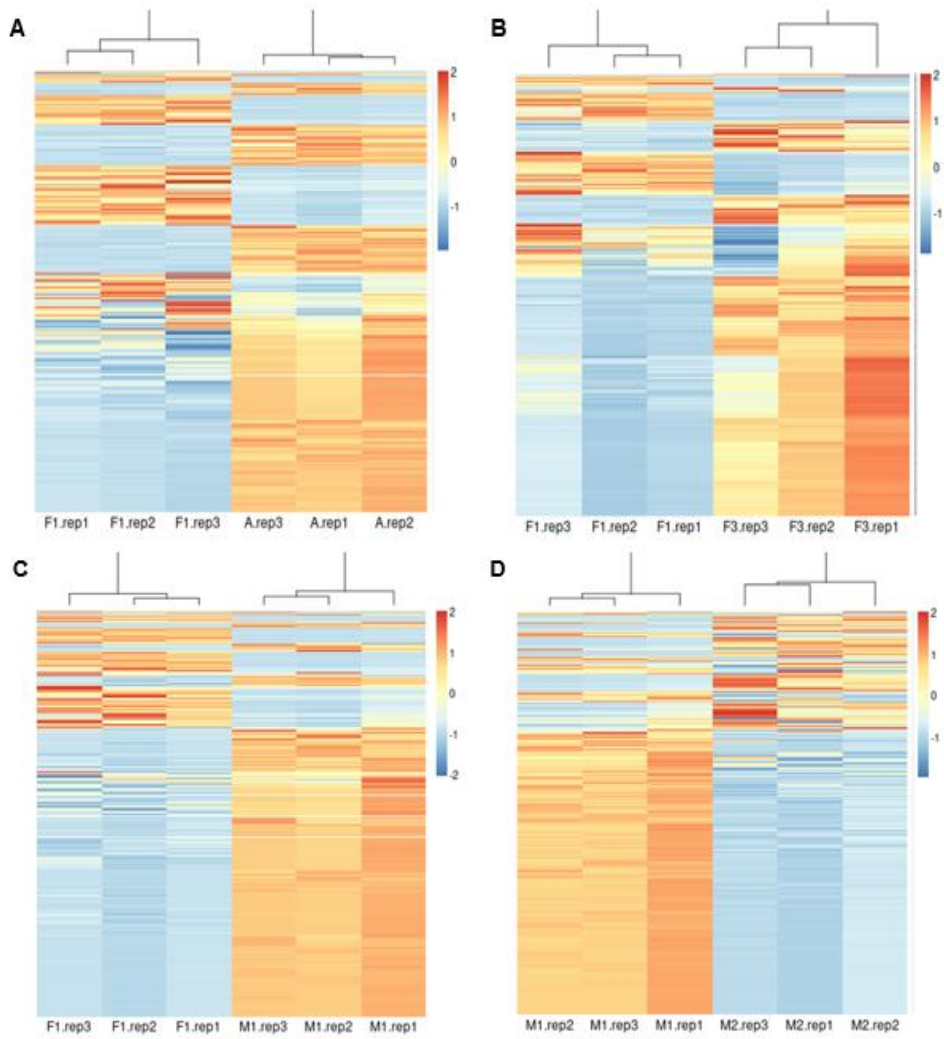


Figure 9. Heatmap of DEGs on each comparison. A: A vs F1, B: F1 vs F3, C: F1 vs M1, D: M1 vs M2

A vs F1 비교에서 나타난 총 1,623개의 DEG 중 정단(A)에서 716개, 초기 암구과(F1)에서 907개가 각각 높게 발현하였고(upregulated), 유일한 DEG 개수가 1,121개로 나타났다(**Figure 10**). 암구과 발달 단계인 F1 vs F3 비교에서는 총 597개의 DEG 중 초기 암구과(F1)에서 256개, 후기 암구과(F3)에서 341개가 각각 높게 발현하였고, 유일한 DEG는 344개로 가장 적었다.

초기 암구과와 초기 수구과를 비교하는 F1 vs M1 비교에서는 총 4,212개의 DEG 중 암구과(F1)에서 2,104개, 수구과(M1)에서 2,108개가 각각 높게 발현되었으며, 유일한 DEG의 개수가 3,002개로 가장 많았다.

수구과 발달 단계인 M1 vs M2 비교에서는 총 1,343개의 DEG 중 초기 수구과(M1)에서 986개, 후기 수구과(M2)에서 357개가 각각 높게 발현하였으며, 573개의 유일한 DEG가 존재하였다. 정단과 암구과 발달 단계(A vs F1/F1 vs F3)에서 공유하는 DEG의 개수는 58개이고, 암구과와 수구과 발달 단계(F1 vs F3/M1 vs M2)에서 공유하는 DEG의 개수는 22개였다. 모든 비교에서 공통적으로 나타난 DEG의 개수는 3개였다.

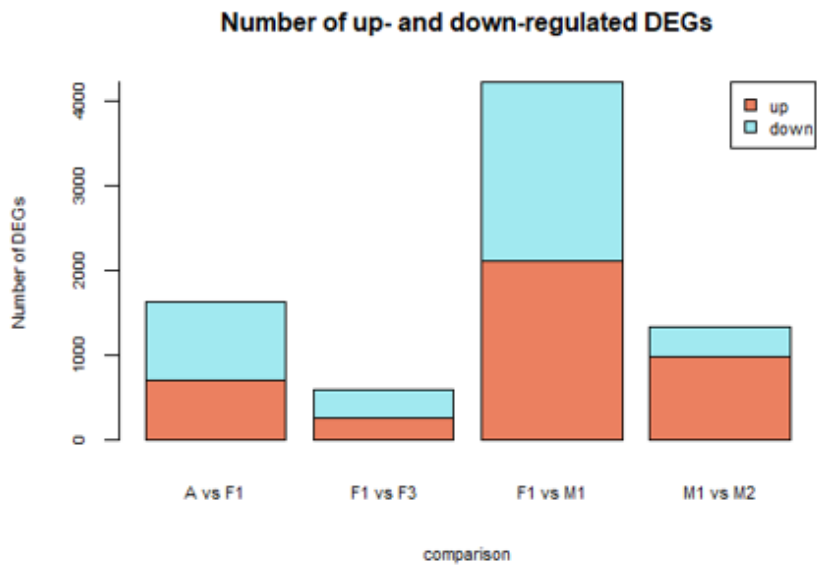
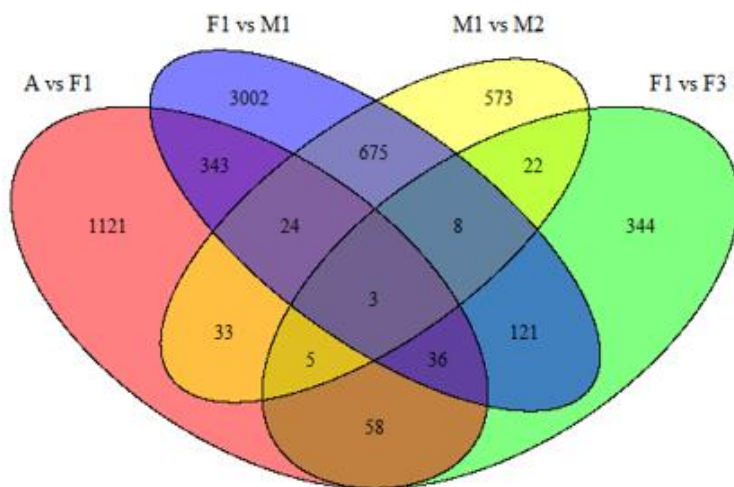
A**B**

Figure 10. The number of DEGs in each comparison. A: histogram of up/down-regulated DEGs, B: Venn diagram indicating the number of DEGs

제3절 소나무 기관별 분자세포생물학적 특성 비교

소나무 기관별 발달과 구과의 단계적 발달에 관여하는 전사체들의 기능을 알아내기 위해 각 비교별 DEG들에 대한 상동성 기반의 functional annotation을 진행하였다. 그리고 각 GO term에 속한 DEG들의 빈도를 전체 유전자에서의 빈도와 비교하여 기관 간 비교별 GO enrichment 분석을 진행하였다. 일반적으로 p -value가 0.05보다 작으면 해당 GO term에서 “enriched” 되었다고 할 수 있다(Ren et al., 2017). 따라서 본 연구에서는 p -value가 작은 순으로 상위 30개의 GO term을 기준으로 하여 각 비교별 분자세포생물학적 특성을 해석하였다.

A vs F1의 경우, 정단(A)에서 ‘positive regulation of programmed cell death(GO:0043068)’, ‘plant-type hypersensitive response(GO:0009626)’, ‘positive regulation of cell death(GO:0010942)’와 같이 세포사와 관련된 GO가 많이 발현되는 것으로 나타났으며, ‘signal transduction(GO:0007165)’, ‘ethylene-activated signaling pathway(GO:0009873)’와 같은 신호 전달 및 호르몬과 관련된 GO가 많았다(**Figure 11**). 그리고 ‘defense response to bacterium(GO:0042742)’, ‘systemic acquired resistance, salicylic acid mediated signaling pathway(GO:0009862)’, ‘EDS1 disease-resistance complex(GO:0106093)’, ‘defense response to Gram-negative bacterium(GO:0050829)’, ‘defense response to insect(GO:0002213)’, ‘positive regulation of defense response to bacterium(GO:1900426)’, ‘defense response to fungus(GO:0050832)’, ‘systemic acquired resistance(GO:0009627)’, ‘response to other organism(GO:0051707)’ 등 외부 생물에 대한 방어 작용이나 저항성과

관련된 GO의 빈도가 높았다.

반면 초기 암구과(F1)에서는 ‘fatty acid biosynthetic process(GO:0006633)’, ‘lignin biosynthetic process(GO:0009809)’, ‘suberin biosynthetic process(GO:0010345)’와 같이 합성과 관련된 GO들이 많았으며, ‘oxidoreductase activity(GO:0016491)’, ‘oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen(GO:0016702)’처럼 oxidoreductase의 활성화와 관련된 GO가 많이 발현하였다(**Figure 12**).

‘oxylipin biosynthetic process(GO:0031408)’, ‘UDP-glycosyltransferase activity(GO:0008194)’, ‘linoleate 9S-lipoxygenase activity(GO:1990136)’, ‘acid phosphatase activity(GO:0003993)’, ‘phenylpropanoid biosynthetic process(GO:0009699)’, ‘cysteine-type peptidase activity(GO:0008234)’, ‘transferase activity, transferring acyl groups other than amino-acyl groups(GO:0016747)’, ‘dioxygenase activity(GO:0051213)’ 등 대사산물이나 효소와 관련된 GO들이 높은 빈도로 존재하였다. 또한 식물 액포와 관련된 ‘plant-type vacuole(GO:0000325)’, 유전자 메틸화를 의미하는 ‘methylation(GO:0032259)’, 물질 이동을 나타내는 ‘transmembrane transport(GO:0055085)’, ‘apoplast(GO:0048046)’, ‘inorganic phosphate transmembrane transporter activity(GO:0005315)’, ‘phosphate ion transport(GO:0006817)’의 발현 정도가 높았다.

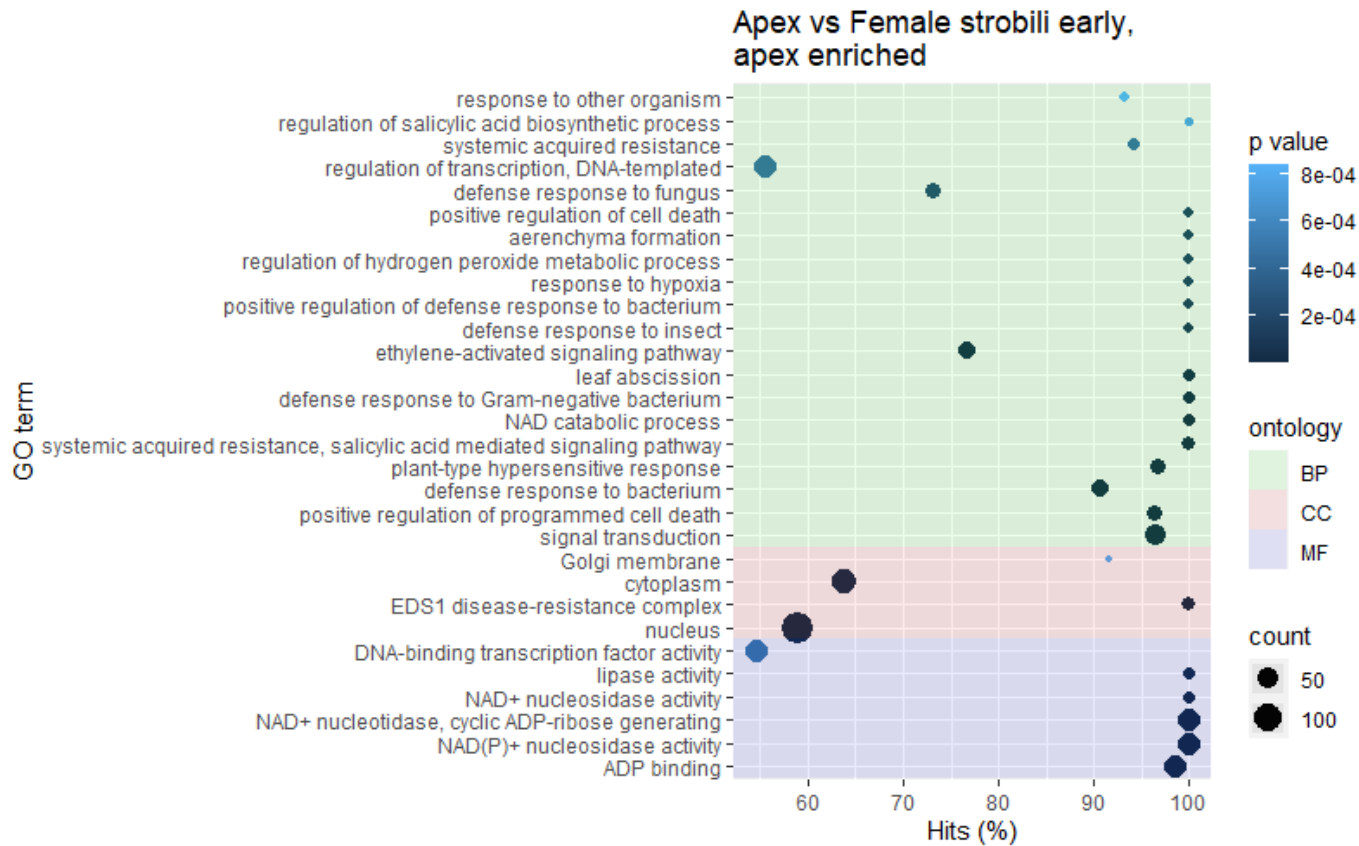


Figure 11. Top 30 GO terms in Gene Ontology analysis for A vs F1 comparison, A enriched

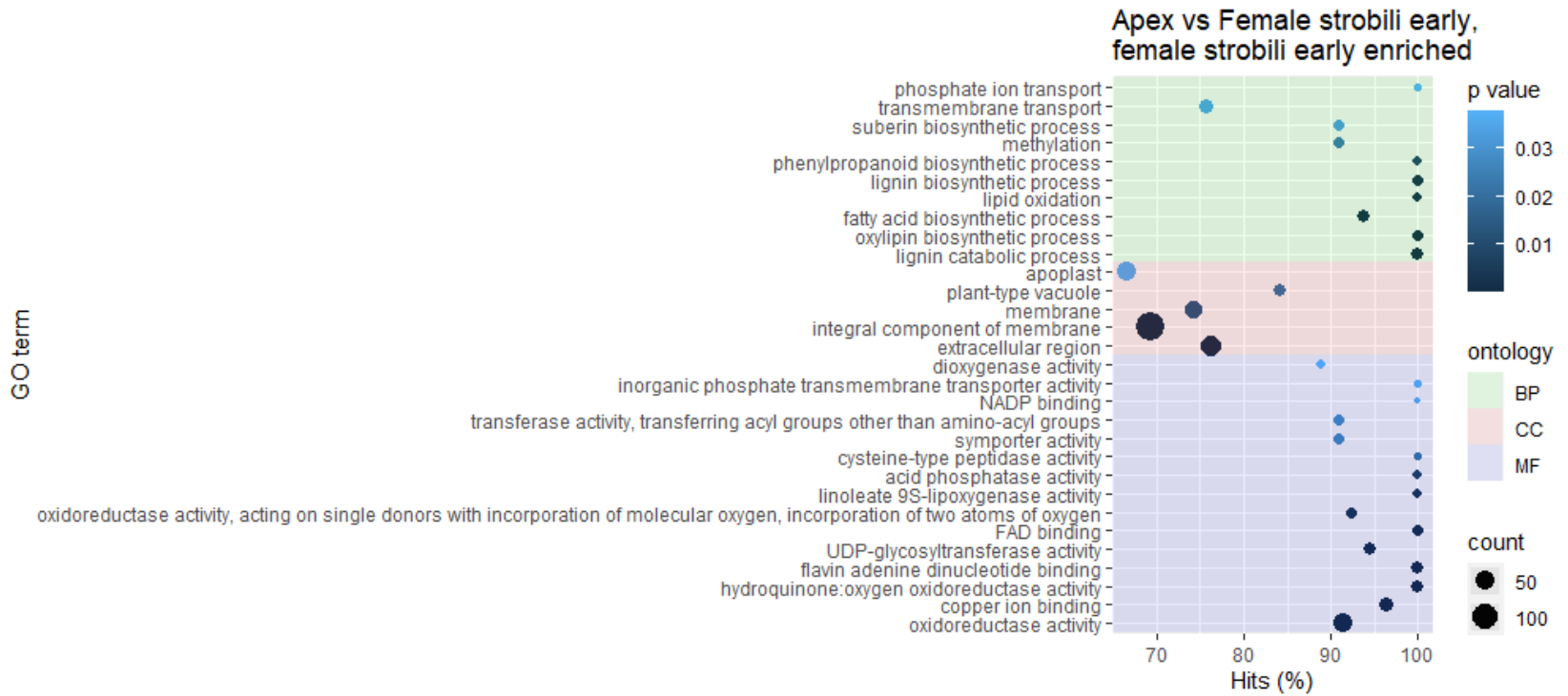


Figure 12. Top 30 GO terms in Gene Ontology analysis for A vs F1 comparison, F1 enriched

암구과 발달 과정을 비교하는 F1 vs F3의 경우, 초기 암구과(F1)에서 꽃잎 형성과 관련된 ‘petal morphogenesis(GO:0048446)’, ‘petal development(GO:0048441)’가 존재했으며, ‘DNA-binding transcription factor activity(GO:0003700)’, ‘regulation of transcription, DNA-templated(GO:0006355)’, ‘DNA binding(GO:0003677)’, ‘transcription regulatory region sequence-specific DNA binding(GO:0000976)’과 같이 DNA 결합이나 전사인사 조절과 관련된 GO의 빈도가 높았다(Figure 13).

단백질의 결합을 의미하는 ‘protein dimerization activity(GO:0046983)’가 많이 발현되었으며, ‘cysteine-type peptidase activity(GO:0008234)’와 같은 효소의 활성도 높게 나타났다. 그리고 ‘endoplasmic reticulum membrane(GO:0005789)’, ‘plant-type secondary cell wall biogenesis(GO:0009834)’, ‘Golgi apparatus(GO:0005794)’ 등 세포 소기관이나 세포벽과 관련된 GO가 존재하였다. ‘diacylglycerol O-acyltransferase activity(GO:0004144)’, ‘long-chain-alcohol O-fatty-acyltransferase activity(GO:0047196)’, ‘arachidoyl-CoA:1-dodecanol O-acyltransferase activity(GO:0019432)’, ‘triglyceride biosynthetic process(GO:0019432)’와 같이 다양한 acyltransferase와 triglyceride(TG)가 높은 발현 정도를 보였다. 또한 설탕 합성과 관련된 ‘sucrose metabolic process(GO:0005985)’, ‘sucrose synthase activity(GO:0016157)’와 지질 합성과 관련된 ‘glycerolipid biosynthesis process(GO:0045017)’가 존재하였다. 특이하게도 화분 형성과 연관된 ‘pollen exine formation(GO:0010584)’가 암구과의 초기 발달 과정에서 발견되었다.

후기 암구과(F3)에서는 ‘ATP binding(GO:0005524)’, ‘ABC-type

transporter activity(GO:0140359)', 'ATPase-coupled transmembrane transporter activity(GO:0042626)' 등 ATP 가수분해를 통한 수송 작용이 많이 나타났다(**Figure 14**). 그리고 효소 활성 및 유기화합물질과 관련된 'dioxygenase activity(GO:0051213)'와 'terpenoid biosynthetic process(GO:0016114)', 'serine-type endopeptidase activity(GO:0004252)', 'aspartic-type endopeptidase activity(GO:0004190)', 'proteolysis(GO:0006508)'의 빈도가 높았으며, 또한 병원체와 상처에 대한 반응인 'defense response to oomycetes(GO:0002229)'와 'response to wounding(GO:0009611)'가 존재하였다.

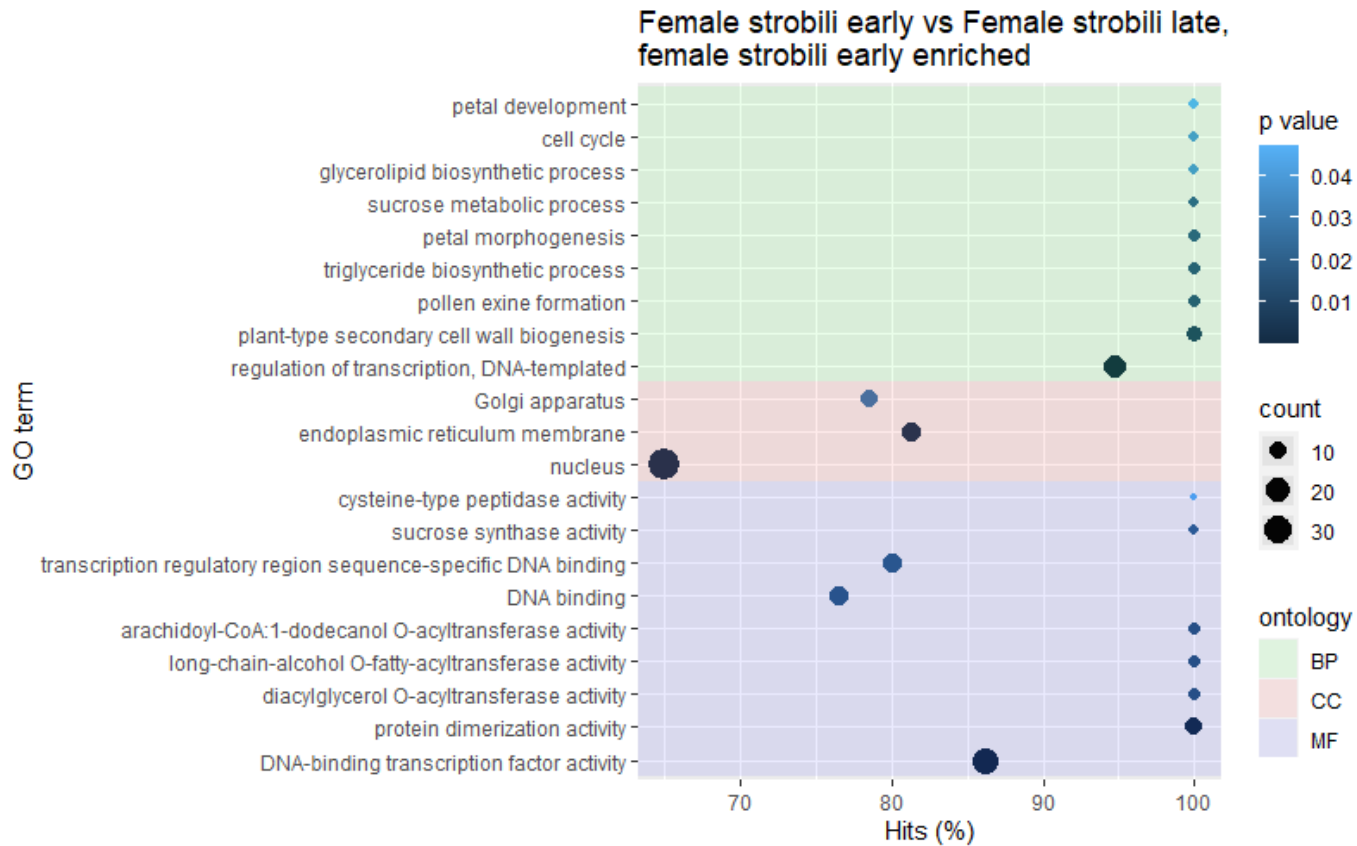


Figure 13. Top 30 GO terms in Gene Ontology analysis for F1 vs F3 comparison, F1 enriched

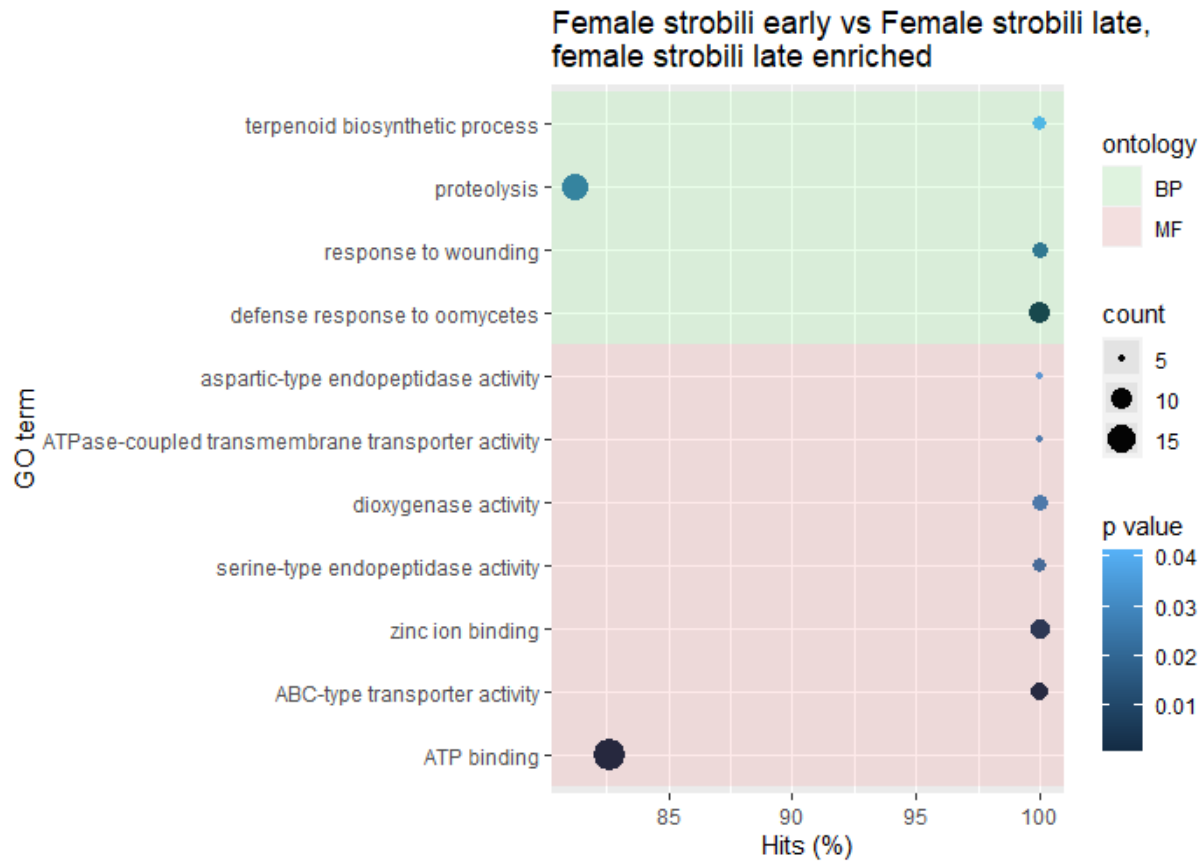


Figure 14. Top 30 GO terms in Gene Ontology analysis for F1 vs F3 comparison, F3 enriched

초기 암구과(F1)와 초기 수구과(M1)를 비교한 F1 vs M1의 경우, 암구과에서 ‘DNA-binding transcription factor activity(GO:0003700)’, ‘regulation of transcription, DNA-templated(GO:0006355)’, ‘transcription regulatory region sequence-specific DNA binding(GO:0000976)’, ‘DNA binding(GO:0003677)’, ‘RNA polymerase II-specific(GO:0000981)’, ‘RNA polymerase II cis-regulatory region sequence-specific DNA binding(GO:0000978)’, ‘sequence-specific DNA binding(GO:0043565)’와 같이 DNA 결합과 전사과정 및 전사인자 조절에 관련된 GO의 빈도가 높았으며, 단백질 결합을 의미하는 ‘protein dimerization activity(GO:0046983)’이 존재했다(Figure 15).

테르펜 합성과 관련된 ‘terpene synthase activity(GO:0010333)’, ‘diterpenoid biosynthetic process(GO:0016102)’, ‘terpenoid biosynthetic process(GO:0016114)’, ‘pinene synthase activity(GO:0050550)’, ‘alpha-pinene metabolic process(GO:0018867)’가 많이 발현하였다. 이온 결합에 대한 ‘magnesium ion binding(GO:0000287)’와 lyase 효소에 대한 ‘lyase activity(GO:0016829)’, ‘pectate lyase activity(GO:0030570)’가 나타났다. 그리고 꽃 발달에 대한 ‘flower development(GO:0009908)’와 색소와 관련된 ‘flavonoid biosynthetic process(GO:0009813)’가 발현하였으며, 빛에 대한 반응인 ‘response to light stimulus(GO:0009416)’가 나타났다. 또한 호르몬 반응을 의미하는 ‘response to gibberellin(GO:0009739)’과 세포사를 의미하는 ‘positive regulation of programmed cell death(GO:0043068)’의 빈도가 높았다. 그리고 인산화반응을 촉매하는 효소인 ‘protein serin/threonine kinase activity(GO:0004674)’가 존재하였다.

수구과에서는 수송 및 이온의 이동과 관련된 ‘ABC-type transporter activity(GO:0140359)’, ‘potassium ion transport(GO:0006813)’, ‘potassium ion transmembrane transport(GO:0071805)’, ‘potassium ion transmembrane transporter activity(GO:0015079)’, ‘calcium ion binding(GO:0005509)’의 빈도가 높았다(Figure 16). 단백질 분해과정을 의미하는 ‘proteolysis(GO:0006508)’, ‘ubiquitin-protein transferase activity(GO:0004842)’와 xylan 분해 과정인 ‘xylan catabolic process(GO:0045493)’, ‘xylan 1,4-beta-xylosidase activity(GO:0009044)’가 나타났으며, arabinan 분해 과정인 ‘arabinan catabolic process(GO:0031222)’, ‘alpha-L-arabinofuranosidase activity(GO:0046556)’ 또한 나타났다. 화분 및 세포벽 형성과 관련된 ‘plant-type primary cell wall biogenesis(GO:0009833)’, ‘sporopollenin biosynthetic process(GO:0080110)’, ‘cellulose synthase (UDP-forming) activity(GO:0016760)’이 많이 발현되었다.

호르몬 전달과 연관된 ‘regulation of auxin mediated signaling pathway(GO:0010928)’이 나타났고, ATP 전달을 통한 인산화반응과 관련된 ‘calmodulin-dependent protein kinase activity(GO:0004683)’, ‘calcium-dependent protein serine/threonine kinase activity(GO:0009931)’와 인산화를 의미하는 ‘peptidyl-serine phosphorylation(GO:0018105)’에 대한 GO가 존재하였다. 탄수화물 결합인 ‘carbohydrate binding(GO:0030246)’와 핵산 결합인 ‘nucleic acid binding(GO:0003676)’이 더 많이 발현하였다. 특히하게도 결뿌리 또는 측근 형성과 관련된 ‘lateral root morphogenesis(GO:0010102)’가 초기 수구과에서 나타났다.

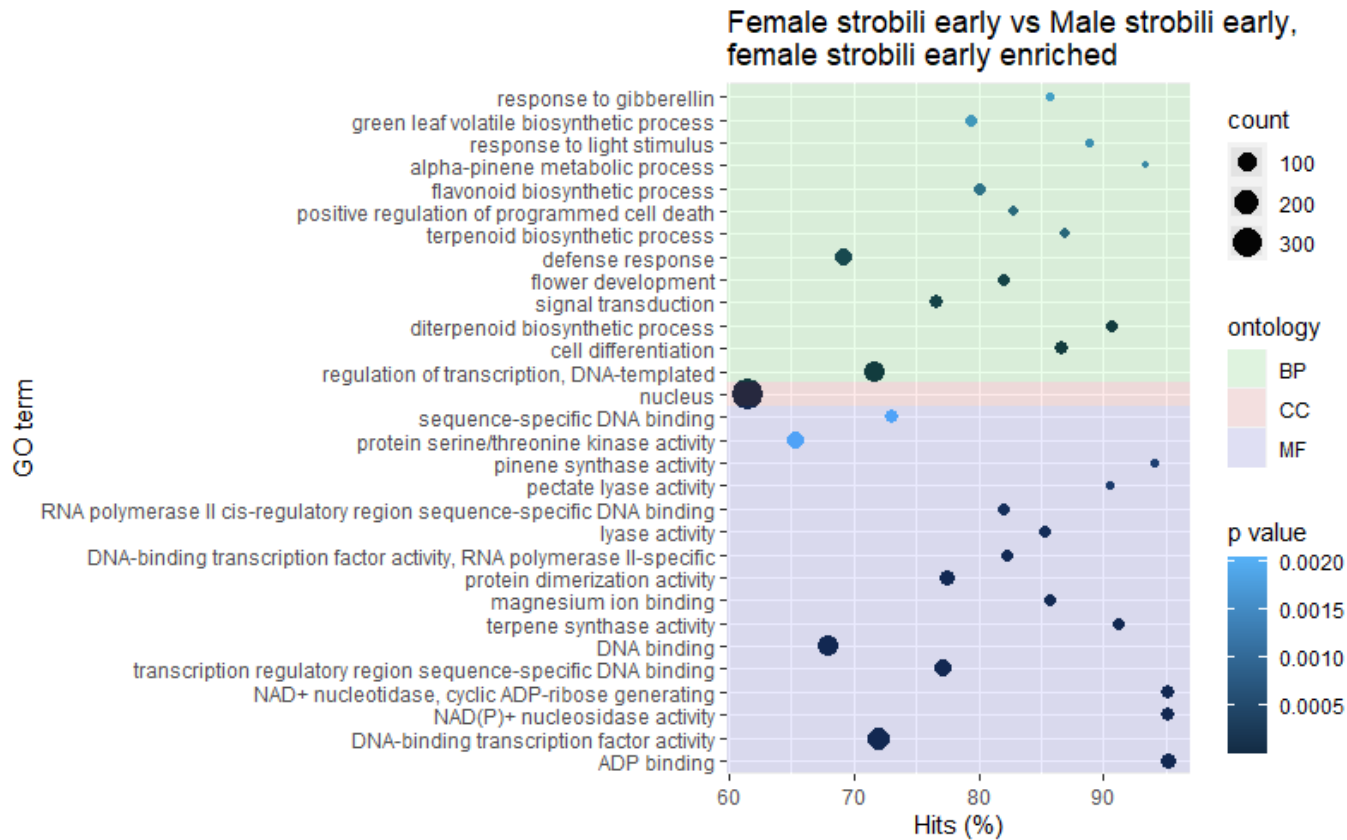


Figure 15. Top 30 GO terms in Gene Ontology analysis for F1 vs M1 comparison, F1 enriched

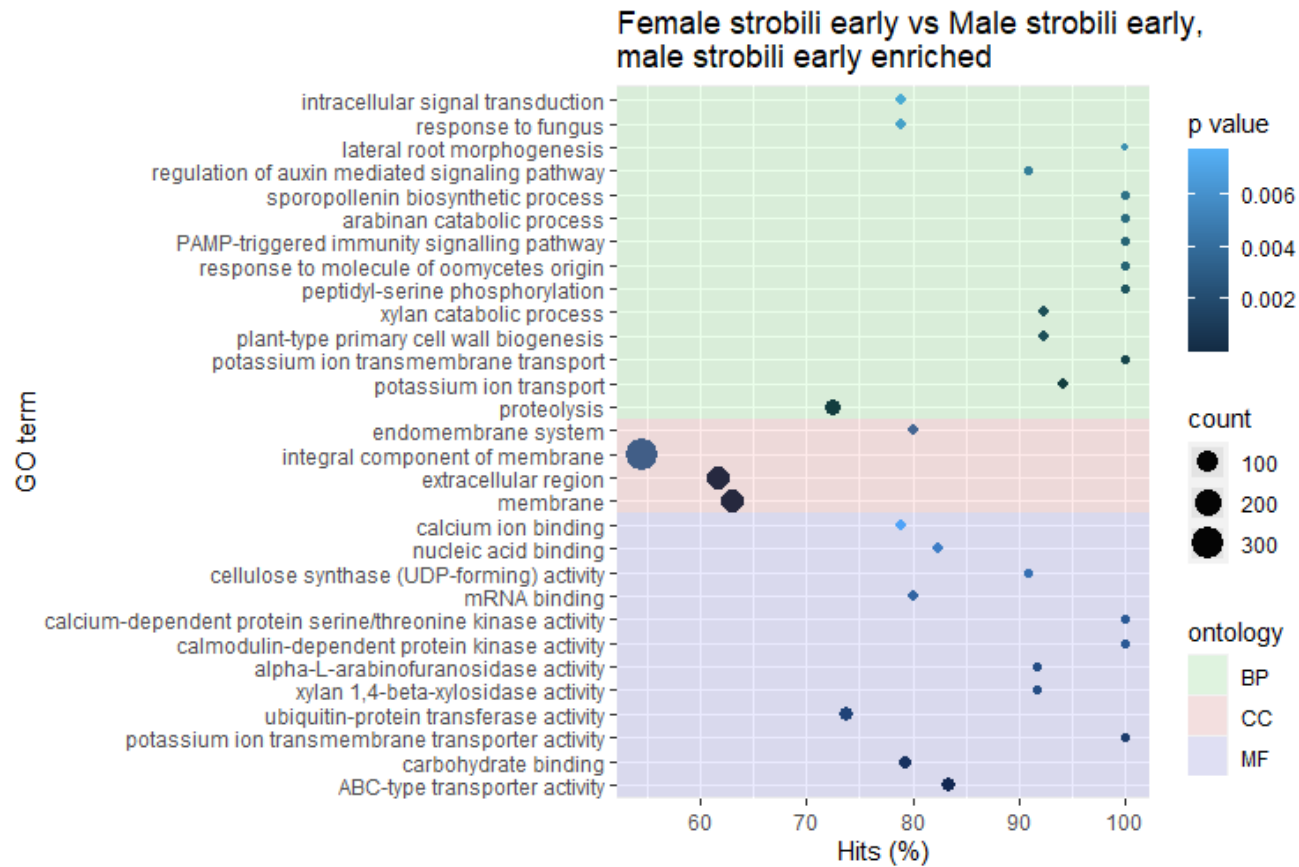


Figure 16. Top 30 GO terms in Gene Ontology analysis for F1 vs M1 comparison, M1 enriched

수구과 발달 과정을 비교하는 M1 vs M2의 경우, 초기 수구과(M1)에서 ‘diacylglycerol O-acyltransferase activity(GO:0004144)’, ‘long-chain-alcohol O-fatty-acyltransferase activity(GO:0047196)’, ‘arachidoyl-CoA:1-dodecanol O-acyltransferase activity(GO:0102966)’와 같이 O-acyltransferase의 활성이 높았다 (Figure 17). 지방산 합성과 관련된 ‘fatty acid biosynthetic process(GO:0006633)’, ‘very long-chain fatty acid biosynthetic process(GO:0042761)’, ‘triglyceride biosynthetic process(GO:0019432)’의 발현이 높았으며, 지질 분해를 나타내는 ‘lipid catabolic process(GO:0016042)’도 나타났다.

단백질 분해과정에 관여하는 ‘ubiquitin-protein transferase activity(GO:0004842)’와 효소의 활성을 의미하는 ‘transferase activity, transferring acyl groups other than amino-acyl groups(GO:0016747)’, ‘oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen(GO:0016705)’, ‘monooxygenase activity(GO:0004497)’가 존재하였다. 세포막을 통한 물질 수송과 관련된 ‘ATPase-coupled transmembrane transporter activity(GO:0004842)’, ‘ABC-type transporter activity(GO:0140359)’의 발현 빈도가 높았으며, ‘heme binding(GO:0020037)’, ‘DNA-binding transcription factor activity(GO:0003700)’와 같은 결합 현상이 나타났다. 또한 세포 소기관을 나타내는 ‘endoplasmic reticulum membrane(GO:0005789)’와 수분 부족에 대한 반응인 ‘response to water deprivation(GO:0009414)’, 수베린 합성 과정인 ‘suberin biosynthetic process(GO:0010345)’가 발현하였다.

중기 수구과(M2)에서는 체세포분열 과정과 관련된 ‘cell

division(GO:0051301)', 'phragmoplast(GO:0009524)', 'microtubule(GO:0005874)', 'microtubule binding(GO:0008017)', 'microtubule motor activity(GO:0003777)', 'microtubule-based movement(GO:0007018)', 'cell cycle(GO:0007049)', 'spindle(GO:0005819)', 'mitotic cell cycle(GO:0000278)', 'mitotic cell cycle phase transition(GO:0044772)', 'nuclear microtubule(GO:0005880)', 'regulation of mitotic spindle organization(GO:0060236)', 'mitotic spindle(GO:0072686)'의 빈도가 높았다(Figure 18).

세포분열 시 DNA 형태 변화와 관련된 'DNA topoisomerase type II (double strand cut, ATP-hydrolyzing) activity(GO:0003918)', 'DNA topological change(GO:0006265)', 'double-strand break repair via homologous recombination(GO:0000724)', 'DNA repair(GO:0006281)', 'double-stranded DNA binding(GO:0003690)'이 나타났다. ATP 결합을 의미하는 'ATPase activity(GO:0016887)', 'DNA-dependent ATPase activity(GO:0008094)', 'ATP binding(GO:0005524)'이 많이 발현하였으며, 인산화효소 활성을 나타내는 'cyclin-dependent protein kinase holoenzyme complex(GO:0000307)', 'regulation of cyclin-dependent protein serine/threonine kinase activity(GO:0000079)', 'cyclin-dependent protein serine/threonine kinase regulator activity(GO:0016538)', 'protein kinase activator activity(GO:0030295)'가 존재했다. 또한 식물에서 합성되는 유기화합물인 'phenylpropanoid metabolic process(GO:0009698)'가 나타났다.

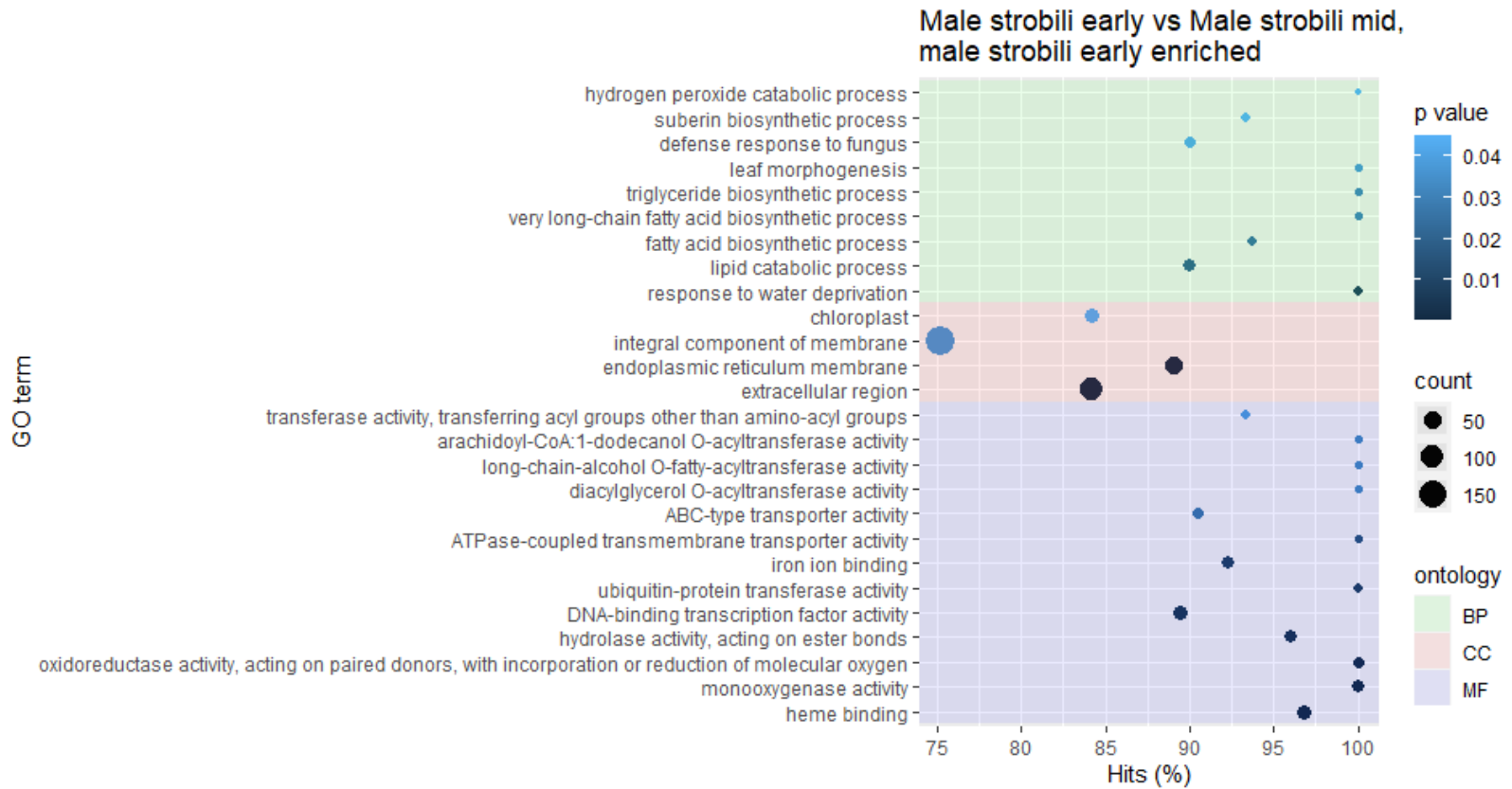


Figure 17. Top 30 GO terms in Gene Ontology analysis for M1 vs M2 comparison, M1 enriched

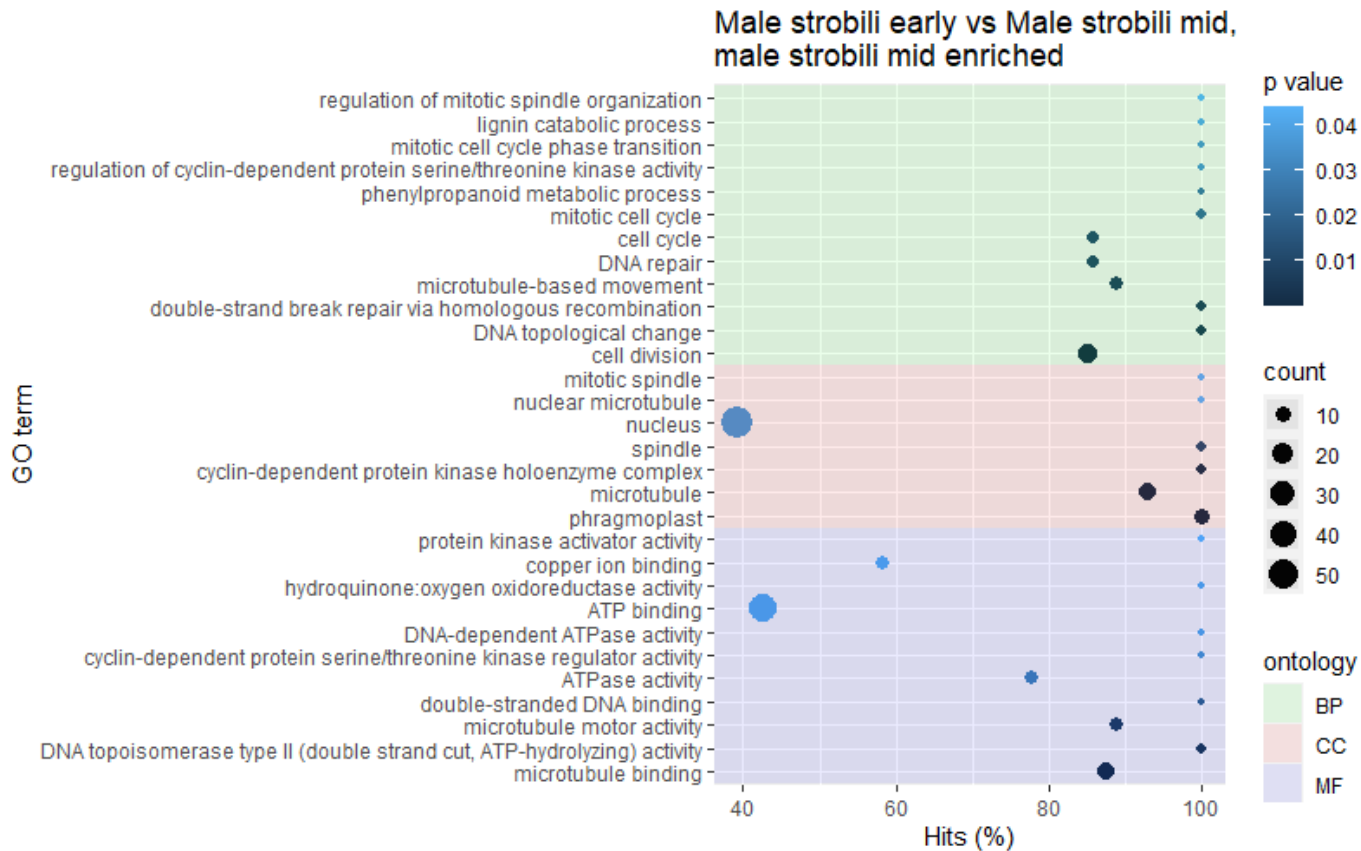


Figure 18. Top 30 GO terms in Gene Ontology analysis for M1 vs M2 comparison, M2 enriched

제4절 소나무의 후보 개화 조절 유전자 탐색 및 기관별 비교

GO enrichment 분석 결과를 바탕으로 개화 조절과 직접적으로 연관된 GO term들을 추출하였다. 그리고 해당 GO term에 포함되어 있는 전사체 서열들을 대상으로 하여 이미 알려진 다른 수종들의 개화 조절 단백질을 중심으로 각각의 상동성을 확인하였다. ‘F1 vs F3, F1 enriched’ 카테고리에서 ‘petal morphogenesis(GO:0048446)’와 ‘petal development(GO:0048441)’에 해당하는 전사체들의 서열을 추출하여 blastx를 실행한 결과, 총 8개의 서열 중 7개 서열이 *Wax Ester Synthase/Diacylglycerol Acyltransferase(WSD)*와 가장 유사하였고, basic Helix-Loop-Helix(bHLH) 전사인자와 유사한 서열이 1개 존재했다. 하지만 직접적으로 개화 조절 단백질과 관련된 서열은 나타나지 않았다(**Table S1**).

‘F1 vs M1, F1 enriched’ 카테고리에서 ‘flower development(GO:0009908)’에 해당하는 전사체들의 서열을 추출하여 blastx를 실행한 결과, 총 48개의 서열 중 2개가 *AP2*, 9개가 *DEFICIENTS AGAMOUS-LIKE(DAL)*, 13개가 *MADS*, 7개가 *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1(SOC1)*, 2개가 *UFO*, 2개가 *WUSCHEL(WUS)*, 1개가 *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE(SPL)*와 가장 유사하였다. 또한 3개가 Gibberellin 20-oxidase(GA20ox), 2개가 peroxidase와 유사하였으며, AINTEGUMENTA, AT-hook protein, GATA, HEADING DATE REPRESSPR, SODIUM POTASSIUM ROOT DEFECTIVE, R2R3MYB1, Snf1 kinase interactor와 유사한 서열은 각각 1개씩 나타났다(**Table S2**).

‘F1 vs M1, M1 enriched’ 카테고리에서 ‘sporopollenin biosynthetic process(GO:0080110)’에 해당하는 전사체들의 서열을 추출하여 blastx를 실행한 결과, 총 20개의 서열 중 6개가 ABC transporter, 2개가 chalcone synthase, 5개가 cinnamoyl-CoA reductase, 2개가 flavonoid 3’ hydroxylase, 2개가 tetraketide alpha-pyrone reductase 단백질과 가장 유사하였고, acyl-CoA synthetase와 cytochrome P450과 유사한 서열이 각각 1개씩 나타났다. 하지만 직접적인 개화 조절 단백질과 관련된 서열은 나타나지 않았다(**Table S3**).

각 개화 조절 단백질에 대해 스코어(score)가 가장 높은 서열을 대표적으로 추출하여 이들을 소나무의 후보 개화 조절 유전자로 선정하였다(**Table 5**). 최종적으로 소나무 구과 발달 과정에 관여하는 7종류의 후보 개화 조절 유전자를 발견하였으며, 그 종류는 *AP2L*(*AP2L1*, *AP2L3*), *DAL*(*DAL3*, *DAL4*, *DAL10*, *DAL19*), *MADS*(*MADS3*, *MADS4*, *MADS5*, *MADS6*, *MADS7*), *UFO*(*UFO1*, *UFO2*), *SOC1*, *SPL3*, *WUS*이다.

Table 5. BLASTX best-hit results of candidate flowering regulatory genes of *Pinus densiflora*

| Species | Protein | Accession | Transcript ID | Identity (%) | E value | Score (bits) |
|---------------------------|---------|------------|---------------------------|--------------|---------|--------------|
| <i>Pinus tabuliformis</i> | AP2L1 | AJP06234.1 | TRINITY_DN2887_c0_g3_i1 | 99 | 0 | 753 |
| <i>Pinus tabuliformis</i> | AP2L3 | AJP06236.1 | TRINITY_DN2887_c0_g2_i2 | 99 | 0 | 1170 |
| <i>Pinus tabuliformis</i> | DAL10 | AJP06266.1 | TRINITY_DN5374_c0_g1_i5 | 100 | 0 | 530 |
| <i>Picea abies</i> | DAL19 | AGR53805.1 | TRINITY_DN7611_c0_g1_i11 | 71 | 1E-92 | 192 |
| <i>Pinus tabuliformis</i> | DAL3 | AJP06276.1 | TRINITY_DN1774_c1_g1_i23 | 76 | 4E-123 | 363 |
| <i>Pinus tabuliformis</i> | DAL4 | AJP06277.1 | TRINITY_DN7611_c0_g1_i7 | 98 | 7E-136 | 407 |
| <i>Pinus tabuliformis</i> | MADS1 | AJP06319.1 | TRINITY_DN14061_c1_g2_i1 | 99 | 6E-166 | 481 |
| <i>Pinus tabuliformis</i> | MADS3 | AJP06321.1 | TRINITY_DN8544_c0_g2_i5 | 99 | 4E-171 | 498 |
| <i>Pinus tabuliformis</i> | MADS4 | AJP06322.1 | TRINITY_DN10705_c0_g1_i19 | 61 | 7E-40 | 150 |
| <i>Pinus radiata</i> | MADS5 | AAB80808.1 | TRINITY_DN14326_c0_g1_i1 | 92 | 8E-134 | 388 |
| <i>Pinus tabuliformis</i> | MADS6 | AJP06324.1 | TRINITY_DN3223_c1_g1_i9 | 96 | 4E-124 | 375 |
| <i>Pinus tabuliformis</i> | MADS7 | AJP06325.1 | TRINITY_DN3223_c1_g1_i15 | 97 | 1E-167 | 476 |
| <i>Pinus taeda</i> | SOC1 | AZA14799.1 | TRINITY_DN1774_c1_g1_i25 | 92 | 5E-141 | 411 |
| <i>Larix kaempferi</i> | SPL3 | UXE52212.1 | TRINITY_DN43672_c0_g1_i2 | 88 | 2E-89 | 285 |
| <i>Pinus tabuliformis</i> | UFO1 | AJP06363.1 | TRINITY_DN15623_c0_g1_i1 | 99 | 0 | 791 |
| <i>Pinus tabuliformis</i> | UFO2 | AJP06362.1 | TRINITY_DN27618_c0_g1_i2 | 99 | 0 | 779 |
| <i>Pinus pinaster</i> | WUS | ALN42231.2 | TRINITY_DN12004_c0_g1_i2 | 98 | 6E-154 | 444 |

소나무에서 발견한 7종류의 후보 개화 조절 유전자들의 TMM 발현값을 각각 추출하여 기관별로 비교하였다(Figure 19). 이때 후기 수구과(M3)는 앞서 진행한 샘플간 상관분석에서 다른 수구과 샘플들과 낮은 유사성을 보였으므로 결과 해석에서 제외하였다.

AP2L 유전자들은 정단과 암구과에서 높은 발현을 보였다. 그중 *AP2L3*는 잎, 뿌리, 줄기에서도 발현 정도가 상당히 높았다. *DAL*에 속한 유전자들은 그 발현 정도가 서로 다르게 나타났다. *DAL10*은 암구과에서 특히 많이 발현하였고 *DAL19*는 정단, 잎, 줄기에서 상대적으로 많이 발현하였다. 그리고 *DAL4*는 정단, 잎, 뿌리와 줄기에서 높게 발현하였다.

*MADS*에 속한 유전자들 또한 서로 다른 발현 양상을 보였다. *MADS1*은 암구과에서만 발현하는 것으로 나타났고, *MADS3*은 정단과 암구과에서 높게 발현하였다. 그리고 *MADS6*은 정단과 잎, 줄기에서 발현하였으며 뿌리에서 가장 많이 발현하는 것으로 나타났다. *SOC* 유전자는 정단, 잎, 줄기에서 높은 발현을 보였고, *SPL* 유전자는 정단, 암구과, 잎에서 높게 발현하였으며 뿌리와 줄기에서도 발현됨을 확인하였다. *UFO* 유전자들 중에서 *UFO1*은 정단과 잎에서 가장 많이 발현하였고 *UFO2*는 정단과 암구과에서 특히 높게 발현하였다. 마지막으로 *WUS* 유전자는 암구과에서 특히 많이 발현하는 것으로 나타났다.

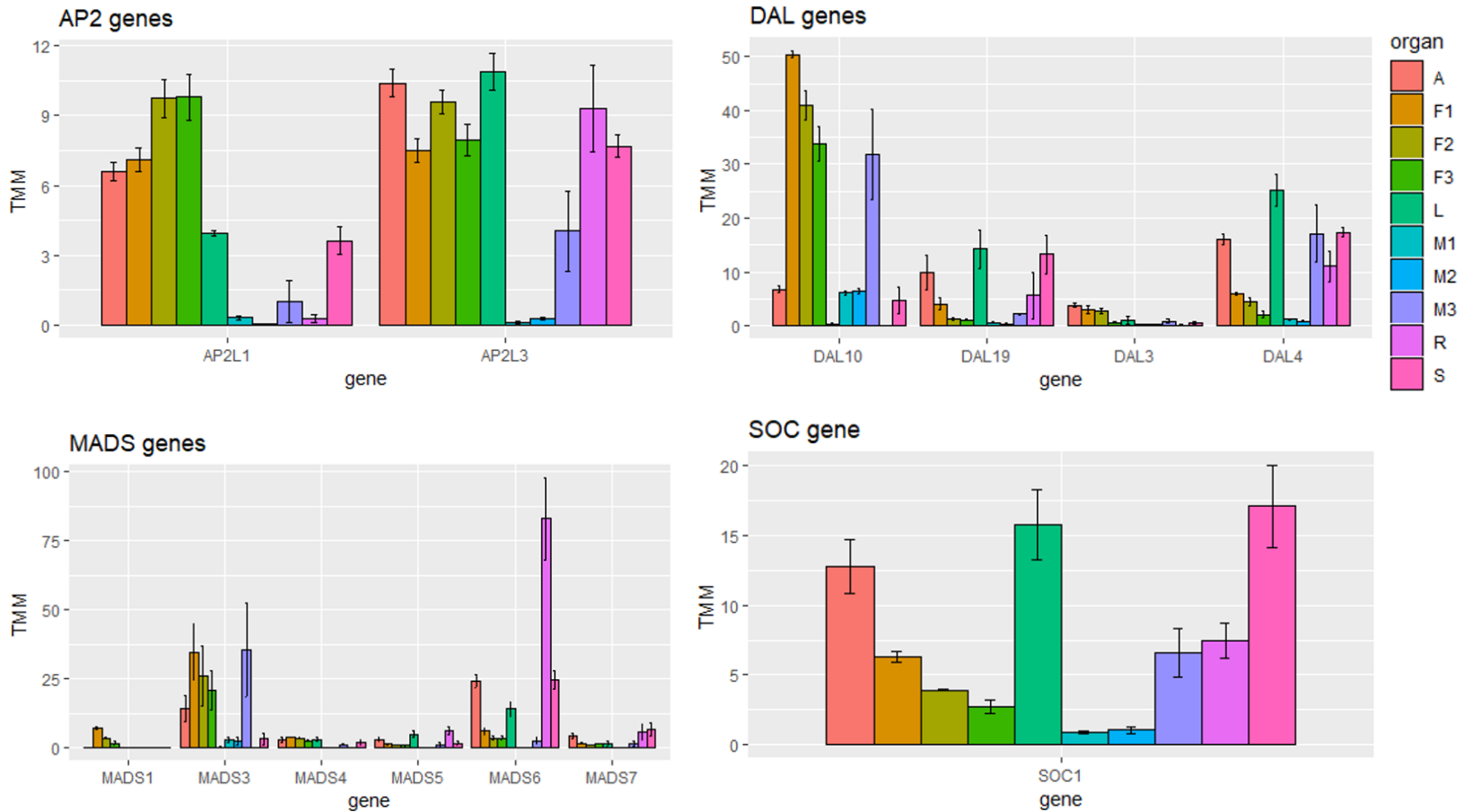


Figure 19. (continued)

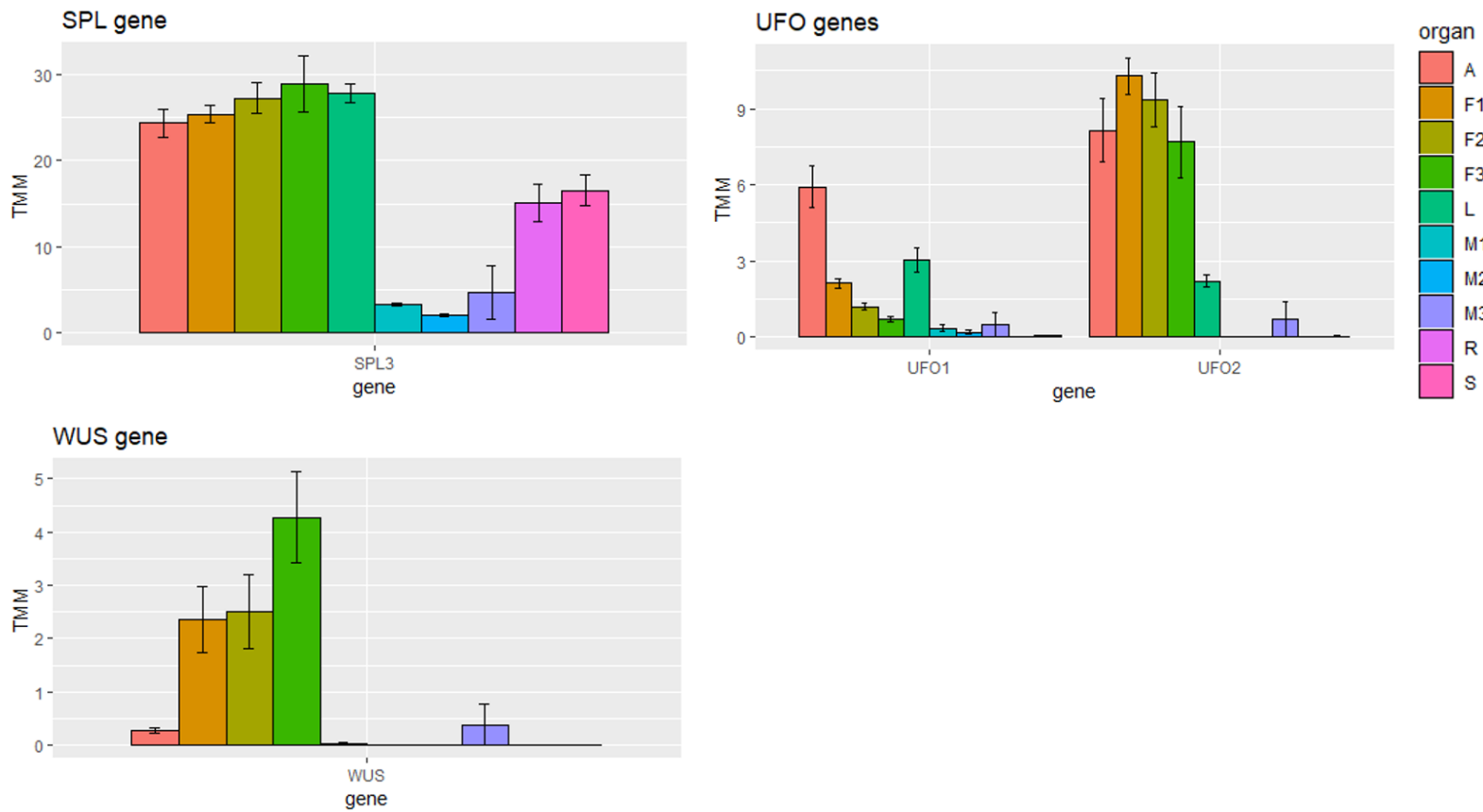


Figure 19. Expression of candidate flowering regulatory genes between reproductive organs of *Pinus densiflora*

본 연구에서 발견한 총 17개의 후보 개화 조절 유전자들의 발현값을 사용하여 PCA plot을 생성하였다(Figure 20). 소나무의 암구과와 수구과 샘플들의 경우 세 번의 발달 단계(초기, 중기, 후기)를 모두 합쳐 각각 하나의 기관으로 분석하였다. 최종적으로 해당 PCA plot 결과를 통해 후보 유전자들의 발현값으로 소나무의 각 기관 조직들을 구분할 수 있음을 확인하였다.

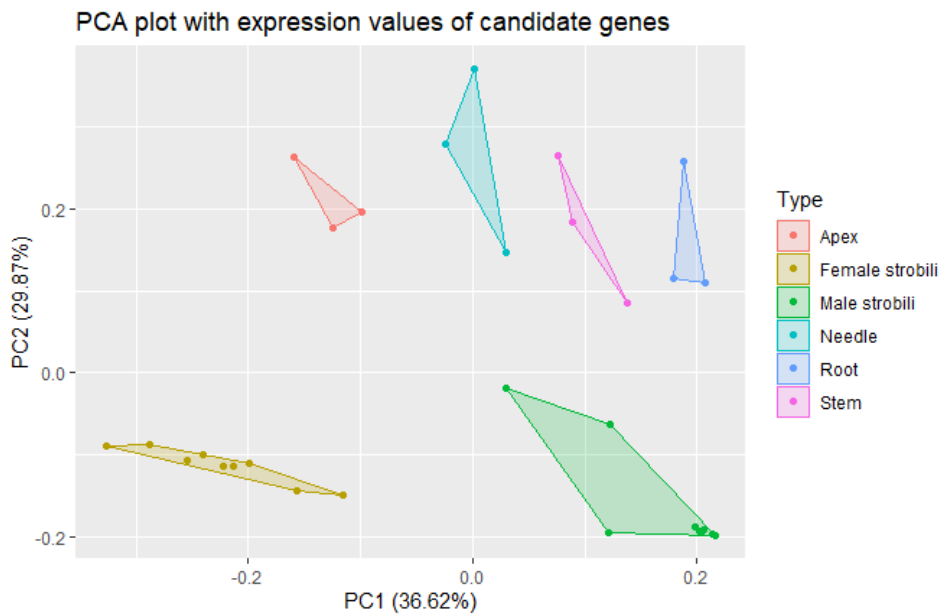


Figure 20. PCA plot with expression values of the candidate flowering regulatory genes of *Pinus densiflora*

제5절 DEG 발현량에 대한 qRT-PCR 검증

본 연구에서 RNA-seq 데이터를 이용하여 전사체 발현 정도를 정확하게 측정하였음을 확인하고 유전자 발현 분석을 검증하기 위해, 12개의 DEG 서열을 임의적으로 선정하여 qRT-PCR로 증폭하였다(Figure 21). 해당 서열들에 대한 qRT-PCR 증폭 값과 RNA-seq 발현량을 비교한 결과, 상관계수(R^2) 0.74로 각 샘플에서의 발현 패턴이 서로 유사하다는 것을 확인하였다. 즉 RNA-seq 데이터를 분석하여 해석한 본 연구에서의 결과들을 신뢰할 수 있다고 판단하였다.

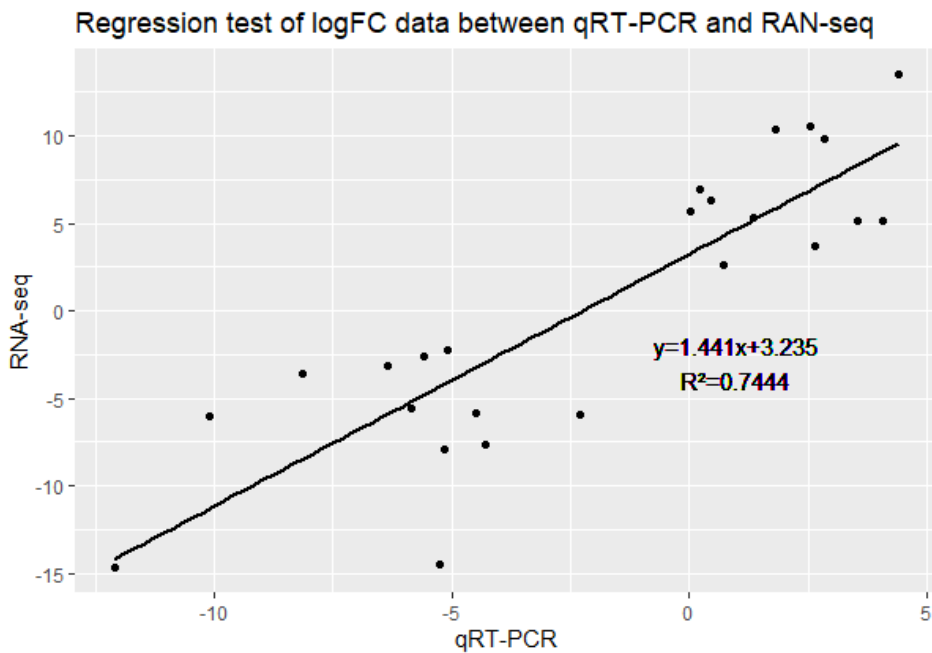


Figure 21. Validation of RNA-sequencing results by qRT-PCR

제5장 고 찰

제1절 소나무 구과 발달 과정의 생물학적 및 생리적 특성

<소나무 정단 분열조직에서의 신호전달 및 방어작용>

세포예정사(Programmed Cell Death, PCD)는 식물의 영양생장 단계와 생식생장 단계에서 모두 발견되는 현상으로 성 결정, 배우체 발달, 배 발생에 필수적인 요소이다(Wan et al., 2002). 식물체 전체의 노화와 상관없이 잎, 꽃잎, 줄기 정단 등의 특정 부분에서 PCD가 일어날 수 있다(Xu and Hanson, 2000). 양성화를 만들지 않는 겉씨식물의 경우 선택적으로 암구과 또는 수구과를 결정하여 구과 원시세포(cone primordia)를 생산해야 하는데, 이때 PCD가 중요한 역할을 하는 것으로 추정된다.

식물은 병원체의 공격에 대응하기 위한 방어 메커니즘을 갖도록 진화하였으며, 미생물 등의 병원성 신호가 인식되면 즉각 유도된다. 이때 식물은 감염 부위뿐만 아니라 시스템적으로 식물체 전체에 대한 반응이 가능하다(Tornero et al., 1997). 식물에 침입하는 병원체를 감지하는 데 *Resistance(R)* 유전자가 중요한 역할을 한다. 식물의 *R* 유전자와 병원체의 *avirulence(avr)* 유전자가 상응하게 되면 병원체가 인식되고 식물의 방어작용이 활성화된다(Baker et al., 1997; Bonas and Anckerveken, 1999). 병원체가 감지된 이후 감염 부위와 인접한 조직에서 병원체의 생장을 제한하기 위해 부분적으로 조직이 괴사하는데, 이를 과민반응(Hypersensitive Response, HR)이라고 한다(Matthews, 1991). HR이 시작되면 이후 발생할 다양한 병원체 감염에 대한 더 강

한 저항성을 갖기 위해 여러 신호전달 과정이 활성화된다. 이를 전신획득저항성(Systemic Acquired Resistance, SAR)이라 하며 공존할 수 없는 식물과 병원체의 상호작용에서 흔하게 발생한다(Ross, 1961). SAR 과정에서 발현되는 유전자들은 방어와 관련된 단백질들을 암호화하는데, Pathogenesis-Related(PR) 단백질이 여기에 해당된다(Ward et al., 1991).

식물 호르몬 중 하나인 에틸렌은 식물의 다양한 발달 과정과 스트레스 반응에 중요한 역할을 담당하며 줄기 정단 및 엽액 분열조직과 생식 분열조직에서 끝눈(terminal bud)을 형성하고 줄기 정단 분열조직(Shoot Apical Meristem, SAM)의 활성을 일시적으로 억제한다(Bleecker and Kende, 2000; Ruonala et al., 2006). 에틸렌은 꽃의 성장에서부터 꽃잎의 노화에 이르기까지 개화의 전반적인 과정에 관여한다(Reid and Wu, 1992).

<소나무 암구과 발달 과정에서의 물질 대사>

원엽록소 산화환원효소(Protochlorophyllide OxidoReductase, POR)는 식물 발달에 중요한 효소이다. POR은 빛을 사용하여 프로토클로로필리드(ProtoChlorophyllide)를 클로로필리드(Chlorophyllide)로 환원시킴으로써 엽록소 합성의 후기 과정에서 촉매 역할을 한다(Begley and Young, 1989; Wilks and Timko, 1995). 엽록소는 광합성 작용의 구성요소이고 빛을 모아서 에너지를 변환하는 기능을 하므로, 잎과 줄기에 상당한 양의 엽록소가 존재한다. 꽃을 생산하는 식물의 경우에는 일반적으로 꽃잎이 성숙하면서 엽록소의 함량은 낮아지며 안토시아닌, 카로티노이드 등 다른 색소가 축적되지만, 꽃잎 발달 단계 초기에는 엽록소가 존재하는 것으로 알려져 있다(Tanaka et al., 2008; Ohmiya et al., 2014). 본 연구에서 성장 초기의 정단과 비교했을 때 초기 암구과에서

원엽록소 산화환원효소의 발현이 높았으므로 소나무의 경우 정단보다 암구과 발달 초기에 상대적으로 엽록소 함량이 많아진다고 추정할 수 있다.

지방산은 세포막의 필수 구성요소일 뿐만 아니라, 종자와 같은 저장조직의 주요 에너지 저장소이다(Mu et al., 2008). 지방산은 식물체에서 잎, 줄기, 꽃과 종자 등 다양한 조직과 기관에서 발견되며(Aziz and Mitu, 2019), 특히 꽃 구조에서는 꽃잎, 자방, 암술, 꽃밥에 지방산이 존재한다. 꽃 형성 초기에 암술의 총 지방산 함량이 가장 높고, 꽃이 발달하면서 암술의 지방산 함량은 낮아지지만 다른 기관들보다는 항상 높게 나타난다. 꽃잎에서도 꽃이 성숙하면서 지방산의 함량이 낮아지지만, 자방에서는 그 함량이 거의 변하지 않는다(Matsuzaki et al., 1983).

단성화와 양성화 식물에서 성 발현은 유전적으로 결정되며 무기양분, 광주기, 온도, 식물호르몬 등 환경적인 요인에 의해서도 변화가 가능하다(Kumar, 1983; Durand and Durand, 1984). 이 중에서 식물호르몬은 성전환에 있어 가장 효과적인 요인으로 여겨지는데(Malik and Bhattacharya, 1979), 호르몬에 의해 다양한 효소의 발현에 차이가 생기며 이는 성결정과 관련된 유전자의 활성화에 차이를 가져온다(Jaiswal et al., 1984; Lal and Jaiswal, 1987). 애기장대의 꽃에서 많은 종류의 Purple Acid Phosphatase(PAP) 효소가 나타나는데, 그중 *AtPAP23*은 꽃분열조직에서 우세하게 발현하지만 완전히 발달한 꽃에서는 꽃잎과 수술대에서만 발현한다(Zhu et al., 2005). 이처럼 식물호르몬에 의해 생성된 여러가지 효소들은 그 종류에 따라 꽃 발달에 중요한 역할을 한다고 할 수 있다.

정유(Essential Oil, EO)는 식물로부터 추출한 휘발성이 있는 방향성 화합물이 포함된 식물 농축액이다. 정유는 일반적으로 잎과 줄기의 특정 세포에서 생성되며 잎과 수피, 과실과 같은 특정 부분에 밀집되어 있다(Oussalah et al., 2006). 페닐프로파노이드(phenylpropanoid)는 식물

정유의 주요한 구성요소 중 하나이다(Friedrich, 1976). 이는 화분 매개자를 유혹하는 역할을 하며, 미생물과 곤충의 접근을 제한하고 살균 작용도 있어 식물의 생식 기관을 보호할 수 있다(De Moraes et al., 2001; Friedman et al., 2002; Hammer et al., 2003; Reinhard et al., 2004). 또한 페닐프로파노이드, 아세트산 벤질(benzyl acetate) 등 다양한 휘발성 복합물들의 조합에 따라 꽃의 향기가 결정된다(Raguso and Pichersky, 1995; Dudareva et al., 2003; Shalit et al., 2004).

액포는 식물의 발달에 중요한 여러가지 기능을 수행하는 기관이다. 식물 액포는 색소를 포함하여 다양한 이온과 대사산물 등을 저장하고 해독 작용을 수행함으로써 세포의 항상성을 유지하는 데 필수적인 역할을 한다(Marty, 1999). 안토시아닌은 식물 조직에 빨간색, 파란색, 보라색을 나타나게 하는 플라보노이드(flavonoid) 색소이다(Passeri et al., 2016). 안토시아닌은 합성된 후 액포 내부로 이동하여 저장되며 이러한 구조를 Anthocyanic Vacuolar Inclusion(AVI)라고 하는데, 많은 식물들의 꽃잎 구조에서 발견된다(Markham et al., 2000; Grotewold, 2006; Zhang et al., 2006).

식물이 영양생장 단계에서 생식생장 단계로 전환되면서 대사작용에 활발한 변화가 생기며, 특히 꽃을 생성하는 과정에서는 탄수화물과 아미노산 등 다양한 물질의 이동에 의한 대사적 재배열이 일어나야 한다(Borghesi et al., 2022). 인(Phosphorus, P)은 식물의 생장에 있어 필수적인 대량 영양소 중 하나로, 토양에 높은 농도로 존재하지만 식물이 사용가능한 형태인 무기인산(inorganic Phosphate, Pi)은 아주 낮은 농도로 토양에 포함되어 있다(Vance et al., 2003). 따라서 식물은 무기인산의 항상성을 유지하기 위해 형태적 및 대사적 변화를 통해 무기인산이 제한적인 환경에 적응하여 살아남을 수 있게 된다(Abel et al., 2002; Rouached et al., 2010). 이처럼 식물은 양분의 농도에 민감하게 반응하

는데, 낮은 인산염(Low Phosphate, LP) 조건에서 식물의 개화 시기가 지연된다는 연구 결과가 보고되었다(Ren et al., 2019).

아포플라스트(apoplast)는 기체와 수분으로 채워진 세포사이 공간으로서 다양한 기능을 수행한다. 양분과 수분을 이동시키고 세포벽을 구성하며 다양한 물질 합성에 관여하는데, 특히 생물적 및 비생물적 스트레스에 대응하는 방어작용과 관련된 물질들을 합성하는 장소이다(Farvardin et al., 2020). 한편 식물은 개화를 통해 화분을 날리며 수분과 수정 과정을 거쳐 최종적으로 종자를 생산한다. 대부분의 식물에서 꽃은 영구적으로 열려 있지만 일부 수종에서는 닫힘 현상이 일어나기도 하며, 꽃잎이 시들어 떨어지면서 개화 현상이 종결된다. 이처럼 개화는 빠른 속도의 세포 팽창과 외부적 및 내부적 요소들의 복잡한 조절 과정이 수반되어 일어난다. 세포가 팽창하기 위해서는 삼투성 용질의 함량이 증가해야 하는데, 이를 위해서는 다당류가 단당류로 전환되거나 아포플라스트로부터 당을 흡수하는 과정이 필요하다(van Doorn and van Meeteren, 2003).

옥시리핀(oxylin)은 지방산의 이화작용에 의해 생성된 산화 물질이며, 식물에서 다양한 방어 작용과 발달 과정을 조절한다. 옥시리핀의 한 종류인 자스몬산(Jasmonic Acid, JA)은 식물의 노화를 촉진하고 성장을 조절한다(Creelman and Mulpuri, 2002). JA 함량은 성장하고 있는 어린 조직과 꽃 분열조직, 꽃 줄기 등 꽃 기관의 특정 부위에서 높게 나타난다(Hause et al., 2000). JA뿐만 아니라 JA의 대사산물까지 모두 통틀어 자스모네이트(jasmonates)라고 부르는데, 식물 호르몬과 함께 스트레스 반응에서 활성을 나타내며 식물의 발달 과정에서 뿌리 성장, 꽃 기관 발달, 노화 등에 관여한다(Browse, 2009; Wasternack et al., 2013). 특히, 자스모네이트의 농도가 높아지면 꽃 발달에 관여하는 유전자들의 발현에 변화가 생기면서 꽃과 종자의 성장에 영향을 준다

(Mandaokar et al., 2006).

<소나무 수구과에서의 화분 형성 과정>

종자식물의 배우체 단계는 감수분열 완료 후에 시작되며, 동물과 달리 식물의 감수분열 산물은 유사분열을 겪어 배우자를 포함하고 있는 다세포 구조인 배우체를 형성한다. 이때 수배우체는 화분립, 그리고 암배우체는 배낭이라고 부른다. 수분 현상 이후 접합자가 형성되는 수정이 일어나면 배우체 단계는 끝나고 다시 포자체 단계가 시작된다(Chen and McCormick, 1996). 개화 식물에서 화분이 형성되는 과정은 여러 단계가 필요하다(McCormick, 1993). 먼저 화분 모세포가 감수분열을 겪어 네 개의 소포자를 생성한다. 단핵 소포자는 빠르게 크기가 증가하여 비대칭적 유사분열을 겪어 영양세포와 생식세포 두 종류의 세포를 생산한다. 영양세포는 화분립에 필요한 대부분의 단백질을 제공하며, 화분 발아 과정 동안 화분관을 형성하면서 더 이상의 세포 분열은 일어나지 않는다. 생식세포는 영양세포의 세포질로 둘러싸여 있으며, 한 번 더 분열하여 두 개의 정자세포를 생산한다. 즉 성숙한 화분은 하나의 영양세포와 또 다른 영양세포의 세포질로 감싸진 두 개의 정자세포로 구성된다.

대부분의 식물 수종에서 두 번째 유사분열은 화분이 꽃밥에서 날리기 전에 화분관 내에서 발생한다. 유사분열 과정에서는 방추 미세소관의 역할이 중요한데, 미세소관은 염색체와 상호작용하여 동원체에 부착해서 염색체 분리에 필요한 동력을 만들어낸다(McIntosh et al., 2002). 본 연구에서는 중기 수구과 발달 과정에서 유사분열 및 방추 미세소관 활동과 관련된 유전자 기능이 많이 나타났으며, 이를 통해 소나무 화분 형성 과정에서 유사 분열로 인한 염색체 분리가 많이 일어났음을 확인하였다.

식물이 정상적인 수분과 수정을 하기 위해서는 꽃밥이 적절한 시기에 열개하여 성숙한 화분립을 내보내는 과정이 필요하다. 이때 옥신이 꽃밥

의 열개 시기를 조절하는 중요한 역할을 한다고 알려져 있다(Cardarelli and Costantino, 2018). 애기장대에서는 옥신의 농도가 꽃밥의 열개 직전에 가장 높고, 내피(endothecium)의 리그닌화가 일어날 때 감소하는 것으로 나타났다. 또한 옥신은 자스몬산의 생합성을 조절함으로써 구변세포(stomium)의 열림을 제한하여 꽃밥 열개에 영향을 주는 것으로 확인되었다(Cecchetti et al., 2013). 여기서 자스몬산은 꽃밥 열개의 최종 단계에 관여하며, 화분 성숙과 수술대 신장을 유도하는 것으로 알려져 있다(Cecchetti et al., 2008). 소나무의 경우 초기 수구과 발달 단계가 옥신에 의해 조절되는 것으로 나타났으며, 이를 통해 소나무에서도 옥신과 자스몬산이 화분 발달 및 수분 현상에 중요한 역할을 하는 것으로 판단하였다.

식물의 세포벽을 구성하고 있는 헤미셀룰로스(hemicellulose)는 지구상에 풍부하게 존재하는 탄소원 중 하나이다(Thomson, 1993). 자일란(xylan)은 헤미셀룰로스의 주된 구성성분으로 beta-1,4-결합으로 연결되어 있는 다당류이다(Coughlan and Hazlewood, 1993; Damásio et al., 2011). 자일란 분해를 담당하는 효소 중 하나인 xylan 1,4-beta-xylosidase는 소포자와 화분의 발달 과정에 관여한다고 알려져 있다. 해당 효소에 대한 유전자가 소포자의 감수분열 이후 초기에 발현하여 세포벽을 느슨하게 함으로써 화분 팽창을 유도하며, 후기에 다시 발현하면서 화분 발아 과정 동안 세포를 팽창시키는 역할을 한다(Hrubá et al., 2004).

제2절 소나무 개화 조절 유전자의 발현 특성

생식생장 과정에 대한 많은 연구가 이루어진 속씨식물과 비교했을 때, 겉씨식물의 경우에는 생식생장을 조절하는 전반적인 대사과정에 대한 이해가 매우 부족한 편이다. 하지만 애기장대와 같은 모델식물에서 발견된 일부 유전적 메커니즘이 침엽수를 포함한 목본 식물에게도 보존되어 있음이 연구되었다(Uddenberg et al., 2013; Wendling et al., 2014). 애기장대에서는 수십개의 유전자가 개화 시기를 직접적으로 조절한다고 알려져 있는데(Srikanth and Schmid, 2011), 이중 일부는 침엽수에도 존재하는 것으로 밝혀졌다(Purugganan, 1997; Sundstrom et al., 1999; Carlsbecker et al., 2004). 겉씨식물 모델에 따르면, B와 C-class 유전자가 공동 발현하면 수구과 발달을 유도하고, C-class 유전자만 발현하면 암구과 발달을 유도하는데, 두 경우 모두 A-class 단백질을 포함해야 한다(Theiben et al., 2016). 이처럼 겉씨식물에서도 MADS-box 단백질의 다양한 물리적 상호작용에 의해 생식기관이 형성된다고 할 수 있다(Niu et al., 2022). 본 연구에서는 다른 소나무속의 수종들과 상동성이 높은 여러 MADS 후보 유전자들(MADS1,3,4,5,6,7)을 소나무에서 발견하였으며 정단, 암구과뿐만 아니라 잎, 줄기, 뿌리와 같은 영양기관에서도 나타났다. 만주곰솔(*P. tabuliformis*)의 *PtMADS11*은 생식생장으로의 전환을 조절하는 역할을 하는 것으로 확인되었는데(Ma et al., 2021), 이를 통해 소나무에서도 생식생장을 유도하는 다양한 MADS 유전자들이 존재한다고 추정할 수 있다.

*FLOWERING LOCUS T(FT)*와 *SOC1*은 다양한 개화 조절 과정에서 발생하는 신호들을 통합하는 역할을 하는데, 여기서 FT 단백질은 모든 개화 식물에 보존되어 있는 개화 유도 신호 물질이다(Samach et al., 2000; Yang et al., 2007; Turck et al., 2008; Amasino and Michaels,

2010). 애기장대의 경우 광주기성 개화 과정에서 *CO(CONSTANS)*가 *FT*의 발현을 촉진하고 *FT*는 *SOC1*의 발현을 촉진함으로써 꽃분열조직에서 다양한 개화 조절 유전자들을 활성화시키는데, 이때 *LFY*의 발현을 직접적으로 유도한다(Liljegren et al., 1999; Wigge et al., 2005; Yoo et al., 2005). 본 연구에서는 소나무의 *SOC1* 후보 유전자를 발견하였으며 정단과 잎, 줄기에서 높은 발현 양상을 보였다. 애기장대의 *SOC1* 유전자는 *MADS-box* 전사인자를 암호화하며 정단 분열조직에서 개화를 유도한다고 알려져 있다(Parenicova et al., 2003). 게다가 *SOC1*은 성장하는 잎과 줄기에서도 상당한 정도로 발현되며 식물이 발달함에 따라 증가한다고 보고되었다(Samach et al., 2000; Jaudal et al., 2018). 따라서 본 연구에서 발견한 소나무의 *SOC1* 후보 유전자와 속씨식물의 *SOC1*은 유사한 발현 패턴을 가지고 있다고 할 수 있다.

침엽수에서 발견된 *MADS-box* 유전자인 *DAL1*은 속씨식물의 E-class 유전자인 *AGL6*의 homolog로서, 생식생장으로의 전환을 유도하며 임목이 성장하면서 그 발현량이 증가하는 경향을 보인다(Tandre et al., 1995). 또한 *DAL10*은 독일가문비(*Picea abies*)의 정단 분열조직에서 생식생장을 조절하며 성장중인 구과에서 높게 발현하는 것으로 확인되었다. *DAL19* 또한 독일가문비의 구과 형성 시작 단계에서 높게 발현한다고 한다(Carlsbecker et al., 2003; Carlsbecker et al., 2004; Uddenberg et al., 2013). 게다가 만주곰솔(*P. tabuliformis*)에서 *PtDAL1*과 *PtDAL10*을 포함한 다양한 *MADS-box* 유전자들이 발견되었다(Ma et al., 2021). 계통발생학적 연구를 통해 겉씨식물의 *DAL10*과 애기장대의 *SOC1*은 서로 ortholog 유전자라는 것이 밝혀졌으며, 따라서 *DAL10* 유전자는 겉씨식물에 특이적인 *MADS-box* 유전자라고 할 수 있다(Uddenberg et al., 2013). 본 연구에서는 소나무의 다양한 *DAL* 후보 유전자를 발견하였으며 이들은 정단, 암구과, 잎, 줄기, 뿌리

에서 높게 발현하였다. 이를 통해 소나무의 *DAL* 유전자는 생식기관 형성을 위해 정단과 구과에서 높게 발현하며 활발히 성장하는 잎, 줄기, 뿌리에서도 나타나는 것으로 확인되었다.

애기장대의 *MADS-box* 유전자 중 A-class 유전자인 *API*이 활성화되면 분열조직에서 생식생장을 촉진한다고 알려져 있다(Benlloch et al., 2007). 또한 *AP2* 전사인자는 개화 시기를 조절하는 데 많은 영향을 주며, 비활성화될 경우 조기 개화가 유도되고 지나치게 활성화될 경우 개화가 늦어진다는 연구결과가 존재한다(Yant et al., 2010; Aukerman and Sakai, 2003). 겉씨식물에서도 *AP2*의 homolog인 *PaAP2L1*, *PaAP2L2*, *PaAP2L3*가 독일가문비에서 확인되었으며, 해당 유전자들은 정단 분열조직과 암구과뿐만 아니라 잎과 뿌리 등 영양기관에서도 발현하는 것으로 나타났다(Vahala et al., 2001; Nilsson et al., 2006). 본 연구에서는 소나무의 *AP2L1*과 *AP2L3* 후보 유전자를 발견하였으며 정단과 암구과뿐만 아니라 잎, 줄기, 뿌리에서도 발현됨을 확인하였다.

한편 *SPL* 유전자는 식물 발달에 관여하는 다양한 유전자들의 발현을 조절한다. 생식생장을 시작하기 전에는 microRNA의 한 종류인 miR156가 *SPL*의 억제자로 작용하여 생식발달을 제한한다(Cardon et al., 1997; Poethig, 2013). 애기장대에서 *SPL*은 *CO*와 유사한 기능을 가져 *FT*의 발현을 증가시키며, 직접적으로 정단 분열조직에서 *LFY*와 *MADS-box* 유전자들을 활성화하여 개화를 유도한다고 알려져 있다(Wang et al., 2009; Yamaguchi et al., 2009; Yu et al., 2012). 또한 줄기 성장 과정에서 잎의 형태 변화에 관여하며 뿌리 성장에도 영향을 주는 것으로 보고된다(Yu et al., 2015; He et al., 2018). 본 연구에서 소나무의 *SPL* 후보 유전자를 발견하였으며 정단과 암구과뿐만 아니라 잎, 뿌리, 줄기 등 영양기관에서도 나타났다. 따라서 겉씨식물의 *SPL* 유전자는 구과 형성 초기에 생식 기관을 유도하며 영양기관의 형태적 변화에

도 관여한다고 추정할 수 있다.

애기장대의 B-class 유전자인 *AP3*의 활성을 위해 *LFY*는 *UFO*로 알려진 F-box 단백질에 부착해야 한다(Chae et al, 2008). *UFO*는 복합적인 꽃 발달 과정에 관여하는데, 특히 꽃분열조직의 형성을 유도하며 꽃 기관들 사이의 경계를 조절한다(Joshua et al., 1995; Levin et al., 1995; Ni et al., 2004). 완두콩의 *UFO* ortholog인 *Fim*은 정상적인 꽃과 잎의 발달에 필수적이라고 여겨지는데(Taylor et al., 2001), 이처럼 *UFO* 유전자 서열은 다양한 식물에 보존되어 있는 것으로 추정된다. 본 연구에서 소나무의 *UFO* 후보 유전자를 발견하였으며 정단과 암구과, 그리고 잎에서 발현하는 것으로 나타났다. 이에 따라 겉씨식물의 *UFO* ortholog는 정단 분열조직에서 초기 암구과 형성에 관여하며 잎의 발달에도 영향을 준다고 할 수 있다.

겉씨식물은 뿌리 정단 분열조직과 줄기 정단 분열조직(SAM)인 서로 다른 두 종류의 정단 분열조직을 가진다. SAM은 잎, 줄기, 꽃을 생산하는데 영양생장 단계 동안 잎 시원체(primordia)를 형성하며 영양생장이 끝날 때 꽃을 형성하는 생식생장 단계로 전환된다. 줄기가 신장되는 동안 잎의 엽액에 2차 SAM이 생성되고 꽃분열조직은 1차와 2차 SAM의 측면에 생성된다. 1차 생식적 SAM은 꽃차례분열조직(inflorescence meristem)이라고도 하며, 네 종류의 꽃 기관을 형성한 뒤 꽃분열조직의 활성은 끝나지만 SAM은 무한정으로 성장한다(Fletcher, 2002). 이렇듯 SAM과 비교했을 때 꽃분열조직의 생장은 유한적이며, 이때 *WUS* 유전자가 애기장대의 꽃분열조직 발달에 필수적인 역할을 한다고 알려져 있다(Laux et al., 1996). 꽃 발달 과정에서 *WUS*는 C-class 유전자인 *AG*와 상호작용을 하는데, *WUS*가 *AG*의 발현을 유도하고 반대로 *AG*에 의해 *WUS*의 발현이 억제됨으로써 꽃분열조직의 생장이 멈추게 된다(Lenhard et al., 2001). 본 연구에서 발견한 소나무의 *WUS* 후보 유전

자는 암구과에서 특히 많이 발현하였다. 이는 소나무의 *WUS*가 꽃분열 조직에서 암구과의 생식생장에 필수적인 유전자들의 발현을 유도하기 때문이라고 추정된다.

제6장 결 론

본 연구에서는 소나무의 구과 발달 과정에 관여하는 7종류(*PdAP2L*, *PdDAL*, *PdMADS*, *PdSOC1*, *PdSPL3*, *PdUFO*, *PdWUS*)의 후보 개화 조절 유전자들을 발견하였다. 이들은 정단과 암구과뿐만 아니라 잎, 줄기, 뿌리 등 다양한 영양기관에서도 발현하였으며, 이미 밝혀진 다른 수종들의 개화 조절 유전자들과 그 기능이나 발현 양상이 비슷한 특성을 보였다. 해당 후보 유전자들의 기관별 발현값을 이용하여 소나무의 기관 구분 가능성이 가능하며, 이를 통해 다양한 MADS-box 전사인자들의 서로 다른 조합으로 인해 소나무의 각 기관이 형성된다고 판단할 수 있었다. 게다가 임목은 유전적인 변화뿐만 아니라, 외부 환경과 생물에 반응하면서 개화 과정이 진행된다는 것을 확인할 수 있었다.

하지만 본 연구에서 발견한 각 유전자의 정확한 기능과 작용 과정을 이해하기 위해서는 추가적인 연구가 필요할 것이다. 소나무의 후보 개화 조절 유전자들을 Rapid Amplification of cDNA Ends(RACE)-PCR과 클로닝 등을 통해 분리한 후, 유전자 과발현(overexpression)이나 기능 상실(loss-of-function)을 통해 애기장대와 같은 모델식물에 형질전환함으로써 각 유전자의 기능을 검증할 수 있을 것이다. 이러한 과정을 통하여 소나무의 개화 조절 유전자들을 NCBI와 같은 공개 데이터베이스에 최종적으로 등록 가능할 것이라고 기대한다. 소나무의 개화 조절 유전자에 대한 연구가 활발히 이루어진다면 임목 개량 과정에서 세대를 단축시키거나 구과 생산량이 우수한 품종을 개발하여 사회적인 요구에 대응할 수 있을 것이다. 또한 소나무 개화 과정에 대한 전체적인 유전생리학적 정보를 통해 소나무의 기초적인 개화 메커니즘을 이해함으로써 임목 개량이나 유전적 마커 개발 등에 필요한 지식을 얻을 수 있을 것이다.

참고 문헌

- Abel, S., Ticconi, C.A. and Delatorre, C.A. 2002. Phosphate sensing in higher plants. *Physiologia Plantarum* 115(1): 1-8.
- Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J. and Yanofsky, M.F. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* 97(10): 5328-5333.
- Amasino, R.M. and Michaels, S.D. 2010. The timing of flowering. *Plant Physiology* 154(2): 516-520.
- Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. (2019. 01. 08).
- Aukerman, M.J. and Sakai, H. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-Like target genes. *Plant Cell* 15(11): 2730-2741.
- Avia, K., Kärkkäinen, K., Lagercrantz, U. Savolaninen, O. 2014. Association of FLOWERING LOCUS T/TERMINAL FLOWER 1-like gene FTL2 expression with growth rhythm in Scots pine (*Pinus sylvestris*). *New Phytologist* 204(1): 159-170.
- Aziz, S. and Mitu, T.K. 2019. Analysis of fatty acid and determination of total protein and phytochemical content of *Cassia sophera* Linn leaf, stem, flower, and seed. *Beni-Suef University Journal of Basic and Applied Sciences* 8(3).
- Baker, B., Zambryski, P., Staskawicz, B. and Dinesh-Kumar, S.P. 1997. Signaling in plant-microbe interactions. *Science* 276(5313): 726-733.
- Begley, T.P and Young, H. 1989. Protochlorophyllide reductase. 1. Determination of the regiochemistry and the stereochemistry of the reduction of protochlorophyllide to chlorophyllide. *Journal of the American Chemical Society* 111(8): 3095–3096.
- Benlloch, R., Berbel, A., Serrano-Mislata, A. and Madueño, F. 2007. Floral initiation and inflorescence architecture: A comparative view. *Annals of Botany* 100(3): 659-676.
- Blázquez, M.A., Ferrándiz, C., Madueño, F. and Parcy, F. 2006. How floral

- meristems are built. *Plant Molecular Biology* 60(6): 855-870.
- Bleecker, A.B. and Kende, H. 2000. Ethylene: a gaseous signal molecule in plants. *Annual Review of Cell and Development Biology* 16: 1-18.
- Böhlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A.M., Jansson, S., Strauss, S.H. and Nilsson, O. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* 312(5776): 1040-1043.
- Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–2120.
- Bonas, U. and Anckerveken, G.V. 1999. Gene-for-gene interactions: Bacterial avirulence proteins specify plant disease resistance. *Current opinion in microbiology* 2(1): 94-98.
- Bond, B.J. 2000. Age-related changes in photosynthesis of woody plants. *Trends in Plant Science* 5(8): 349–353.
- Bonnet-Masimbert, M. 1987. Floral induction in conifers: a review of available techniques. *Forest Ecology and Management* 19(1-4): 135-146.
- Borghini, M. et al. 2022. High-energy-level metabolism and transport occur at the transition from closed to open flowers. *Plant Physiology* 190(1): 319-339.
- Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. 1991. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112: 1-20.
- Browse, J. 2009. Jasmonate passes muster: a receptor and targets for the defense hormone. *Annual Review of Plant Biology* 60: 183-205.
- Bryant, D.M. et al. 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Reports* 18(3): 762–776.
- Cardarelli, M. and Costantino, P. 2018. An auxin switch for male fertility. *Nature plants* 4(7): 408-409.
- Cardarelli, M. and Costantino, P. 2018. An auxin switch for male fertility. *Nature plants* 4(7): 408-409.
- Cardon, G.H., Höhmann, S., Nettlesheim, K., Saedler, H. and Huijser, P. 1997. Functional analysis of the *Arabidopsis thaliana* SBP-box gene SPL3: A novel gene involved in the floral transition. *Plant Journal* 12(2): 367-377.
- Carlsbecker, A., Sundström, J., Tandré, K., Englund, M., Kvarnheden, A., Johanson, U. and Engström, P. 2003. The DAL10 gene from Norway spruce (*Picea abies*)

- belongs to a potentially gymnosperm specific subclass of MADS-box genes and is specifically active in seed cones and pollen cones. *Evolution and Development* 5(6): 551-561.
- Carlsbecker, A., Tandre, K., Johanson, U., Englund, M. Engström, P. 2004. The MADS-box gene *DAL1* is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *The Plant Journal* 40(4): 546-557.
- Cecchetti, V., Altamura, M.M., Brunetti, P., Petrocelli, V., Falasca, G., Ljung, K., Costantino, P. and Cardarelli, M. 2013. Auxin controls *Arabidopsis* anther dehiscence by regulating endothecium lignification and jasmonic acid biosynthesis. *The Plant Journal* 74(3): 411-422.
- Cecchetti, V., Altamura, M.M., Falasca, G., Costantino, P. and Cardarelli, M. 2008. Auxin regulates *Arabidopsis* anther dehiscence, pollen maturation, and filament elongation. *The Plant Cell* 20(7): 1760-1774.
- Chae, E., Tan, Q.K.G., Hill, T.A. Irish, V. 2008. An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development* 135(7): 1235-1245.
- Chen, Y., Shen, Q., Lin, R., Zhao, Z., Shen, C. Sun, C. 2017. De novo transcriptome analysis in *Dendrobium* and identification of critical genes associated with flowering. *Plant Physiology and Biochemistry* 119: 319-327.
- Chen, Y.C. and McCormick, S. 1996. Sidecar pollen, an *Arabidopsis thaliana* male gametophytic mutant with aberrant cell divisions during pollen development. *Development* 122(10): 3243-3253.
- Chung, H. et al. 2019. Development of polymorphic SSR markers from *Pinus densiflora* (Pinaceae) natural population in Korea. *Plant Breeding and Biotechnology* 7(1): 67-71.
- Coughlan, M.P. and Hazlewood, G.P. 1993. beta-1,4-D-Xylan-degrading enzyme systems: biochemistry, molecular biology and applications. *Biotechnology and Applied Biochemistry* 17(3): 259-289.
- Creelman, R.A. and Mulpuri, R. 2002. The oxylipin pathway in *Arabidopsis*. *Arabidopsis Book* 1: e0012.
- Cremer, K.W. 1992. Relations between reproductive growth and vegetative growth of *Pinus radiata*. *Forest Ecology and Management* 52(1-4): 179-199.
- Damásio, A.R. et al. 2011. Heterologous expression of an *Aspergillus niveus*

- xylanase GH11 in *Aspergillus nidulans* and its characterization and application. *Process Biochemistry* 46(6): 1236-1242.
- Day, M.E. and Greenwood, M.S. 2011. Regulation of ontogeny in temperate conifers. pp. 91–119. In : Meinzer, F.C., Lachenbruch, B., and Dawson, T.E. (Ed.). Size- and age-related changes in tree structure and function. Springer. The Netherlands.
- De Moraes, C.M., Mescheer, M.C., Tumlinson, J.H. 2001. Caterpillar-induced nocturnal plant volatiles repel nonspecific females. *Nature* 410: 577–580.
- Denoeud F. et al. 2008. Annotating genomes with massive-scale RNA sequencing. *Genome Biology* 9(R175).
- Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorenstein, N., Faldt, J., Miller, B. and Bohlmann, J. 2003. (E)-b-Ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new TPS-subfamily. *Plant Cell* 15(5): 1227–1241.
- Durand, R. and Durand, B. 1984. Sexual differentiation in higher plants. *Physiologia Plantarum* 60(3): 267-274.
- Eddy, S.R. 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7(10): e1002195.
- El-Kassaby, Y.A. and Barclay, H.J. 1991. Cost of reproduction in Douglas-fir. *Canadian Journal of Botany* 70(7): 1429–1432.
- Farvardin, A., González-Hernández, A.I., Llorens, E., García-Agustín, P., Scalschi, L. and Vicedo, B. 2020. The apoplast: a key player in plant survival. *Antioxidants* 9(7): 604.
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M. and Colombo, L. 2003. MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant Cell* 15(11): 2603-2611.
- Feng, X., Xue-mei, T., Yang, Z. and Fu-hua, F. 2008. Transcriptome analysis of *Pinus massoniana* Lamb. microstrobili during sexual reversal. *Open Life Sciences* 13(1): 97–106.
- Fenning, T.M. and Gershenzon, J. 2002. Where will the wood come from? Plantation forests and the role of biotechnology. *Trends in Biotechnology* 20(7): 291-296.

- Finn, R.D. et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44(D1): D279–285.
- Fletcher, J.C. 2002. Shoot and floral meristem maintenance in *Arabidopsis*. *Annual Review of Plant Biology* 53: 45-66.
- Friedman, M., Henika, P.R. and Mandrell, R.E. 2002. Bactericidal activities of plant essential oils and some of their isolated constituents against *Campylobacter jejuni*, *Escherichia coli*, *Listeria monocytogenes*, and *Salmonella enterica*. *Journal of Food Protection* 65(10): 1545–1560.
- Friedrich, H. 1976. Phenylpropanoid constituents of essential oils. *Lloydia* 39(1): 1-7.
- Fritsche, S., Salgado, L.R., Boron, A.K., Hanning, K.R., Donaldson, L.A. and Thorlby, G. 2022. Transcriptional regulation of pine male and female cone initiation and development: key players identified through comparative transcriptomics. *Frontiers in Genetics* 13: 815093.
- Gene Ontology Consortium. 2015. Gene ontology consortium: going forward. *Nucleic Acids Research* 43: D1049–1056.
- Graherr, M.G. et al. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29(7): 644–652.
- Gramzow, L and Theissen, G. 2010. A hitchhiker's guide to the MADS world of plants. *Genome Biology* 11(214).
- Greenwood, M.S. 1995. Juvenility and maturation in conifers: Current concepts. *Tree Physiology* 15(7-8): 433–438.
- Grotewold, E. 2006. The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology* 57: 761–780.
- Gyllenstrand, N., Clapham, D., Källman, T., Lagercrantz, U. 2007. A Norway spruce FLOWERING LOCUS T homolog is implicated in control of growth rhythm in conifers. *Plant Physiology* 144(1): 248-257.
- Haas, B.J. et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.
- Hackett, W. and Murray, J. 1993. Maturation and rejuvenation in woody species. In : Ahuja, M.R. (Ed.). *Micropropagation of Woody Plants*. *Forestry Sciences* 41: 93-105.

- Hackett, W.P. 1987. Juvenility and maturity. pp. 216–231. In : Cell and tissue culture in forestry. Springer Netherlands. Dordrecht, The Netherlands.
- Hause, B., Stenzel, I., Miersch, O., Maucher, H., Kramell, R., Ziegler, J. and Wasternack, C. 2000. Tissue-specific oxylipin signature of tomato flowers: allene oxide cyclase is highly expressed in distinct flower organs and vascular bundles. *The Plant Journal* 24(1): 113-26.
- He, J. et al. 2018. Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in *Arabidopsis thaliana*. *PLoS Genetics* 14(4): e1007337.
- Houle, G and Filion, L. 1993. Interannual variations in the seed production of *Pinus banksiana* at the limit of the species distribution in northern Quebec, Canada. *American Journal of Botany* 80(11): 1242-1250.
- Hrubá, P., Honys, D., Twell, D., Čapková, V. and Tupý, J. 2004. Expression of β -galactosidase and β -xylosidase genes during microspore and pollen development. *Planta* 220: 931-940.
- Hrubá, P., Honys, D., Twell, D., Čapková, V. and Tupý, J. 2004. Expression of β -galactosidase and β -xylosidase genes during microspore and pollen development. *Planta* 220(6): 931-940.
- Huerta-Cepas, J. et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* 44(D1): D286–293.
- Irish, V.F. 2010. The flowering of *Arabidopsis* flower development. *The Plant Journal* 61(6): 1014-1028.
- Jackson, D.I. and Sweet, G.B. 1972. Flower initiation in temperate woody plants. *Horticultural Science Abstracts* 42: 9-24.
- Jaiswal, V.S., Narayan, P. and Lal, M. 1984. Activities of acid and alkaline phosphatases in relation to sex differentiation in *Carica papaya* L.. *Biochemie und Physiologie der Pflanzen* 179(9): 799-801.
- Jaudal, M., Zhang, L., Che, C., Li, G., Tang, Y., Wen, J., Mysore, K.S. and Putterill, J. 2018. A SOC1-like gene MtSOC1a promotes flowering and primary stem elongation in *Medicago*. *Journal of Experimental Botany* 69(20): 4867-4880.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids*

- Research 44(D1): D457–462.
- Kang, K.S. 2000. Clonal and annual variation of flower production and composition of gamete gene pool in a clonal seed orchard of *Pinus densiflora*. Canadian Journal of Forrest Research 30(8): 1275–1280.
- Kang, K.S. and Lindgren, D. 1998. Fertility variation and its effect on the relatedness of seeds in *Pinus densiflora*, *Pinus thunbergii* and *Pinus koraiensis* clonal seed orchards. Silvae Genetica 47(4): 196-201.
- Kato, Y., Miyake, I. and Ishikawa, H. 1958. Initiation of flower bud by gibberellin in *Cryptomeria japonica*. Journal of Japanese Forrest Society 40: 35-36.
- Korpelainen, E., Tuimala, J., Somervuo, P., Huss, M. and Wong, G. 2014. RNA-seq Data Analysis: A Practical Approach. Chapman and Hall. New York.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305(3): 567–580.
- Kumar, A. 1983. Some aspects of physiology and biochemistry of sex expression in flowering plants. (Dissertation). Varanasi, India. Banaras Hindu University.
- Lal, M. and Jaiswal, V. 1988. Modification of flower sex and acid phosphatase activity by phthalimides in female plants of *Morus nigra* L.. Plant Growth Regulation 7: 29-37.
- Laux, T., Mayer, K.F., Berger, J. Jürgens, G. 1996. The WUSCHEL gene is required for shoot and floral meristem integrity in *Arabidopsis*. Development 122(1): 87-96.
- Lee, S.W., Jang, S.S., Chang, K.H., Kim, C.S. 2003. Estimation of mating system parameters in the natural population of *Pinus densiflora* of Anmyun Island, Korea using allozyme markers. Journal of Korean Forestry Society 92(2): 121-128.
- Lenhard, M., Bohnert, A., Jürgens, G. and Laux, T. 2001. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between WUSCHEL and AGAMOUS. Cell 105(6): 805-814.
- Levin, J.Z. and Meyerowitz, E.M. 1995. UFO: an *Arabidopsis* gene involved in both floral meristem and floral organ development. The Plant Cell 7(5): 529-548.
- Li, B., Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq

- data with or without a reference genome. *BMC Bioinformatics* 12(323).
- Li, W. and Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13): 1658-1659.
- Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S. and Yanofsky, M.F. 1999. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* 11(6): 1007-1018.
- Liu, C., Thong, Z. Yu. H. 2009b. Coming into bloom: the specification of floral meristems. *Development* 136(20): 3379-3391.
- Liu, C., Xi, W., Shen, L., Tan, C. and Yu, H. 2009a. Regulation of floral patterning by flowering time genes. *Developmental Cell* 16(5): 711-722.
- Ma, J.J. et al. 2021. MADS-box transcription factors MADS11 and DAL1 interact to mediate the vegetative-to-reproductive transition in pine. *Plant Physiology*. 187(1): 247-262.
- Malik, C.P. and Bhattacharya, S. 1979. Sex expression and sex differentiation in flowering plants. pp 1-42. In : Malik, C.P. (Ed.). *Current Advances in Plant Reproduction Biology I*. Kalyani Publishers. New Delhi.
- Mandaokar, A. et al. 2006. Transcriptional regulators of stamen development in *Arabidopsis* identified by transcriptional profiling. *The Plant Journal* 46(6): 984-1008.
- Mardis, E.R. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9:387-402.
- Markham, K.R., Gould, K.S., Winefield, C.S., Mitchell, K.A., Bloor, S.J. and Boase, M.R. 2000. Anthocyanic vacuolar inclusions—their nature and significance in flower colouration. *Phytochemistry* 55(4): 327–36.
- Marty, F. 1999. Plant vacuoles. *The Plant Cell* 11(4): 587–599.
- Matsuzaki, T., Koiwai, A. and Kawashima, N. 1983. Total fatty acid and polar lipid content in developing flower of *Nicotiana tabacum*. *Plant and Cell Physiology* 24(2): 199-206.
- Matthews, R.E.F. 1991. *Plant Virology*. 3d Ed. Academic Press. SanDiego, CA.
- McCormick, S. 1993. Male gametophyte development. *The Plant Cell* 5(10): 1265-1275.
- McIntosh, J.R., Grishchuk, E.L. and West, R.R. 2002. Chromosome-microtubule interactions during mitosis. *Annual Review of Cell and Developmental Biology*

18: 193-219.

- Mouradov, A., Glassick, T., Hamdorf, B. and Teasdale, R.D. 1998. NEEDLY, a *Pinus radiata* ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. Proceedings of the National Academy of Sciences of the United States of America 95(11): 6537-6542.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R. and Parcy, F. 2010. LEAFY blossoms. Trends in Plant Science 15(6): 346-352.
- Moyroud, E., Tichtinsky, G. and Parcy, F. 2009. The LEAFY floral regulators in angiosperms: conserved proteins with diverse roles. Journal of Plant Biology 52: 177-185.
- Mu, J. et al. 2008. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiology 148(2): 1042-1054.
- Mutz, K.O., Heilkenbrinker, A., Lönne, M., Walter, J.G. and Stahl, R. 2013. Transcriptome analysis using next-generation sequencing. Current Opinion in Biotechnology 24(1): 22-30.
- Ng, M. and Yanofsky, M. Function and evolution of the plant MADS-box gene family. 2001. Nature Reviews Genetics 2: 186-195.
- Ni, W., Xie, D., Hobbie, L., Feng, B., Zhao, D., Akkara, J. Ma, H. 2004. Regulation of flower development in *Arabidopsis* by SCF complexes. Plant Physiology 134(4): 1574-1585.
- Nilsson, L., Carlsbecker, A., Sundås-Larsson, A. and Vahala, T. 2007. APETALA2 like genes from *Picea abies* show functional similarities to their *Arabidopsis* homologues. Planta volume 225: 589-602.
- Niu, S. et al. 2022. The Chinese pine genome and methylome unveil key features of conifer evolution. Cell 185(1): 204-217.e14.
- Niu, S., Yuan, H., Sun, X., Porth, I., Li, Y., El-Kassaby, Y.A. and Li, W. 2015. A transcriptomics investigation into pine reproductive organ development. New Phytologist 209(3): 1278-1289.
- O'Leary, N.A. et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44(D1): D733-745.
- Ohmiya, A., Hirashima, M., Yagi, M., Tanase, K. and Yamamizo, C. 2014. Identification of genes associated with chlorophyll accumulation in flower

- petals. PLoS One 9(12): e113738.
- Oussalah, M., Caillet, S., Saucier, L., Lacroix, M. 2006. Inhibitory effects of selected plant essential oils on the growth of four pathogenic bacteria: *E. coli* O157:H7, *Salmonella Typhimurium*, *Staphylococcus aureus* and *Listeria monocytogenes*. Food Control 18(5): 414-420.
- Pařenicová, L. et al. 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis* new openings to the MADS world. Plant Cell 15(7): 1538-1551.
- Passeri, V., Koes, R. and Quattrocchio, F.M. 2016. New challenges for the design of high value plant products: stabilization of anthocyanins in plant vacuoles. Frontiers in Plant Science 7.
- Petersen, L.M., Martin, I.W., Moschetti, W.E., Kershaw, C.M and Tsongalis, G.J. 2019. Third-generation sequencing in the clinical laboratory: Exploring the advantages and challenges of nanopore sequencing. Journal of Clinical Microbiology 58(1): e01315-19.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8(10): 785–786.
- Pharis, R.P. 1975a. Promotion of flowering in conifers by gibberellins. The Forest Chronicle 51: 244-248.
- Pharis, R.P. and Morf, W. 1967. Experiments on the precocious flowering of western red cedar and four species of *Cupressus* with gibberellin A3 and A4/7 mixture. Canadian Journal of Botany 65: 1519-1524.
- Pharis, R.P., Wample, R.L. and Kamienska, A. 1975b. Growth, development and sexual differentiation in *Pinus* with emphasis on the role of the plant hormone, gibberellin. pp. 106-134. In: Baum-gartner, D.M. (Ed.). Management of Lodgepole Pine Ecosystems. Washington State University Press. Washington, UAS.
- Poethig, R. 1990. Phase change and the regulation of shoot morphogenesis in plants. Science 250(4983): 923–930.
- Poethig, R.S. 2013. Vegetative phase change and shoot maturation in plants. Current Topics in Developmental Biology 105: 125-152.
- Purugganan, M.D. 1997. The MADS-box floral homeotic gene lineages predate the

- origin of seed plants: Phylogenetic and molecular clock estimates. *Journal of Molecular Evolution* 45(4) :392-396.
- R Core Team. 2018. R: A language and environment for statistical computing. <https://www.R-project.org/>. (2022. 10. 31).
- Raguso, R.A. and Pichersky, E. 1995. Floral volatiles from *Clarkia breweri* and *C. concinna* (Onagraceae): recent evolution of floral scent and moth pollination. *Plant Systematics Evolution* 194(1/2): 55-67.
- Reid, M.S. and Wu, M.J. 1992. Ethylene and flower senescence. *Plant Growth Regulation* 11: 37-43.
- Reinhard, J., Srinivasan, M.V. and Zhang, S. 2004. Scent-triggered navigation in Honeybees. *Nature* 427: 411.
- Ren, X., Hu, Q., Liu, S., Wang., J and Miecznikowski, J.C. 2017. Gene set analysis controlling for length bias in RNA-seq experiments. *BioData Mining* 10(5).
- Ren, Z. et al. 2019. QTL analysis of delayed maize flowering in response to low phosphate across multi-environments. *Euphytica* 215(128).
- Riechmann, J.L. and Meyerowitz, E.M. 1997. MADS domain proteins in plant development. *Biological Chemistry* 378(10): 1079-1101.
- Ross, A.F. 1961. Systemic acquired resistance induced by localized virus infections in plants. *Virology* 14(3): 340-358.
- Rouached, H., Arpat, A.B. and Poirier, Y. 2010. Regulation of phosphate starvation responses in plants: signaling players and cross-talks. *Molecular Plant* 3(2): 288-299.
- Ruonala, R., Rinne, P.L.H., Baghour, M., Moritz, T., Tuominen, H and Kangasjärvi, J. 2006. Transitions in the functioning of the shoot apical meristem in birch (*Betula pendula*) involve ethylene. *The Plant Journal* 46(4): 628-640.
- Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. and Coupland, G. 2000. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. *Science* 288(5471): 1613-1616.
- Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74(12): 5463-5467.
- Shalit, M. et al. 2004. Volatile compounds emitted by rose cultivars: fragrance perception by man and honeybees. *Israel Journal of Plant Sciences* 52: 245–255.

- Singh, H. 1978. Embryology of Gymnosperms. Gebrüder Borntraeger. Berlin, Germany.
- Srikanth, A. and Schmid, M. 2011. Regulation of flowering time: All roads lead to Rome. *Cellular and Molecular Life Science* 68(12): 2013–2037.
- Strauss, S.H., Rottmann, W.H., Brunner, A.M. and Sheppard, L.A. 1995. Genetic engineering of reproductive sterility in forest trees. *Molecular Breeding* 1: 5–26.
- Sundström, J., Carlsbecker, A., Svensson, M.E., Svenson, M., Johanson, U., Theissen, G. and Engström, P. 1999. MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Developmental Genesis* 25(3): 253-266.
- Tanaka, Y., Sasaki, N. and Ohmiya, A. 2008. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal* 54(4): 733-749.
- Tandre, K., Albert, V.A., Sundas, A. and Engstrom, P. 1995. Conifer homologues to genes that control floral development in angiosperms. *Plant Molecular Biology* 27(1): 69-78.
- Taylor, S., Hofer, J. and Murfet, I. 2001. *Stamina pistilloida*, the pea ortholog of *Fim* and *UFO*, is required for normal development of flowers, inflorescences, and leaves. *The Plant Cell* 13(1): 31-46.
- Theissen, G. and Becker, A. 2004. Gymnosperms orthologs of class B floral homeotic genes and their impact on understanding flower origin. *Critical Review in Plant Sciences* 23(2): 129-148.
- Theissen, G., Kim, J.T. and Saedler, H. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution* 43(5): 484-516.
- Theißen, G., Melzer, R. and Rümpler, F. 2016. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* 143(8): 3259–3271.
- Thomson, J.A. 1993. Molecular biology of xylan degradation. *FEMS Microbiology Reviews* 10(1-2): 65-82.
- Tornero, P., Gadea, J., Conejero, V and Vera, P. 1997. Two PR-1 genes from tomato are differentially regulated and reveal a novel mode of expression for a pathogenesis-related gene during the hypersensitive response and development.

- Molecular Plant-Microbe Interactions 10(5): 624-634.
- Uddenberg, D., Reimegård, J., Clapham, D., Almqvist, C., von Arnold, S., Emanuelsson, O. Sundström, J.F. 2013. Early cone setting in *Picea abies* acrocona is associated with increased transcriptional activity of a MADS box transcription factor. *Plant Physiology* 161(2): 813-823.
- Vahala, T., Oxelman, B. and von Arnold, S. 2001. Two APETALA2-like genes of *Picea abies* are differentially expressed during development. *Journal of Experimental Botany* 52(358): 1111-1115.
- van Doorn, W.G. and van Meeteren, U. 2003. Flower opening and closure: a review. *Journal of Experimental Botany* 54(389): 1801-1812.
- Vance, C.P., Uhde-Stone, C. and Allan, D.L. 2003. Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytologist* 157(3): 423-447.
- Walker, J.M and Marsh, S. (Ed.). 2007. Pyrosequencing Protocols. Humana Press. New York.
- Wan, L., Xia, Q., Qiu, X. and Selvaraj, G. 2002. Early stages of seed development in *Brassica napus*: a seed coat-specific cysteine proteinase associated with programmed cell death of the inner integument. *The Plant Journal* 30(1): 1-10.
- Wang, J.W., Czech, B. Weigel, D. 2009. miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138(4): 738-749.
- Ward, E.R., Uknes, S.J., Williams, S.C., Dincher, S.S., Wiederhold, D.L., Alexander, D.C., Ahl-Goy, P., Métraux, J.P. and Ryals, J.A. 1991. Coordinate gene activation in response to agents that induce systemic acquired resistance. *Plant Cell* 3(10): 1085-1094.
- Wasternack, C., Forner, S., Strnad, M. and Hause, B. 2013. Jasmonates in flower and seed development. *Biochimie* 95(1): 79-85.
- Weigel, D. and Meyerowitz, E.M. 1993. Activation of floral homeotic genes in *Arabidopsis*. *Science* 261(5129): 1723-1726.
- Wendling, I., Trueman, S.J. and Xavier, A. 2014. Maturation and related aspects in clonal forestry-Part I: Concepts, regulation and consequences of phase change. *New Forest* 45(4): 449-471.
- Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U. and

- Weigel, D. 2005. Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* 309(5737): 1056-1059.
- Wilks, H.W. and Timko, M.P. 1995. A light-dependent complementation system for analysis of NADPH:protochlorophyllide oxidoreductase: identification and mutagenesis of two conserved residues that are essential for enzyme activity. *Proceedings of the National Academy of Sciences of the United States of America* 92(3): 724–728.
- Williams, C.G. 2009. *Conifer reproductive biology*. Springer. Dordrecht, The Netherlands. P. 169.
- Xu, Y. and Hanson, M.R. 2000. Programmed cell death during pollination-induced petal senescence in *Petunia*. *Plant Physiology* 122(4): 1323-1333.
- Yamaguchi, A., Wu, M.F., Yang, L., Wu, G., Poethig, R.S. and Wagner, D. 2009. The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. *Development Cell* 17(2): 268-278.
- Yang, Y., Klejnot, J., Yu, Z., Liu, X. and Lin, C. 2007. Florigen (II): It is a mobile protein. *Journal of Integrative Plant Biology* 49(12): 1665-1669.
- Yant, L., Mathieu, J., Dinh, T.T., Ott, F., Lanz, C., Wollmann, H., Chen, X. and Schmid, M. 2010. Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor APETALA2. *Plant Cell* 22(7): 2156-2170.
- Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S. Ahn, J.H. 2005. CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in *Arabidopsis*. *Plant Physiology* 139(2): 770-778.
- Yu, N., Niu, Q.W., Ng, K.H. and Chua, N.H. 2015. The role of miR156/SPLs modules in *Arabidopsis* lateral root development. *The Plant Journal* 83(4): 673-685.
- Yu, S. et al. 2012. Gibberellin regulates the *Arabidopsis* floral transition through miR156-targeted SQUAMOSA promoter binding-like transcription factors. *Plant Cell* 24(8): 3320-3332.
- Zhang, H., Wang, L., Deroles, S., Bennett, R. and Davies, K. 2006. New insight into the structures and formation of anthocyanic vacuolar inclusions in flower

petals. BMC Plant Biology 6(29).

Zhu, H., Qian, W., Lu, X., Li, D., Liu, X., Liu, K. and Wang, D. 2005. Expression patterns of purple acid phosphatase genes in *Arabidopsis* organs and functional analysis of AtPAP23 predominantly transcribed in flower. Plant Molecular Biology 59(4): 581–594.

Appendix

Table S1. BLASTX best-hit results of transcripts in ‘petal morphogenesis (GO:0048446)’ and ‘petal development (GO:0048441)’

| Species | Protein | Accession | Transcript ID | Identity (%) | E value | Score (bits) |
|---------------------------|----------------------|----------------|--------------------------|--------------|---------|--------------|
| | transcription factor | | | | | |
| <i>Camellia sinensis</i> | bHLH79 isoform X4 | XP_028084308.1 | TRINITY_DN8967_c0_g1_i5 | 56 | 4E-41 | 155 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN326_c0_g1_i1 | 58 | 0 | 610 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN326_c0_g1_i24 | 60 | 0 | 600 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN4142_c0_g1_i39 | 42 | 4E-135 | 421 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN1911_c0_g1_i9 | 59 | 2E-128 | 390 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN4142_c0_g1_i56 | 43 | 3E-103 | 330 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN4968_c0_g1_i1 | 56 | 2E-61 | 213 |
| <i>Pinus tabuliformis</i> | WSD1 | AJP06367.1 | TRINITY_DN4968_c0_g1_i3 | 57 | 3E-60 | 199 |

Table S2. BLASTX best-hit results of transcripts in ‘flower development (GO:0009908)’

| Species | Protein | Accession | Transcript ID | Identity (%) | E value | Score (bits) |
|---------------------------|---------------------------------|----------------|---------------------------|--------------|---------|--------------|
| <i>Ginkgo bilob</i> | AINTEGUMENTA-like protein | BAE48515.1 | TRINITY_DN631_c0_g1_i1 | 63 | 0 | 557 |
| <i>Pinus tabuliformis</i> | AP2L1 | AJP06234.1 | TRINITY_DN2887_c0_g3_i1 | 99 | 0 | 753 |
| <i>Pinus tabuliformis</i> | AP2L3 | AJP06236.1 | TRINITY_DN2887_c0_g2_i2 | 99 | 0 | 1170 |
| | AT-hook motif nuclear-localized | | | | | |
| <i>Coffea arabica</i> | protein 24 | XP_027080473.1 | TRINITY_DN783_c4_g1_i1 | 65 | 5E-59 | 210 |
| <i>Pinus tabuliformis</i> | DAL10 | AJP06266.1 | TRINITY_DN5374_c0_g1_i5 | 100 | 0 | 530 |
| <i>Pinus tabuliformis</i> | DAL19 | AJP06271.1 | TRINITY_DN8993_c0_g1_i9 | 97 | 2E-29 | 124 |
| <i>Picea abies</i> | DAL19 protein | AGR53805.1 | TRINITY_DN7611_c0_g1_i11 | 71 | 1E-92 | 192 |
| <i>Picea abies</i> | DAL19 protein | AGR53805.1 | TRINITY_DN6053_c3_g1_i12 | 55 | 2E-51 | 177 |
| <i>Picea abies</i> | DAL19 protein | AGR53805.1 | TRINITY_DN6053_c3_g1_i11 | 55 | 2E-50 | 171 |
| <i>Pinus tabuliformis</i> | DAL3 | AJP06276.1 | TRINITY_DN1774_c1_g1_i23 | 76 | 4E-123 | 363 |
| <i>Pinus tabuliformis</i> | DAL4 | AJP06277.1 | TRINITY_DN7611_c0_g1_i7 | 98 | 7E-136 | 407 |
| <i>Pinus tabuliformis</i> | GA2ox1, partial | AHW42465.1 | TRINITY_DN3160_c0_g1_i1 | 99 | 0 | 731 |
| <i>Pinus tabuliformis</i> | GA2ox12 | AHW42462.1 | TRINITY_DN13996_c0_g1_i8 | 57 | 2E-116 | 354 |
| <i>Pinus tabuliformis</i> | GA2ox12 | AHW42462.1 | TRINITY_DN13996_c0_g1_i16 | 52 | 2E-99 | 310 |
| <i>Elaeis guineensis</i> | GATA transcription factor 19 | XP_010925915.1 | TRINITY_DN314629_c0_g1_i1 | 54 | 2E-18 | 98.2 |
| <i>Pinus tabuliformis</i> | MADS1 | AJP06319.1 | TRINITY_DN14061_c1_g2_i1 | 9 | 6E-166 | 481 |
| <i>Pinus tabuliformis</i> | MADS1 | AJP06319.1 | TRINITY_DN12240_c0_g1_i14 | 69 | 3E-51 | 187 |
| <i>Pinus tabuliformis</i> | MADS1 | AJP06319.1 | TRINITY_DN12240_c0_g1_i10 | 53 | 3E-56 | 183 |
| <i>Pinus tabuliformis</i> | MADS3, partial | AJP06321.1 | TRINITY_DN8544_c0_g2_i5 | 99 | 4E-171 | 498 |
| <i>Pinus tabuliformis</i> | MADS3, partial | AJP06321.1 | TRINITY_DN8544_c0_g2_i8 | 99 | 5E-131 | 383 |

| | | | | | | |
|------------------------------|---|----------------|---------------------------|----|--------|-----|
| <i>Pinus tabuliformis</i> | MADS3, partial | AJP06321.1 | TRINITY_DN8544_c0_g2_i2 | 99 | 2E-104 | 331 |
| <i>Pinus tabuliformis</i> | MADS3, partial | AJP06321.1 | TRINITY_DN8544_c0_g2_i4 | 96 | 8E-49 | 172 |
| <i>Pinus tabuliformis</i> | MADS4 | AJP06322.1 | TRINITY_DN10705_c0_g1_i19 | 61 | 7E-40 | 150 |
| <i>Pinus tabuliformis</i> | MADS6, partial | AJP06324.1 | TRINITY_DN3223_c1_g1_i9 | 96 | 4E-124 | 375 |
| <i>Pinus tabuliformis</i> | MADS7, partial | AJP06325.1 | TRINITY_DN3223_c1_g1_i15 | 97 | 1E-167 | 476 |
| <i>Pinus taeda</i> | MADS-box protein SOC1 | AZA14799.1 | TRINITY_DN1774_c1_g1_i25 | 92 | 5E-141 | 411 |
| <i>Pinus taeda</i> | MADS-box protein SOC1 | AZA14799.1 | TRINITY_DN1774_c1_g1_i1 | 90 | 2E-126 | 374 |
| <i>Pinus taeda</i> | MADS-box protein SOC1 | AZA14799.1 | TRINITY_DN1774_c1_g1_i7 | 85 | 9E-118 | 354 |
| <i>Mucuna pruriens</i> | MADS-box protein SOC1, partial MADS-box transcription factor | RDX68490.1 | TRINITY_DN3584_c3_g1_i3 | 69 | 9E-28 | 112 |
| <i>Pinus koraiensis</i> | GbMADS5 MADS-domain transcription | UXW62812.1 | TRINITY_DN4170_c1_g2_i2 | 91 | 7E-105 | 321 |
| <i>Picea abies</i> | factor dal3 variant 2 MADS-domain transcription | ADM73188.1 | TRINITY_DN1774_c0_g1_i6 | 70 | 1E-88 | 282 |
| <i>Picea abies</i> | factor dal3 variant 2 | ADM73188.1 | TRINITY_DN1774_c0_g1_i12 | 63 | 2E-83 | 269 |
| <i>Pinus koraiensis</i> | peroxidase | UXW62802.1 | TRINITY_DN4869_c0_g1_i10 | 81 | 0 | 539 |
| <i>Pinus koraiensis</i> | peroxidase | UXW62802.1 | TRINITY_DN4869_c0_g1_i16 | 80 | 0 | 535 |
| | protein HEADING DATE | | | | | |
| <i>Elaeis guineensis</i> | REPRESSOR 1 | XP_010929182.1 | TRINITY_DN7119_c0_g1_i9 | 49 | 1E-38 | 149 |
| <i>Cinnamomum micranthum</i> | protein SODIUM POTASSIUM f. ROOT DEFECTIVE 2-like | | | | | |
| <i>kanehirae</i> | protein | RWR93040.1 | TRINITY_DN5375_c0_g1_i2 | 50 | 9E-38 | 145 |
| <i>Pinus radiata</i> | putative MADS box transcription | AAB80808.1 | TRINITY_DN14326_c0_g1_i1 | 92 | 8E-134 | 388 |

| | | | | | | |
|------------------------------|-------------------------------------|------------|---------------------------|-----|--------|-----|
| | factor PrMADS5 | | | | | |
| | putative MADS box transcription | | | | | |
| <i>Pinus radiata</i> | factor PrMADS5 | AAB80808.1 | TRINITY_DN14326_c0_g1_i10 | 86 | 2E-119 | 355 |
| <i>Ginkgo biloba</i> | R2R3MYB1 | ASR18100.1 | TRINITY_DN921_c0_g1_i14 | 58 | 6E-61 | 214 |
| <i>Parasponia andersonii</i> | Snf1 kinase interactor-like protein | PON77759.1 | TRINITY_DN7119_c0_g1_i5 | 54 | 1E-46 | 168 |
| <i>Picea abies</i> | SOC1 | AJA34534.1 | TRINITY_DN16588_c0_g1_i11 | 65 | 7E-63 | 210 |
| <i>Picea abies</i> | SOC1 | AJA34534.1 | TRINITY_DN16588_c0_g1_i12 | 57 | 4E-61 | 206 |
| <i>Picea abies</i> | SOC1 | AJA34534.1 | TRINITY_DN3584_c3_g1_i2 | 60 | 1E-57 | 190 |
| <i>Larix kaempferi</i> | SPL3 | UXE52212.1 | TRINITY_DN43672_c0_g1_i2 | 88 | 2E-89 | 285 |
| <i>Pinus tabuliformis</i> | UFO1 | AJP06363.1 | TRINITY_DN15623_c0_g1_i1 | 99 | 0 | 791 |
| <i>Pinus tabuliformis</i> | UFO2 | AJP06362.1 | TRINITY_DN27618_c0_g1_i2 | 99 | 0 | 779 |
| <i>Pinus pinaster</i> | WUSCHEL homeobox protein | ANC94876.1 | TRINITY_DN26733_c0_g1_i1 | 100 | 2E-117 | 348 |
| | WUSCHEL homeobox protein | | | | | |
| <i>Pinus pinaster</i> | WUS | ALN42231.2 | TRINITY_DN12004_c0_g1_i2 | 98 | 6E-154 | 444 |

Table S3. BLASTX best-hit results of transcripts in ‘sporopollenin biosynthetic process (GO:0080110)’

| Species | Protein | Accession | Transcript ID | Identity (%) | E value | Score (bits) |
|---------------------------------|---|----------------|---------------------------|--------------|---------|--------------|
| <i>Citrus sinensis</i> | ABC transporter G family member 2 | XP_006474829.1 | TRINITY_DN5858_c0_g1_i13 | 56 | 0 | 746 |
| <i>Carica papaya</i> | ABC transporter G family member 26 | XP_021905991.1 | TRINITY_DN45193_c0_g1_i2 | 67 | 0 | 862 |
| <i>Phoenix dactylifera</i> | ABC transporter G family member 26 | XP_038976678.1 | TRINITY_DN5858_c0_g1_i11 | 56 | 0 | 729 |
| <i>Phoenix dactylifera</i> | ABC transporter G family member 26 | XP_038976678.1 | TRINITY_DN5858_c0_g1_i16 | 56 | 0 | 727 |
| <i>Amborella trichopoda</i> | ABC transporter G family member 26 | XP_006857737.1 | TRINITY_DN45193_c0_g1_i5 | 69 | 1E-134 | 419 |
| <i>Quercus lobata</i> | ABC transporter G family member 26 isoform X1 | XP_030928991.1 | TRINITY_DN5858_c0_g1_i8 | 55 | 0 | 729 |
| <i>Cryptomeria japonica</i> | acyl-CoA synthetase 5 | BDR24743.1 | TRINITY_DN11910_c0_g1_i1 | 74 | 0 | 746 |
| <i>Pinus radiata</i> | chalcone synthase homolog PrChS1 | AAB80804.1 | TRINITY_DN11961_c0_g1_i3 | 99 | 0 | 761 |
| <i>Pinus radiata</i> | chalcone synthase homolog PrChS1 | AAB80804.1 | TRINITY_DN11961_c0_g1_i1 | 98 | 3E-110 | 333 |
| <i>Cunninghamia lanceolata</i> | cinnamoyl-CoA reductase | AFX98067.1 | TRINITY_DN39924_c0_g1_i6 | 59 | 2E-125 | 378 |
| <i>Cephalotaxus hainanensis</i> | cinnamoyl-CoA reductase | UUJ74919.1 | TRINITY_DN39924_c0_g1_i21 | 80 | 0 | 551 |

| | | | | | | |
|-----------------------------|---------------------------------------|----------------|---------------------------|----|--------|-----|
| <i>Cephalotaxus</i> | | | | | | |
| <i>hainanensis</i> | cinnamoyl-CoA reductase | UUJ74919.1 | TRINITY_DN39924_c0_g1_i4 | 78 | 3E-147 | 395 |
| <i>Cephalotaxus</i> | | | | | | |
| <i>hainanensis</i> | cinnamoyl-CoA reductase | UUJ74919.1 | TRINITY_DN39924_c0_g1_i10 | 76 | 2E-82 | 283 |
| <i>Cephalotaxus</i> | | | | | | |
| <i>hainanensis</i> | cinnamoyl-CoA reductase | UUJ74919.1 | TRINITY_DN39924_c0_g1_i34 | 76 | 1E-85 | 283 |
| <i>Amborella trichopoda</i> | cytochrome P450 704B1 isoform X1 | XP_006827309.2 | TRINITY_DN13087_c2_g1_i1 | 73 | 0 | 812 |
| <i>Ginkgo biloba</i> | flavonoid 3' hydroxylase-like protein | AJO67233.1 | TRINITY_DN5588_c0_g1_i4 | 76 | 0 | 763 |
| <i>Ginkgo biloba</i> | flavonoid 3' hydroxylase-like protein | AJO67233.1 | TRINITY_DN5588_c0_g1_i3 | 76 | 0 | 763 |
| <i>Cannabis sativa</i> | tetraketide alpha-pyrone reductase 1 | XP_030485843.1 | TRINITY_DN8134_c0_g1_i3 | 71 | 1E-136 | 404 |
| <i>Citrus sinensis</i> | tetraketide alpha-pyrone reductase 1 | KAH9654674.1 | TRINITY_DN8134_c0_g1_i4 | 67 | 3E-123 | 369 |
| <i>Citrus sinensis</i> | tetraketide alpha-pyrone reductase 1 | KAH9654674.1 | TRINITY_DN8134_c0_g1_i1 | 71 | 4E-159 | 466 |

Abstract

Transcriptome analysis and identification of flowering regulatory genes using RNA–seq in the strobilus development of *Pinus densiflora*

Dayoung Lee

Department of Agriculture, Forestry and Bioresources

The Graduate School

Seoul National University

In the case of seed plants, flowering is an essential process for survival and reproduction, and various genetic changes occur during the transition from vegetative to reproductive growth. Compared to angiosperms, in which studies on various flowering regulatory genes have been conducted, research in coniferous species is lacking. Increasing the expression of the flowering control genes in conifers will increase cone production, induce early flowering, and accelerate tree breeding program. In this study, samples were collected from the reproductive organs of *Pinus densiflora*, and then a reference transcriptome of *P. densiflora* was generated using RNA–seq data. The transcriptomes were compared and analyzed for each reproductive organ, and biological and physiological characteristics and flowering control genes of *P. densiflora* were identified. Seven candidate genes for flowering regulatory (*Pinus*

densiflora APETALA2-LIKE(PdAP2L), DEFICIENTS AGAMOUS-LIKE(PdDAL), MINICHROMOSOME MAINTENANCE1 AGAMOUS DEFICIENTS SERUM RESPONSE FACTOR 1(PdMADS1), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1(PdSOC1), SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3(PdSPL3), UNUSUAL FLORAL ORGANS(PdUFO), WUSCHE(PdWUS)) were found in *P. densiflora*. And these genes were expressed not only in the apex and the female cones, but also in various vegetative organs. In the future, it will be possible to extract the exact nucleotide sequence of each candidate gene and verify its function through transformation. Finally, it is expected that various flowering control genes of *P. densiflora* can be registered in public databases.

Keywords : RNA-seq, *Pinus densiflora*, cone development, transcriptome, flowering regulatory genes

Student Number : 2021-28445