FEMALE STERILITY AND CARPEL DEVELOPMENT IN PLANTS

BY

SETU CHAKRABARTY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Plant Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Ray R Ming, Director of Research
Professor Raymond E Zielinski
Professor Patrick J Tranel
Assistant Professor Steven J Clough

Abstract

Angiosperms, or flowering plants, form the largest group of plants, with more than 350,000 extant species. They exhibit extensive diversity in shape, size, color, structure and organization of their reproductive organs contained within the flowers. With regard to the sexual nature of the flowers, males, females and hermaphrodites occur in nature, with hermaphroditism being the norm. About 6% of flowering plants are dioecious and 5% monoecious, supporting the widely accepted view that bisexual flowers are the ancestral condition. Unisexuality evolves from hermaphroditism by the process of random mutations affecting the female organ (carpel) or male organ (stamen) abortion. The ABCE model of floral development provides a basic underlying developmental framework of individual floral whorls across species. However, it stops short of universally explaining the occurrence of selective organ abortion, as in the case of unisexual flowers. Abortion of either one of the sexual organs is the first step towards evolving a sexually dimorphic species, mostly by a loss-of-function mutation, rarely by a gain-of-function mutation, occurring in any of a number of genes and regulatory elements involved. Ontogenic similarities between the lateral organ leaf and the flower has also led to research demonstrating increasing roles played by the plant hormone auxin in the initiation and patterning of these organs. Reproductive organs are structurally complex and critical to survival, and have been under intense research for the last three decades. However, the genetic elements and interactions between that sculpt the organs are relatively poorly understood, and neither of these models sufficiently explain the occurrence of different sex types in plants.

Stable dioecy results when the two functional genes affecting carpel and stamen development are linked in close proximity and their recessive alleles are linked in repulsion phase. The relatively low frequency of female sterile mutants in nature is indicative of the evolutionary constraints on the female organs by dint of their role in bearing the ovules and providing nutrition and protection to the next generation. It is also caused by the lesser probability of female mutations being fixed in a population as the sedentary recipient nature of carpels would drive the population to extinction. The model plant species *Arabidopsis* also reflects this dearth of female sterile mutants in laboratory studies. Being a hermaphrodite

species with perfect flowers, and with a fully sequenced genome, *Arabidopsis* is ideally suited to study floral organ development.

Carica papaya variety AU9 is an improved dioecious variety with male and female sex types controlled by a pair of nascent sex chromosomes, and with a sequenced genome. It makes for an ideal system to study the underlying genetic basis for floral sex organ development. *Arabidopsis* and papaya are both in the order Brassicales, with papaya having 2 fewer whole genome duplications than Arabidopsis. To explore and identify genomic regions and gene loci involved in floral organ development, we combined the parallel study of gene expression differences between male and female shoot apical meristem tissue of AU9 with that of sequence analysis of EMS generated female-sterile mutants in *Arabidopsis*.

The combined approach of our study identified a host of gene loci in both papaya and Arabidopsis, and 2 distinct genomic regions in Arabidopsis as putatively involved in the developmental program of the carpels. In papaya, significantly higher number of genes were found to be present in the male tissues compared to the female tissues. Known genes involved in organ development showed a distribution among transcription factors, hormone related functions, transporter proteins, and kinase proteins. We also identified chromatin related proteins which presumably work to maintain genome integrity and accuracy. Of the loci identified to be differentially expressed between males and females, a majority was found to be of unknown function. This was expected as many of the critical regulatory elements function upstream and downstream of the ABCE model and auxin responses are largely uncharacterized. In addition, there should also be crosstalk among effectors of the ABCE class genes and those of the auxin related genes as evident from the combined results of our parallel experiments. A large portion of these gene loci code for proteins containing WD40 repeats, ankyrin repeats, penta-, tetra- and tri- copeptide repeats. Although these protein motifs are found to be involved in a wide variety of physiological functions, emerging evidence of sub-functionalization and additional motif based studies are increasingly implicating them in developmental roles.

In our analysis of Arabidopsis mutants, we found genomic regions on the long arms of Chromosome 1 and Chromosome 3 to harbor single nucleotide polymorphisms (SNPs) at a higher

frequency and greater density, compared to the rest of the genome. This was an expected situation given that our mutant FS322 displayed a distorted segregation ratio of 15:1, with the mutant phenotype failing to manifest in the second backcross generation. This is suggestive of more than one gene being affected to generate the mutant phenotype. The gene loci identified to be putative candidate genes show a variety of functional roles, and unknown functions. We identified ribosomal structural and functional protein components, F-Box and U-Box proteins, and ankyrin, penta-, tetra- and tri- copeptide repeat containing proteins. The known loci identified to have SNPs include key players such as HUELLENOS (HLL), MEIOSIS DEFECTIVE 1 (MEI1), ESSENTIAL MEIOTIC ENDONUCLEASE 1B (EME 1B), SPATULA (SPT), RIBOSOMAL RNA PROCESSING 5 (RRP5), PRESEQUENCE PROTEASE 1 (PREP1) AND BRASSINOSTEROID-SIGNALING KINASE 2 (BSK2), all of which are known to play roles in female reproductive development.

Acknowledgements

My interest in the biological world from an early age is almost entirely a result of my parents' thought processes, and I would like to thank them for funding and supporting my first science experiments. Early observations and projects include digging for worms, being amazed at how composting works, noting variations in plants and animals, dissecting plants from my mother's flower beds, and trying to understand the world around me through reading vicariously.

I first developed interest in science with the instilling of a questioning and skeptical nature in me, inspired by my mother. My interest in genetics and plants developed as an undergraduate student and stem from studying and reading about the pioneering discoveries in genetics made in plants. Undergraduate education taught me the basics of classical botany and genetics, and inspired me to gravitate towards molecular biology, developmental biology, biochemistry, the importance of experimental design, and encouraged me to pursue an academic research career. I would especially like to thank my Master's advisor Dr. Rita Mahanta for not kicking me out of her labs for breaking glassware, killing lab mice or using the wrong pipette for an important experiment.

My transition into a young scientist is largely thanks to my PhD advisor Ray Ming and my dissertation committee. Ray gave me the freedom to pursue my own research ideas, and involved me in many exciting projects outside of my thesis work. I have learnt a lot in the last 5 years from Ray.

I would like to thank everyone in the Ming lab for their help over the last five years and give special thanks to Robert VanBuren, Jennifer Wai, Daniel Weber, Julie Nguyen, Neha Pandey, Fan Zhu and Will Wadlington for sharing ideas and trouble shooting. Special thanks also to Dr. Tom Jacobs and Miranda Haus for allowing me to use their space to grow my plants. I would like to acknowledge the immense support and help I received during difficult times from Dr. Kyle Bender and Jesse Miller. I would also like to thank my committee members Dr. Ray Zielinski, Dr. Patrick Tranel and Dr. Steven Clough for helpful discussion in my preliminary exam and thesis defense, and for their extreme patience and consideration. Finally, I would like to thank my friends and family for support. I could not have graduated without you.

Table of contents

CHAPTER 1: INTRODUCTION……………………..……1

CHAPTER 2: DIFFERENTIALLY EXPRESSED GENES PRIOR TO FLOWERING IN
MALE AND FEMALE PAPAYA ……………………………………………………………………………..………………..8

CHAPTER 3: GENERATION AND ANALYSIS OF FEMALE STERILE MUTANTS
IN ARABIDOPSIS…….51

CHAPTER 4: IDENTIFICATION OF GENOMIC REGIONS ASSOCIATED WITH
FEMALE STERILITY MUTANT FS322……………………………………………………………………………….76

CHAPTER 5: CONCLUSIONS…………………………………………………………………………………………….97

REFERENCES…….100

Chapter 1: Introduction

The primary objective of life on earth is to survive and reproduce. This ensures the persistence and propagation of the species, and to this end, life forms on earth have evolved a variety of mechanisms to ensure that organisms are reproduced faithfully and the genetic information is transmitted accurately from one generation to the next. Variation within a species is created and maintained by the process of sexual reproduction with haploid gametes being formed from diploid parents. Utilizing this mode of reproduction required the creation and proper development of distinct sex types- with division of labor and roles played by the 2 sexes. In animals, the general trend is for the 'female' of the species to bear the next generation and nurture it till birth, while the 'male' provides half of the genetic information in the form of the male gamete. In plants, there is a similar pattern of reproduction with males and females playing respective roles in the persistence of the species. However, as with any rule, there are exceptions as well. Hermaphrodites (male and female roles in the same individual) exist in both plants and animals, along with other alternate modes of reproduction.

The reproductive roles played by the male and female counterparts require the presence of specialized organs to carry out the processes of gamete formation, union of the gametes and subsequent protection and nurturing of the young ones. In higher plants that produce flowers (angiosperms), these specialized organs are contained within the highly modified structures of the flowers. The male organs (androecium) and the female organs (gynoecium) each have their own specific developmental programs to ensure the success of reproduction. The gynoecium is

likely the biggest factor in the tremendous success of angiosperms over their 160 million year history, as the source of the next generation, as well as the physical structure providing protection and nutrition for the progeny. The gynoecia are also the source of many different types of food for humans- fruits, nuts, beans, and cereals, enabling the sustenance of human life over the ages. The incredible diversity of the angiosperms, with more than 250,000 extant species also reflects a similar level of diversity in shape, size, and structure of the gynoecia, along with the existence of variations in the sex types of the plants and flowers themselves.

30 years ago, research into the developmental genetics of flowers helped in elucidating the first few genes involved in this process. Intense research in this area over the years, has helped in developing and fine tuning the ABCE model. This model postulated the combinatorial functions of 4 classes of genes (Coen & Meyerowitz 1991). Except for the A class gene APETALA2 (AP2), all other members of these genes classes code for MADS-domain transcription factors. The partially redundant class E genes (SEPALLATA genes) function as higher-order functionally-redundant complexes with the A, B and C class genes to identity of floral organs (Pelaz et al. 2000; Zahn et al. 2005). The B and C class of genes, namely APETALLA3 (AP3), PISTILLATA (PI), and AGAMOUS (AG), work with SEPALLATA genes to specify the anthers and carpels respectively. The C class gene AGAMOUS (AG), specifies carpel identity and floral determinacy. It also has cadastral functions that helps to restrict its expression domain (by action of AP2), and is regulated by a co-repressor complex of LEUNIG (LUG) and SEUSS (SEU) (ÓMaoiléidigh et al. 2013; Ó'Maoiléidigh et al. 2014). HUA and HUA ENHANCER (HEN) functioning in RNA metabolism, also functions in post-transcriptional regulation of AG (Chen & Meyerowitz 1999; Cheng & Chen 2004).

The ABCE model elegantly describes initiation and development of floral organs across species. However, it doesn't explain the occurrence of unisexual flowers on individuals of the same species. The model also does not explain the various defects in carpel development as seen in *yucca*, *pinoid* and *pinformed* mutants (Ellis et al. 2005; Cheng et al. 2006; Cheng & Zhao 2007; Cheng et al. 2008). The past 3 decades of intense research has uncovered a host of other genes that interact both directly and indirectly, with the ABCE genes in a spatial-temporal manner. These genes cover every aspect of development, starting from perception of the flowering signal, the initiation of the floral primordia, to the final stages of flower maturation (Krizek & Fletcher 2005). These studies also revealed the intricate involvement of plant hormones in the developmental processes, specifically of the female organs. YUCCA1, YUCCA4, PIN, PINOID (auxin hormone biosynthesis and response pathways) (Youfa Cheng et al. 2006; Cheng et al. 2007; Xing et al. 2013; Bennett et al. 1995; Benjamins et al. 2001; Furutani et al. 2004; Christensen et al. 2000; Lampugnani et al. 2013b), ACC synthase (ethylene biosynthesis pathway)(De Martinis & Mariani 1999; Boualem et al. 2008; Boualem et al. 2009; Sherif et al. 2009) are some of the well characterized hormone related genes that play roles in the female organ developmental processes. In addition to these, some other non-hormone genes such as COP1-interacting protein-related (Wei & Deng 2003; Stewart et al. 2016) have been identified, while some of the genes with unknown functions include ribosomal proteins (L39), Leucine-rich-repeat transmembrane proteins, zinc finger proteins, AP2/B3-like transcription factor family protein (Krogan et al. 2012; Zik & Irish 2003).

From an evolutionary and ontogenic perspective, flowers are highly modified leaves, containing complex organs that have been derived from leaves. Reproductive organs

(androecium and gynoecium) consist of the innermost whorls of flowers. The gynoecium, or fused carpels, contains the ovules that develop into seeds upon fertilization and the androecium, or stamens, produce the pollen that fertilize the ovules. The carpel or gynoecia itself develops into the 'fruit', and is immense economic value. The widely accepted model of carpel development is thought to represent modified leaves or sporophylls. The resemblance of carpels to leaf-like lateral organs can be seen in the homology between carpels and leaves of certain angiosperm species such as the "mother of thousands" (*Kalanchoe daigremontiana*). Similar to leaves, carpels have an adaxial-abaxial pattern and demonstrate directional auxin transport and localized auxin synthesis. Auxin utilizes a pattern of 'auxin-maxima' and 'auxin-minima' to initiate cell wall loosening and organ primordia initiation. Defects in auxin transport and synthesis has been able to account for the defects in carpel development mentioned before. The apical-basal morphogenesis and development of gynoecium in the model species *Arabidopsis* was previously explained by Nemhauser et al. in 2000. This model relied on an auxin gradient to determine the apical-basal patterning and proper organ development. Although it is an attractive model, the auxin gradient proposed has been found to be inconsistent with later data (Benkova et al. 2003; Sorefan et al. 2009; Girin et al. 2011; Larsson et al. 2013; Grieneisen et al. 2013). The auxin model has been now revised to address the shortcomings of the Nemhauser Model and reflect the similarities with leaf development. The current model is the "early action model", and is based on three main observations: timing of apical-basal patterning, the established evolutionary homology between carpels and leaves, and the emerging roles of auxin in transport, signaling and synthesis in lateral organ and leaf development (Hawkins & Liu 2014). This model emphasizes the early importance of auxin in establishment of the adaxial-abaxial boundaries and the defects

arising from the auxin disrupting mutation as well as mutations in genes that are functionally downstream of auxin function.

The efficacy of the early action model is in describing the proper patterning as a result of abaxial-adaxial boundary establishment through counter-oriented auxin flows, and in explaining the phenotypes observed in YUCCA and PIN and PID mutants. To some extent, it also goes to explain the occurrence of repeating patterns and simplified blueprint for lateral organ development. Combined with the ABCE model, this leaves room for the identification of an increasing number of genes in regulatory hierarchies that function at both transcriptional and post-transcriptional levels controlling specific aspects of developments.

To address these questions, we designed our study from two perspectives. We generated and analyzed the shoot apical meristem (SAM) transcriptome of males and females of the dioecious papaya variety AU9, to look for differences in gene expression profiles between the sexes. Simultaneously, we generated an Ethyl-methyl sulfonate (EMS) mutagenized population of *Arabidopsis* mutants and screened for female-sterile phenotypes to map genomic locations and identify candidate genes necessary for carpel development.

Papaya is an important economic crop of the tropics. It is nutritious and has medicinal values. *Carica papaya* variety AU9 is uniquely suited to our study because of many reasons. It is a dioecious variety, has nascent sex chromosomes, and a relatively small genome ($2n$= 18; 372 Megabases). The lesser number of whole genome duplications in papaya compared to *Arabidopsis* is also beneficial in studying the evolution of genetic control of floral organ development. Papaya and *Arabidopsis belong to* the order Brassicales and they diverged about

72 million years ago (Wikström et al. 2001). This facilitates comparative structural and evolutionary genomics research in both species, with the exception of copy number variation, and therefore, higher likelihood of genetic redundancy in Arabidopsis. We therefore performed RNA Sequencing to explore the differences in the expression patterns of genes in the shoot apical meristem tissue between the male and female sex types at different time points prior to flowering. The early stages were selected to identify genes involved in sex determination and differentiation of the sex types.

We also generated an EMS mutagenized population of the model plant species *Arabidopsis*. We screened the mutants for female sterile phenotypes and defects in carpel development. *Arabidopsis* and papaya being close relatives, we designed this study to map genomic regions and gene loci involved in the female sterile *Arabidopsis* mutants to the genes identified to be differentially expressed in the papaya female SAM tissue. The identified mutants defective in female organ development were used to generate mutant mapping populations bulked into mutant-like and wild type-like groups. Genomic DNA was extracted from these groups and sequenced to obtain both sequence data and allele frequency data, which was then used to genomic locations and gene identities of the candidate genes.

As discussed previously, the ABCE model is not able to fully explain the occurrence of unisexuality in angiosperms. Unisexuality has also evolved multiple times and independently throughout evolutionary history. This has occurred in different ancestral lineages and utilized different mechanisms. The current model of sex determination postulates the hermaphrodite state to be the ancestral state, and unisexual flowers evolve from this state through the selective abortion of either one of the sexual organs, by either gain or loss of function mutation (Ming et

al. 2011). These mutations could be in any of the immense number of genes that control the events at various developmental stage (Wellmer et al. 2004; Zhang et al. 2005). Auxin-based models, by themselves, also cannot explain the different sex types in the plant kingdom. Combining the ABCE model and the auxin based models to fine tune our understanding of the activity and regulation of genes in the development process. Linking the regulatory activities at the various levels of floral organ pattering regulatory hierarchy with the downstream events that lead to terminal differentiation of tissue types will enhance our understanding of these processes. It will also help discern the intricacies of the interactions between these models that lead to the proper sculpting of the different floral organs. Integrating the various pathways and components that specify and maintain floral organ identity, pattering, number, size shape and symmetry into a comprehensive system will be an ambitious task, and can be achieved though further and more extensive studies in this area.

Chapter 2: Differentially expressed genes prior to flowering in male and female papaya

Abstract

Papaya is a major tropical fruit crop originated in Central America and hermaphrodite papaya was selected during the domestication process. All wild papaya is dioecious. Papaya variety AU9 is an improved but not released breeding variety, and it has male and female sex types controlled by a pair of nascent sex chromosomes. It is an ideal system to explore the underlying genetic basis of floral sex organ development. Here we performed differential gene expression analysis of the transcriptomes and compared the relative expression patterns between the shoot apical meristem of male and female plants at 2- and 5-week before flowering. In total 17,868 transcripts were identified, out of which 1187 were differentially expressed genes (DEGs) between the two sex types. These DEGs include 58 transcription factor (TF) coding genes, 22 plant hormone related genes (HRG), 21 transporter proteins (TP), 35 chromatin related proteins and 62 kinase coding genes. DEGs showing increase in expression in Female tissue 2 weeks before flowering (2F) vs Male tissue 2 weeks before flowering (2M) and 2F vs Female tissue 5 weeks before flowering (5F) should be involved in female specification, whereas genes showing increase in 2M vs 2F and 2M vs 5M should be involved in male fate specification. DEGs associated with auxin and gibberellin mostly belong to the hormone response and transport processes, most likely involved in organ fate and sex specificity. A large portion of the upregulated genes are with 'unknown' and 'putative' function, indicating potential novel genes involved in sex determination

and differentiation that are not present in hermaphroditic model plant species. Our study provides a foundation for functional analyses of these genes with unknown functions and will improve our understanding of the developmental programs controlling the specification of male and female sex organs.

Introduction

The current global scenario with regards to the rapid increase in human population brings unique problems to our needs as a society. One of the major challenges is food security. To be able to support the growing human population, food production must increase significantly by the year 2050, to avoid a looming crisis and potentially catastrophic consequences, including the rising levels of CO2 and greenhouse gases leading to a changing climate. Except in the developed countries, human society has been primarily agriculture based and all our food finds its basic source in plants, be it as livestock feed, processed foods, or cereal and other food crops. However, the amount of arable land suitable for agriculture is limited. Therefore, there is a growing need to focus on, and devise sustainable practices and approaches towards the optimal utilization of both land as well as the plant resources available to us.

Most of the produce that we consume as food, is in the form of cereal, fruit, vegetables and tubers. Botanically speaking, the first 2 fall into the technical category of ‘fruit’, made by the female reproductive organs of a plant. In plant species that display sexual dimorphism, the sex ratio is usually 1:1. Typically, this translates to yield losses as well as losses in terms of time,

farming area and resources. It would, therefore, be of tremendous utility if it was possible to re-engineer dioecious crop plants to true breeding hermaphrodites.

In addition to having an important applied direction, the study of the dynamics of floral organ development is also a fundamental area of research that will advance our understanding of the complex nature of sex determination and sexual polymorphisms observed in the plant kingdom. Hermaphroditism is the norm in the plant kingdom, with only 6% of angiosperms having dimorphic sexual nature. Papaya (*Carica papaya* L.) is a major tropical fruit crop with outstanding nutritional and medicinal values. It also has certain characteristics such as small genome ($2n$ = 18, 372 Mb), incipient sex chromosome, and both dioecious and trioecious sex types that make genomic and genetic studies relatively easy and cost effective (Arumuganathan & Earle 1991), (Ming et al. 2008), (Wang et al. 2012). The model plant species Arabidopsis is considered to have perfect or bisexual flowers while papaya exhibits both tri and dimorphic characters. This makes the study of papaya shoot apical meristem in parallel with that of Arabidopsis especially interesting as contrasting features in terms of expression and regulation patterns of genes will expectedly stand out and will be relatively easy to dissect. Arabidopsis, with its extensive knowledgebase of gene families and functional descriptions and characterizations, serves as the ideal contrast to that of papaya, where the evolution of plant sex chromosomes is nascent.

The lack of a recent whole genome duplication in papaya also helps to study angiosperm genome evolution and, in particular, the evolution and genetic control of floral organ development (Gschwend et al. 2012). Papaya and *Arabidopsis belong to* the order Brassicales and they diverged about 72 million years ago (Wikström et al. 2001). This facilitates comparative structural and evolutionary genomics research in both species, with the exception of copy

number variation, and therefore, higher likelihood of genetic redundancy in Arabidopsis. These characteristics make papaya an excellent model for tropical tree fruit crops while drawing comparisons and parallels with Arabidopsis.

Papaya has a short juvenile phase and most cultivars flower in 3-4 months. The objective of this project is to identify genes involved in sex determination and differentiation, which occur before flowers become visible. We designed this study to explore the differences in gene expression and patterns at 5 and 2 weeks before flowering between the 2 sex types in the shoot apical meristem (SAM) of the dioecious papaya variety AU9.

Plant Material

Carica papaya variety AU9 plants were grown in Kunia Station at Hawaii Agriculture Research Center on Oahu, Hawaii and two sex types of shoot apical meristem (SAM) were collected at two developmental stages (2 and 5 weeks before flowering). The SAM samples were then frozen in liquid nitrogen immediately and stored at -80°C.

RNA extraction and library construction

SAM tissues were collected at the two developmental stages, with two biological replications for each sex type. RNA was extracted from SAM tissues using TRIzol Reagent (Cat. No. 15596-026) and genomic DNA was removed by Ambion DNA-*free*™ DNA removal kit (Life technologies, #AM1906). The DNase-treated RNA was then subjected to Illumina TruSeq

Stranded mRNA Sample Preparation Kit LT v2 (Illumina, #RS-122-9004DOC) for library construction according to manufacturer's instruction. The multiplexed libraries were pooled and sequenced in two lanes of 150nt paired-end sequencing using HiSeq 2000. Quality control was performed using fastQC to remove adapters and low quality sequences, allowing retention of high quality reads of approximately 100bp read lengths, averaging 37,853,248 reads per library.

Sequence read alignment and differential expression analysis

The trimmed sequence reads were aligned to *Carica papaya* gene model (based on the draft genome of the 'SunUp' variety) using the splice junction mapping program Tophat version 2.1.1 default settings (Trapnell et al. 2012). The alignment file was then subjected to Cufflinks Tool Suite version 2.2.0 (Trapnell et al. 2014) to assemble the transcriptome, and perform differential expression analysis between the samples according to best practices (a nonparametric, annotation-guided approach to estimate the means and variances of transcript FPKM values under different conditions, using Student's *t*-tests to identify differentially expressed transcripts). In this study, genes with more than 3-fold changes (i.e. log2 fold change ≥ 1.58), p value ≤ 0.05, and false discovery rate (FDR) ≤ 0.05 are classified as differentially expressed genes. For each developmental stage, two pairwise tests were carried out: i) female versus male (2F vs 2M and 5F vs 5M); ii) female versus female (2F vs 5F); and iii) male versus male (2M vs 5M). The first two conditions are aimed at exploring genes regulated in both female and male SAM at similar stages of development, while the latter two conditions are aimed at genes involved in transition from 5 weeks to 2 weeks before flowering in both female and male shoot

apical meristem. Transcripts of interest were then functionally annotated using the Blast2GO software package version 4.0 (Conesa & Stefan 2012).

Functional annotation

Sequence-similarity Blast searches of all papaya predicted protein sequences were conducted with a typical cut-off *E*-value of 10^{-3} against several publicly available protein databases: The National Center for Biotechnology Information (NCBI) non-redundant (Nr) protein database, Clusters of Orthologous Groups (COGs), and Kyoto Encyclopedia of Genes and Genomes (KEGG). Gene Ontology (GO) terms describing biological processes, molecular functions and cellular components were assigned to the predicted genes. Biological pathway enrichment analysis was done based on known confirmed and putative pathways according to the KEGG database against all NR database hits and confirmed by matches with the Arabidopsis biological proteins and pathways.

Results

Differentially expressed genes between two sex types of papaya flowers

Shoot apical meristem tissue from two sex types (female and male) at two developmental stages (2 weeks and 5 weeks before flowering) were sequenced with two biological replicates per samples. In total, 151,412,990 paired end 100nt reads were obtained for all 4x2 samples with an average of 37,853,248 reads.

During late floral development of meristem tissue (2 weeks before flowering), gene expression difference is the least between female and male SAM (2F vs 2M) with 372 genes expressed differently, compared to early stages (5F vs 5M) (Fig. 1.1 and 1.2) with 606 genes expressed differentially. Common differentially expressed genes showed a reversal of expression patterns between the two comparisons, in other words, over-represented genes in one comparison were underrepresented in the other. Similarly, there was a trend of expression pattern reversal between male and female SAM tissue in the transition from early stage to late stage (2F vs 5F and 2M vs 5M) with 408 and 724 genes over-represented respectively. Interestingly, a significantly high number of genes were up-regulated in the female transition compared to the male transition where a large number of genes were down-regulated. It is important to note here that the trend of expression reversal was observed in the same gene loci across all sample comparisons (Fig. 1.1). Many gene loci that qualified under the selection category of 'differential expression' in the above comparisons were expressed commonly across at least 2 sample types. For our purpose, we further filtered our list of gene loci by selecting for loci that are uniquely expressed in only one comparison (Fig. 1.2).

Functional annotation, classification and KEGG mapping

Sequence-similarity Blast searches of all papaya predicted protein sequences were conducted with a typical cut-off *E*-value of 10^{-3} against several publicly available protein databases: The National Center for Biotechnology Information (NCBI) non-redundant (Nr) protein database, Clusters of Orthologous Groups (COGs), and Kyoto Encyclopedia of Genes and

Genomes (KEGG). Gene Ontology (GO) terms describing biological processes, molecular functions and cellular components were assigned to the predicted genes by Blast2GO program (Conesa & Stefan 2012) based on the NR blastp output.

In the comparisons between sex type SAM transcriptomes at 2 weeks before flowering stage, 23 gene loci (out of 73 loci) were upregulated in the female sample compared to the male sample (Fig. 1.3). Upregulated gene loci include *transcription factor TRY*, *senescence specific cysteine protease SAG39*, *respiratory burst oxidase homolog D-like*, *zinc finger 4-like*, *glutathione-s-transferase F13*, *epidermal patterning factor 2* etc. Out of these, *epidermal patterning factor 2* and *respiratory burst oxidase homolog D-like* show the highest difference in fold change (log2 fold change of 3.1 each). In the comparison between sex type SAM transcriptome at 5 weeks before flowering, 51 gene loci (out of 152 loci) were upregulated in the female sample compared to the male sample (Fig. 1.3). Upregulated gene loci include *sacsin*, *midasin*, *dnaJ homolog subfamily C GRV2-like isoform X1*, *auxin transport BIG*, *calcium-dependent lipid binding family isoform 1 and isoform 3*, *BEACH domain containing*, Chromatin remodeling and maintenance, signaling pathways etc. It is interesting to note that the over represented gene loci in the 5F samples mostly represent gene loci that have been implicated in the maintenance of stem cell status of the developing meristem, while those in the 2F samples mostly hint towards differentiation pathways. However, some of the gene loci showing the highest fold changes are either hypothetical proteins or are without a match to the NCBI database. This could be the indication of novel transcript discovery and indicative of hitherto uncharacterized candidate genes responsible for female organ development.

In the comparisons between the two time points within the same sex type (i.e. between 2F and 5F), 9 gene loci (out of 86) were upregulated in the 2F sample compared to the 5F sample (Fig. 1.3). Upregulated gene loci include *Ribosomal RNA small subunit methyltransferase E*, *Ras GTPase-activating IQGAP3*, *heptahelical transmembrane 4-like*, *neutral ceramidase-like*, *carbonic anhydrase family* etc. Most notable is that the gene loci displaying the highest difference in fold change (log2fold change of 4), was found to have no match to the NCBI database. In the comparison between the 2 time points within the same sex type (2M vs 5M) 89 gene loci (out of 231 loci) were upregulated in the 2M sample compared to the 5M sample (Fig. 1.3). These include *RING-H2 finger ATL74-like*, *sacsin*, *serine threonine- kinase TOR-like*, *auxin transport BIG*, *Tudor PWWP MBT superfamily isoform 1*, *calcium and calcium calmodulin-dependent serine threonine-kinase*, *serine threonine- kinase ATM*, *midasin isoform X1*, *sacsin*, *endoribonuclease Dicer homolog 1*. In this case too, one of the loci with the highest fold change in expression is a Cysteine Histidine-rich C1 domain family protein (log2FC=3.7 in 2Fv2M) while the other (*evm.TU.supercontig_2050.1:7-234*; log2FC=4.6 in 2Fv5M), has no blast hit but shows partial hits to AT4G24210.1 (F-box family protein) and AT5G42780.1 (homeobox protein 27), when searched against the Arabidopsis database (TAIR10). A list of the other loci with high fold changes and their matches in the TAIR10 database is shown in Table 1.1.

The SAM in plants can be thought of as having 2 major roles. Maintenance of stem cell for the growth and proliferation of the plant, as well as differentiation of cells on the periphery, towards the fate of organ development. The transcriptome of the SAM should therefore be representative of biological pathways that are involved in these biological processes. Consistent with this notion, we observed an overrepresentation of both DNA repair, chromatin

maintenance, DNA replication and transcription pathways- geared towards maintaining the stem cell niche, as well as the overrepresentation of deterministic pathways such as hormone signaling, hormone biosynthesis, biosynthesis of secondary metabolites that form the backbone of phytohormones, ubiquitin mediated protein degradation, lipid biosynthesis and signaling pathways in our selected list of 1187 DEGs representing the global functional transcriptome.

Global functional analysis of the DEGs was performed using the software Blast2GO, which gave us a total of 842 proteins which had at least one associated GO term. The assigned GO terms belonged to three main classes: molecular function, biological process and cellular component. The sequence distribution in the 'molecular function' was split into 43 categories, out of which, 'ATP binding' (118), 'anion binding' (93), 'protein binding' (92), and 'zinc-ion binding' (81) were the most common (Fig. 1.6). Apart from these broadly general categories, 'signal transducer activity' (59), 'transcription factor activity' (58), DNA repair and chromatin maintenance (41), and 'protein serine/threonine kinase activity' (39) are the other highly represented categories (Fig. 1.6).

In the Biological process class, 'Regulation of biological process' (278), 'response to stimulus and signaling' (256), 'Growth and reproduction' (94), and 'developmental process' (93) are predominant (Fig. 1.7). The 'cellular component' class 'Intracellular component', 'membrane component' and 'protein complex' comprised the majority of the sequence distribution. We then applied GO term enrichment analysis of the 1187 DEGs to understand the functional enrichment of the DEGs, using the Arabidopsis orthologs and comparing against the complete known database of the Arabidopsis functional ontology using PANTHER and GO Consortium (Thomas et al. 2003; Ashburner et al. 2000). A total of 144 GO terms were enriched in 'biological process',

‘molecular function’ and ‘cellular component’ (Table 1.2). Apart from the expected ‘metabolic’ and ‘catalytic’ general categories, the most significantly overrepresented terms were ‘cell differentiation’, ‘cytokinesis’, ‘response to endogenous stimulus’ within ‘biological process’ (Table 1.2); ‘voltage-gated potassium channel activity’, kinase activator activity’, ‘calmodulin binding’ within ‘molecular function’; and, ‘cell part’, ‘intracellular’ and ‘organelle’ in the ‘cellular components’. The enriched categories are consistent with each other, as well as with the expected role played by the SAM in stem cell propagation and differentiation.

We performed analysis of biochemical pathways associated with the DEGs based on the orthology shared with Arabidopsis genes according to the KEGG database. A total of 35 pathways, grouped into 3 categories- metabolism, genetic information processing and signaling, were upregulated (Table 1.3). Among the metabolic pathways, phenylpropanoid biosynthesis (ko00940), diterpenoid biosynthesis (ko00904), terpenoid backbone biosynthesis (ko00900), sesquiterpenoid and triterpenoid biosynthesis (ko0009) and metabolism of xenobiotics by cytochrome P450 (ko00982) are the most biologically significant. Given the role of the shoot apical meristem, upregulation of genetic information processing pathways consisting of base excision repair (ko03410), nucleotide excision repair (ko03420), and mismatch repair (ko03440) are also consistent with the biological processes expected to be upregulated for the maintenance of the stem cell niche of the SAM. Signaling pathways that were upregulated globally were MAPK signaling (ko04071), mTOR signaling (ko04150), hormone signal transduction (ko04075) and sphingolipid signaling (ko04071).

DEGs related to transcription factors, plant hormones, Chromatin -related, transporters and kinases

We categorized the DEGs according to their functional classification and were able to identify 58 transcription factor (TF) coding genes (Table 1.4), 22 plant hormone related genes (HRG) (Table 1.5), 21 transporter proteins (TP) (Table 1.7), 35 chromatin related proteins Table 1.6) and 62 kinase coding genes.

Heat maps constructed from the expression data of these DEGs show interesting dynamic patterns of expression (Fig: 1.1 and 1.2). Hierarchical clusters of these DEG classes contain genes that show a decrease in expression from 5W to 2W regardless of sex type and make a strong suggestion that they should be involved in stem cell maintenance. Genes that show increase from 5W to 2W regardless of sex type, therefore, should be involved in fate determination and differentiation. Similarly, DEGs showing increase in expression in 2F vs 2M and 2F vs 5F should be involved in female specification, whereas genes showing increase in 2M vs 2F and 2M vs 5M should be involved in male fate specification.

The HRGs showing these patterns of dynamic expression difference are related to ABA, Auxins, cytokinins, gibberellin and jasmonic acid pathways. DEGs associated with auxins mostly belong to the hormone response and transport processes, rather than being involved in the biosynthesis of the hormone (Table 1.5). This gives us an indication that appropriate regulation of auxin response and transport of the hormone biomolecule are more important for organ fate and sex specificity prior to the development of the floral organs. This is supported by the evidence of auxin playing a role in loosening of cell wall at locations of maximal accumulation, leading to

the formation of organ primordia (Okada et al. 1991; Friml et al. 2003; Lampugnani et al. 2013b). A similar trend is shown by the DEGs associated with the hormone gibberellin. This trend however, does not follow when it comes to jasmonic acid and ABA. In case of both these hormones, the DEGs are involved in the biosynthesis pathways of these hormones.

Chromatin remodeling, replication, repair and assembly are critical to proper maintenance of stem cell niche in the growing SAM. Differentiation pathways are mostly observed in the peripheral zones of the SAM, where the development of the organs are guided by formation of zones of auxin minima and maxima, in concert with a gradient of cytokinin localization. The identification of the specific DEGs involved in chromatin maintenance and hormone localization are congruent with previous studies in similar areas (Mattsson et al. 2003; Mina & Hitoshi 2009; Lampugnani et al. 2013b)

Putative and uncharacterized DEGs

In our study we have also been able to identify 56 gene loci that are hitherto uncharacterized and 186 gene loci that have no matches with multiple publicly available databases. This suggests that organ development in flowers involves many gene loci that have not yet been identified or characterized. The current study therefore lays down the ground work for future explorations into the function and role of these uncharacterized gene loci in the developmental dynamics of sex specific organs in dioecious species of plants. Future work would involve protein motif identification and specific gene perturbation experiments to study the exact effect of these loci and could unravel a host of regulatory elements in terms of transcriptional

gene regulation, post transcriptional and translational modulations of gene as well as protein-protein interactions that shape the developmental and morphological landscape of floral organ development downstream and upstream of the canonical ABCE model of flower development.

Discussion

The primary objective of this study was to identify genome wide changes in gene expression between the two sex types of papaya at 2 weeks and 5 weeks prior to flowering. Here we carried out paired end sequencing of RNA-Seq libraries prepared from mRNA isolated from the shoot apical meristem of the dioecious papaya variety AU9 grown under the same conditions at Hawaii Agricultural Research Center Kunia station. High throughput sequencing generated more than 151 million filtered reads with nearly 80% of the total reads being uniquely mapped to the papaya reference genome. The resulting transcriptome of unique mapped reads were then used to perform the downstream analysis of normalized gene expression (FPKM), gene ontologies, pathways and other functional categories.

Our results revealed changes in the expression profiles of multiple genes involved in hormone signaling pathways and regulatory networks, chromatin maintenance and DNA repair pathways, transcription factors and kinase mediated processes. In addition to these, there were also a significant portion of the DEGs with putative and predicted functions indicating that floral development after the initiation of primordia involves many uncharacterized players. As expected, these uncharacterized gene loci are also in agreement with differential expression levels in specific tissue types. Some of these loci, such as *evm.TU.supercontig_25.161_179-378*

(*PREDICTED: uncharacterized protein LOC101291165 isoform X1)* display 2 fold increase in expression while transitioning from 5 weeks to 2 weeks before flowering in the female sample and the opposing pattern of 3 fold decrease in expression in going from 5 weeks to 2 weeks in the male sample (supplemental data).

Plant hormones play a vital role in flower sexuality and fertility. Some are required for the proper development and functioning of both male and female organs, while others play specific roles in the male or female organs. Auxins are essential in the development of both male and female organs (Okada et al. 1991), (Sessions & Zambryski 1995), (Y Cheng et al. 2006), (Wu et al. 2006), (Signaling et al. 2013). In Arabidopsis, auxin is required for the formation of all floral organs, acting by promoting cell division and growth via creation of zones of auxin maxima and minima (Lampugnani et al. 2013a), (Li et al. 2016). Disruption of genes associated with auxin signaling, biosynthesis and transport results in flowers with aberrant organs (Okada et al. 1991), (Nagpal et al. 2005), (Cecchetti et al. 2008). The *arf6 arf8* double mutant and mutants with miR167 resistant versions of *ARF6* and *ARF8* display defects such as shortened stamen, gynoecia, indehiscent anthers, abnormal ovules and short petals (Nagpal et al. 2005), (Wu et al. 2006). Abnormal female fertility is also observed in the *Arabidopsis floral organs in carpel* (*foc*) mutant, caused by the increased expression of *ARF 10, 16* and *17* via lack of its negative regulator miR160 (Liu et al. 2010). The auxin biosynthesis *YUCCA* genes in Arabidopsis result in non-functional reproductive organs in the *yuc1yuc4* double mutant and non-elongating stamens in the *yuc1yuc6* double mutant (Y Cheng et al. 2006). The auxin transport genes such as PIN-FORMED (PIN) also result in various phenotypes in the range of *pin* allelic mutants. Here we see a 3-fold increase in expression of *auxin transport BIG* in the female tissue between 5 and 2 weeks before flowering,

while the opposite pattern is seen in the male sample. Similar expression patterns are observed in *auxillin 1*. On the other hand, *auxin-responsive SAUR32-like* and *auxin-responsive SAUR68-like* genes display a decrease in expression in the transition from 5 week to 2 week time points in both male and female tissues Fig. 1.10 and 1.11).

Jasmonic acid is known to promote male but suppress female development in Arabidopsis (Stintzi & Browse 2000), (Mandaokar et al. 2006) and this trend follows in maize as well. Jasmonate biosynthesis is crucial for proper stamen elongation and *opr3* mutants are defective in conversion of linolenic acid to JA. However, in papaya, there seems to be no conclusive evidence towards female suppression by the above-mentioned mechanism. In our study, gene loci involved in the jasmonate pathway exhibit similar patterns of increased expression in the 2-week tissue compared to 5-week tissue in both sex types. However, it should be noted that the female tissue has a 4-fold increase while the male tissue only displays a 2-fold increase in expression (Fig. 1.10 and 1.11). This seems to be similar to tomato where JA is also important for maintaining female fertility (Li et al. 2004), indicating a potential divergence in the roles played by jasmonic acid.

The role played by gibberellin in floral organ development is of contradictory nature. In Arabidopsis, GA is critical for proper male development, however in maize, it induces arrest of stamen development while preventing carpel abortion (Li et al. 2004), (Hu et al. 2008), (Fujioka et al. 1988), (Dellaporta & Calderon-urrea 1994). In Arabidopsis, the *ga3ox1 ga3ox3* double mutant defective in GA biosynthesis exhibits high frequency of sterility on the lower siliques, with varying degree of fertility restoration of siliques higher up on the inflorescence around the 20th inflorescence. This is caused by the abnormal dehiscence of anthers and shortened filaments (Hu

et al. 2008). In maize however, GA deficient mutants fail to suppress or abort male organs on the female ears (Fujioka et al. 1988), (Dellaporta & Calderon-urrea 1994). This GA sensitivity in opposing terms could be due to the inherent differences in monocots and dicots. However, in papaya, gibberellin is not known to be an effector in the development of reproductive organs or sex expression. GA does play roles in enhancing secondary sexual characters such as an increase in peduncle length, number of flowers, and increased branching (Han et al. 2014). This is consistent with the lack of change in expression patterns in GA pathway genes in our RNA-Seq dataset (Fig. 1.10 and 1.11).

Out of the 58 TFs identified as having significantly different expression patterns (Table 1.4), a significant portion (16) exhibit expression changes in the transition from 5 week tissue to 2 week tissue in both male and female sex types. These TFs include *calmodulin binding transcription activators, ERRF114, myb A, NAC domain containing TFs, T-box transcription factor isoform 1, TATA-binding associated factor BTAF1 isoform X1, transcription factor jumonji domain, transcription factor TRY, transcription factor UPBEAT1, truncated transcription factor CAULIFLOWER A-like,* and *translational activator GCN1* (Fig. 1.8 and 1.9). In terms of our study, genes that exhibit a reduction in expression levels during the transition from 5 weeks to 2 weeks regardless of the tissue type, indicate their involvement in the maintenance of the undifferentiated stem cell niche. Following the same argument, genes displaying an increase in expression from 5 weeks to 2 week time points are suggestive of their involvement in deterministic pathways and processes leading to the differentiation of stem cells to either male or female cell types and therefore tissue formation of these sex types. Within this smaller group of genes, some show an increase only in one of the sex types. Some of the TFs we found to show

this pattern of increasing expression are *CCR4-NOT transcription complex subunit 1-like* (2F), *NAC domain containing 35 like* (2M), *Probable WRKY transcription factor 70 isoform X2* (2M), *transcription repressor OFP6-like* (2M), *transforming growth factor beta receptor associated 1* (2M) (Fig. 1.8 and 1.9). Overall, the number of TFs that have their expression modulated in only one sex type are in the male subset rather than the female subset. This agrees with previous studies which demonstrate that a larger number of gene loci are involved in the processes leading to the proper development of male sexual organs, compared to female development programs.

In addition to the TFs, most of the gene loci identified to be involved in chromatin maintenance and DNA repair are overrepresented in the 5 week samples regardless of sex type of the tissue. The only exceptions are *DNAJ heat shock N-terminal domain containing, general DNA repair, chromatin structure remodeling BSH,* and *chromatin assembly factor 1 subunit FAS1 like*, that exhibit higher expression values in the 2-week tissue regardless of sex type (Fig. 1.12 and 1.13). This suggests that these gene loci are involved in chromatin remodeling and repair processes, which are critical in ensuring reproductive success.

Our study identified a large portion of known genes and the processes they are involved in various aspects of growth and development. The patterns of expression are consistent with previous studies and logical expectations, and provide support to the validity and accuracy of the RNA-Seq data. We were also able to identify a significant number of unknown gene loci and gene loci with putative and predicted functions. We demonstrated that many biological processes and genes involved in them are shared between developmental programs of male and female sex organs. Many of the DEGs that were upregulated in one sample compared to the other represent hormone pathways, stem cell maintenance pathways, cell differentiation pathways, and cellular

signaling pathways. Further study of the unknown gene loci will be required to identify these unknown loci and understand their physiological and molecular function in relation to the developmental programs of sexual organs. These genes may represent a host of potential targets for manipulation to re-engineer dioecious plant species to the ancestral hermaphroditic state, to improve our understanding of the intricate and complex gene regulatory networks underlying sex specification, and therefore, the evolution of sex chromosomes in land plants.

Tables and Figures

Table 1.1: Manually curated orthologs for 25 gene loci with high degree of expression variance

Locus	BLAST hit	Accession No.	TAIR 10 BLAST hit (Protein DB)	2f fpkm	2m fpkm	5f fpkm	5m fpkm
evm.TU.supercontig_2050.1_7-234	NA	AT5G56420.2	F-box/RNI-like/FBD-like domains-containing protein	2.78352	36.4194	70.1816	32.314
evm.TU.supercontig_823.2_14-182	NA	AT5G62230.2	ERECTA-like 1	18.4288	96.8251	NA	9.83469
evm.TU.supercontig_107.93_40-163	NA	AT5G40010.1	AAA-ATPase 1	582.799	165.28	21.9093	922.202
evm.TU.supercontig_109.10_0-315	NA	AT4G15460.1	glycine-rich protein	5.37432	1.44545	NA	4.58014
evm.TU.supercontig_126.23_0-124	NA	AT4G33210.1	F-box family protein	510.816	168.821	325.489	916.172
evm.TU.supercontig_14.14_0-157	NA	AT1G10640.1	Pectin lyase-like superfamily protein	978.871	386.224	427.412	1425.85
evm.TU.supercontig_16.102_57-235	NA	AT5G64420.1	DNA polymerase V family	95.4006	24.1906	46.2781	129.866
evm.TU.supercontig_17.109_4-291	NA	AT5G10370.1	helicase domain-containing protein	7.63361	2.61402	4.81939	11.9331
evm.TU.supercontig_2.140_280-540	NA	AT2G20580.1	26S proteasome regulatory subunit S2	3.33959	4.74931	1.52091	15.7291
evm.TU.supercontig_229.9_4-237	NA	AT1G10930.1	DNA helicase (RECQl4A)	14.3714	2.90226	5.28943	13.8696
evm.TU.supercontig_30.69_811-969	NA	AT5G61940.1	Ubiquitin carboxyl-terminal hydrolase-related protein	38.2854	29.3737	17.8128	123.067
evm.TU.supercontig_44.105_56-182	NA	AT2G42790.1	citrate synthase 3	466.335	48.4247	206.922	276.826
evm.TU.supercontig_53.62_38-428	NA	AT5G66240.3	Transducin/WD40 repeat-like superfamily protein	4.79881	1.70075	1.68574	5.95007
evm.TU.supercontig_58.56_0-149	NA	AT5G19700.1	MATE efflux family protein	NA	85.0463	90.3973	252.965
evm.TU.supercontig_59.48_29-563	NA	AT4G34830.1	Pentatricopeptide repeat (PPR) superfamily protein	NA	0.484586	0.512513	2.39222
evm.TU.supercontig_6.344_1-192	NA	AT4G39040.2	RNA-binding CRS1	33.086	9.21766	3.11954	28.3273
evm.TU.supercontig_7.223_14-364	NA	AT2G42100.1	Actin-like ATPase superfamily protein	5.03872	2.43462	0.524374	7.62703
evm.TU.supercontig_74.36_2-330	NA	AT2G37020.2	Translin family protein	5.20955	1.04345	NA	3.85557
evm.TU.supercontig_96.4_39-246	NA	AT3G61380.1	Phosphatidylinositol N-acetyglucosaminlytransferase subunit P-related	10.8382	7.17524	2.27645	29.8552
evm.TU.supercontig_107.93_40-163	NA	AT3G12720.1	myb domain protein 67	582.799	165.28	21.9093	922.202
evm.TU.supercontig_14.117_61-442	NA	AT5G12400.1	DNA binding;zinc ion binding;DNA binding	1.74902	1.30187	0.738198	3.9398
evm.TU.supercontig_14.14_0-157	NA	AT3G10030.2	aspartate/glutamate/uridylate kinase f	978.871	386.224	427.412	1425.85
evm.TU.supercontig_4.87_5-154	NA	AT5G38950.1	RmlC-like cupins superfamily protein	69.1913	69.7786	25.3172	241.376
evm.TU.supercontig_6.344_1-192	NA	AT1G58037.	Cysteine/Histidine-rich C1 domain family protein	33.086	9.21766	3.11954	28.3273
evm.TU.supercontig_7.223_14-364	NA	AT2G42100.1	Actin-like ATPase superfamily protein	5.03872	2.43462	0.524374	7.62703

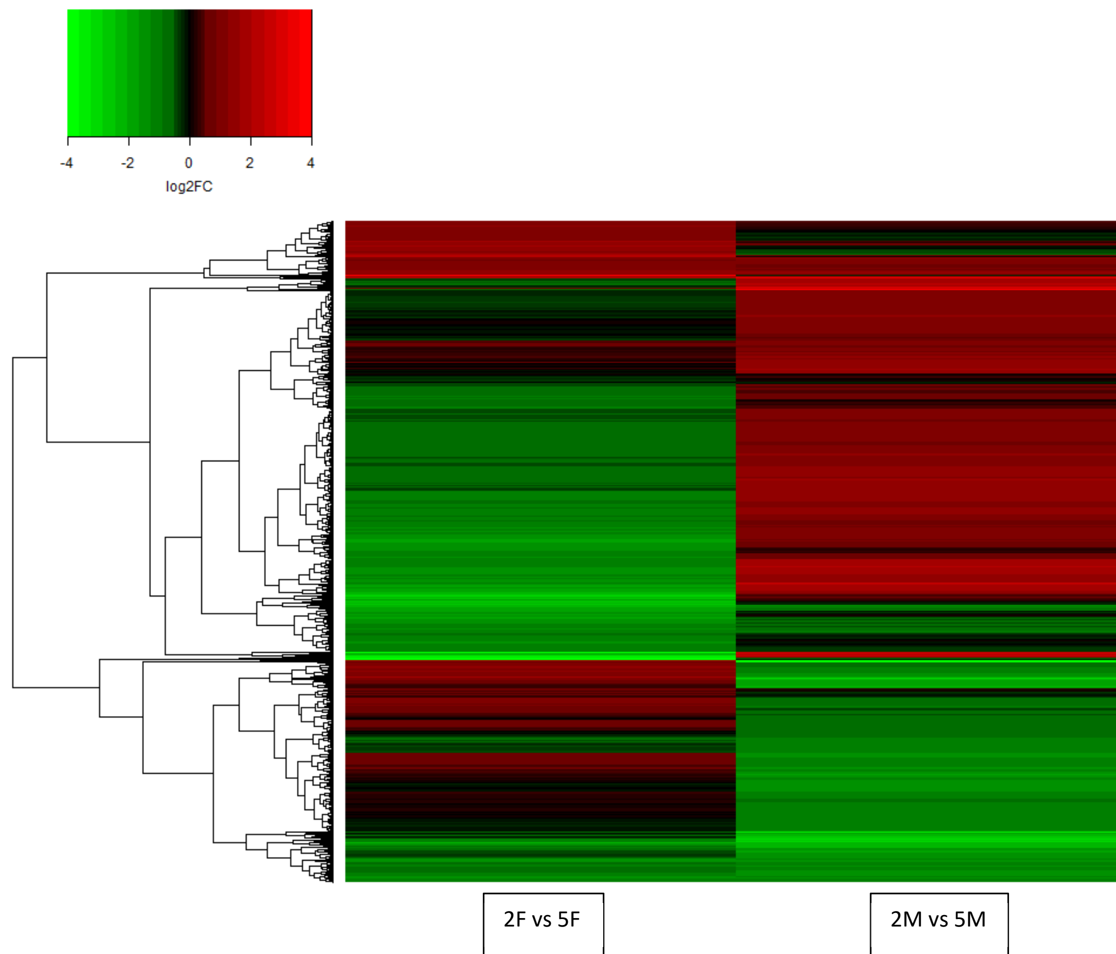

Figure 1.1: Hierarchical heat map showing log2 transformed fold change in expression (fpkm). Left panel shows expression difference between timepoints of 2 weeks and 5 weeks before flowering event in the female SAM (2F vs 5F). Right panel shows expression difference bewteen same timepoints in the male SAM (2m vs 5M). Green depicts downregulation and red depicts upregulation.

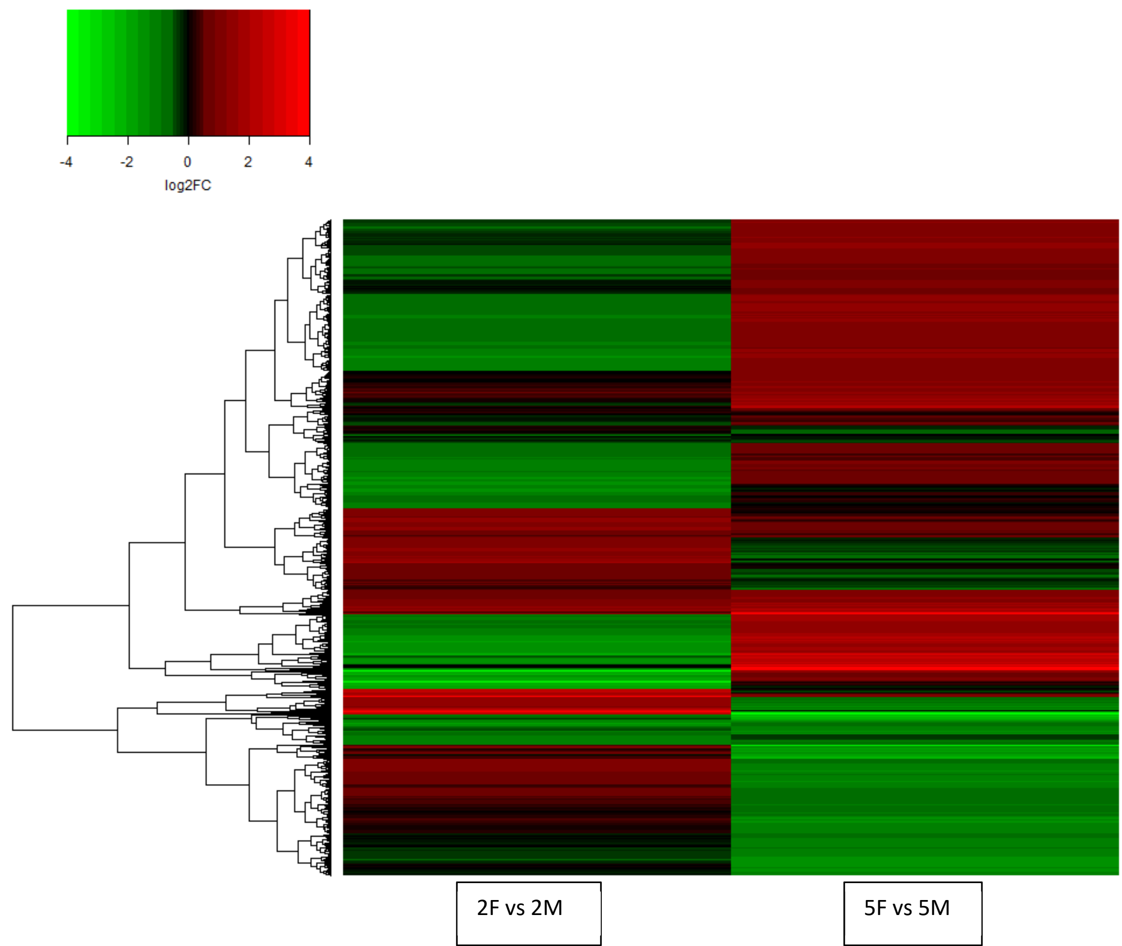

Figure 1.2: Hierarchical heat map showing log2 transformed fold change in expression (fpkm) between sex types. Left panel shows expression difference between SAM of female vs male plants at 2 weeks before flowering event (2F vs 2M). Right panel shows expression difference bewteen female and male SAM at 5 weeks before flowering event (5F vs 5M). Green depicts downregulation and red depicts upregulation.

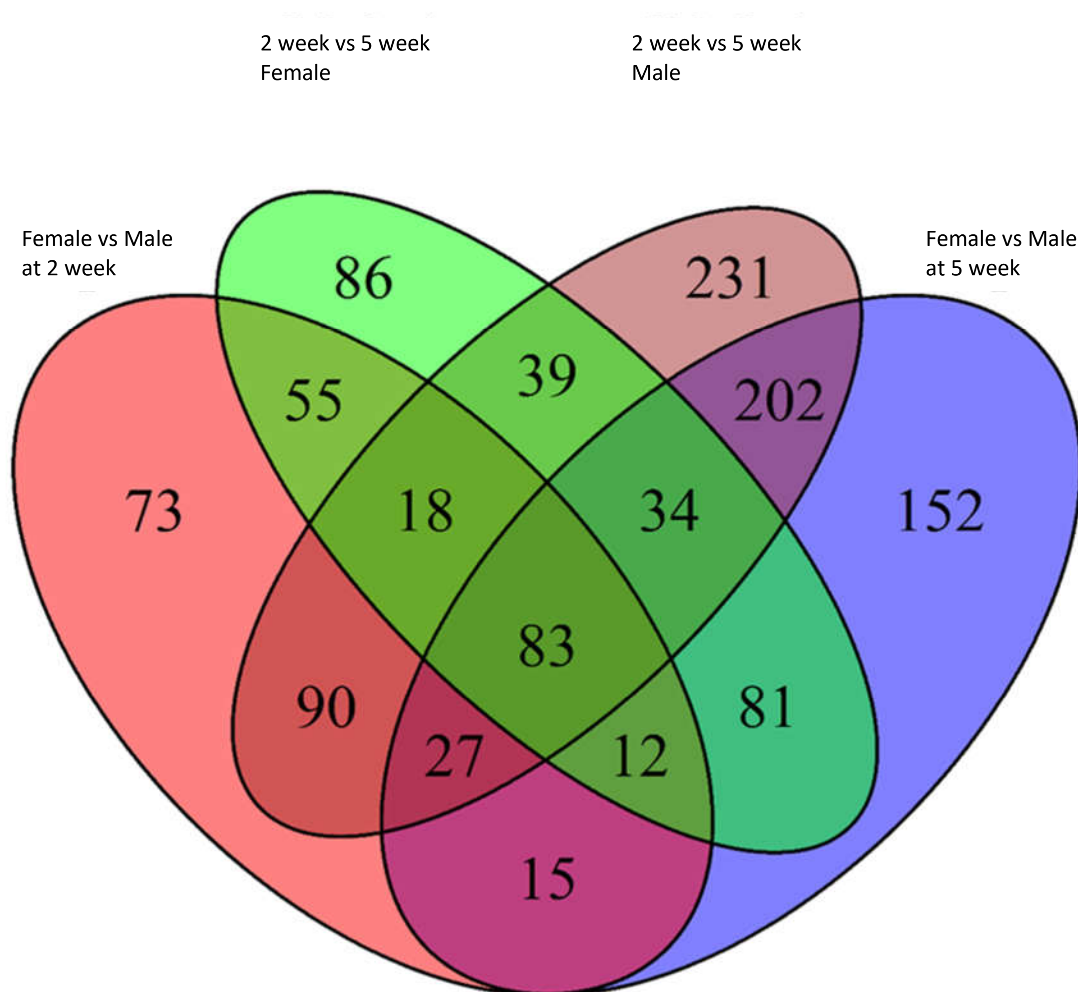

Figure 1.3: Venn diagram showing relative differences in overlapping and exclusively expressed gene counts in the SAM tissue. 73 genes are exclusively expressed in the female tissue 2 week before flowering compared to the male, 86 genes exclusively in the female SAM 2 weeks before flowering compared to 5 weeks before flowering, 231 genes exclusively in the male SAM 2 weeks before flowering compared to 5 weeks, and 152 genes exclusively expressed in the female SAM tissue at 5 weeks before flowering stage compared to the male SAM.

GO distribution: Top 20

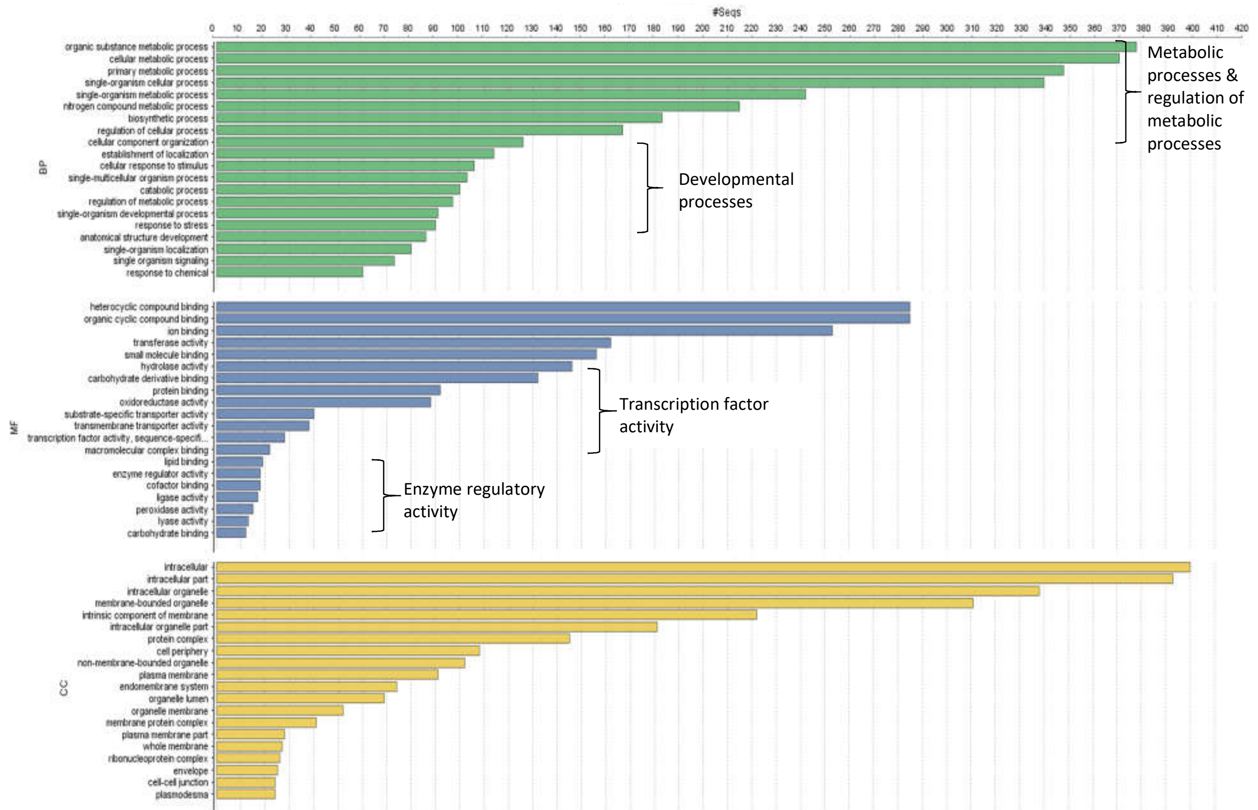

Figure 1.4: Distribution of the top 20 GO annotation terms (level 3) by number of DEGs. We sorted the distribution by 3 major GO ontologies biological process (BP), molecular function (MF) and cellular component (CC)

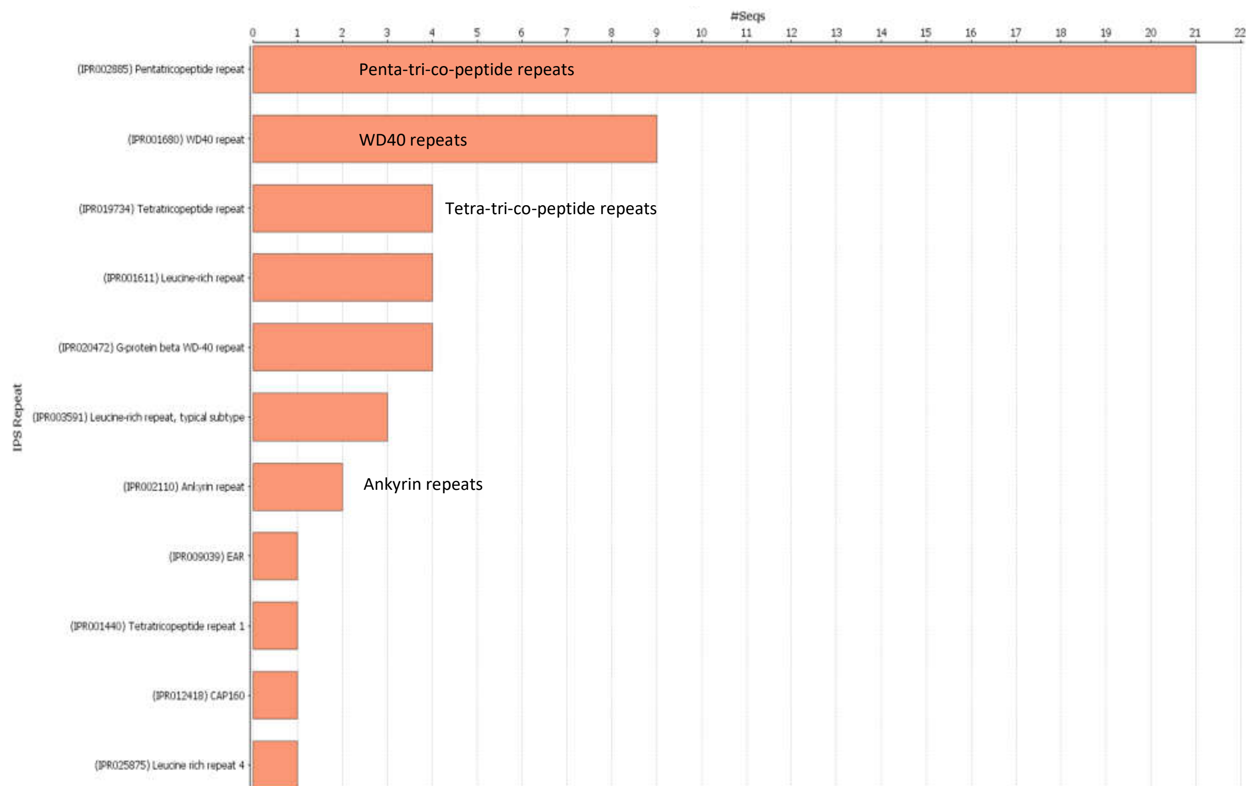

Figure 1.5: InterPro scan result, sequence distribution among the DEGs. The most commonly found protein motifs are Pentatricopeptide repeats, WD40 repeats, Leucine-rich repeat, and Ankyrin repeats.

Number of Sequences [Molecular Function]

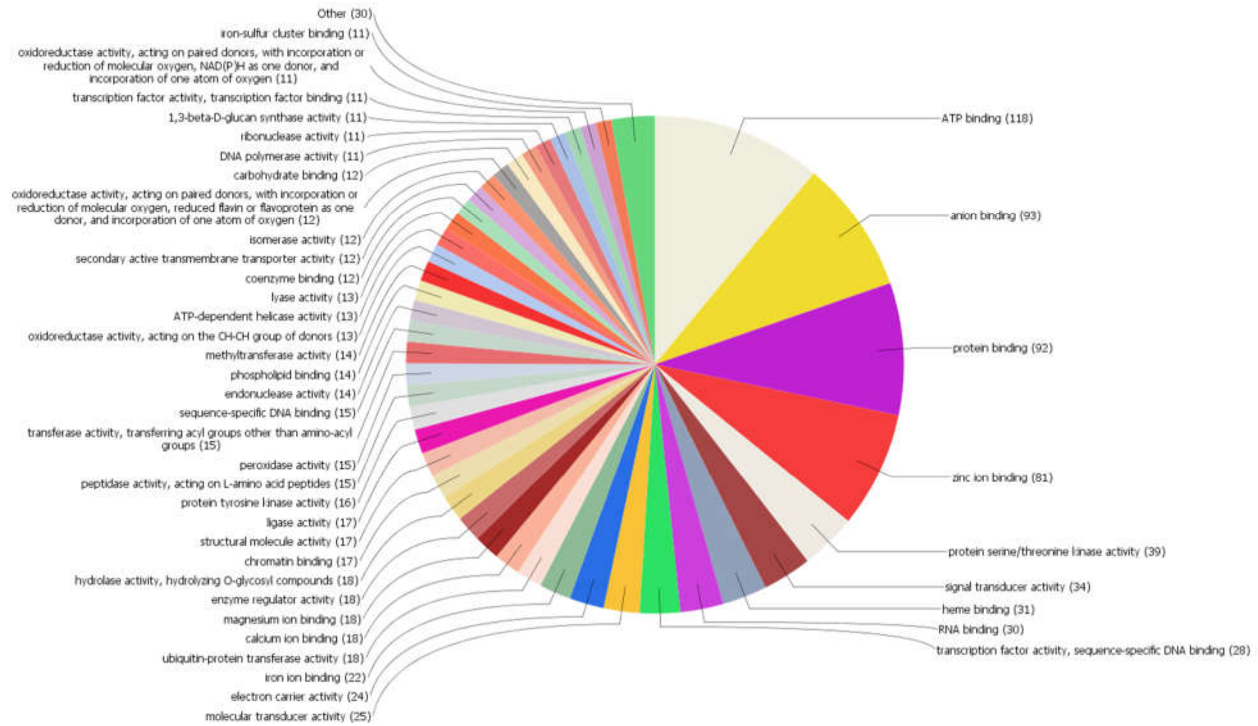

Figure 1.6: Sequence distribution in class Molecular Function': ATP binding' (118), 'anion binding' (93), 'protein binding' (92), and 'zinc-ion binding' (81) were the most common. Apart from these broadly general categories, 'signal transducer activity' (59), 'transcription factor activity' (58), DNA repair and chromatin maintenance (41), and 'protein serine/threonine kinase activity' (39) are the other highly represented categories.

Number of Sequences [Biological Process]

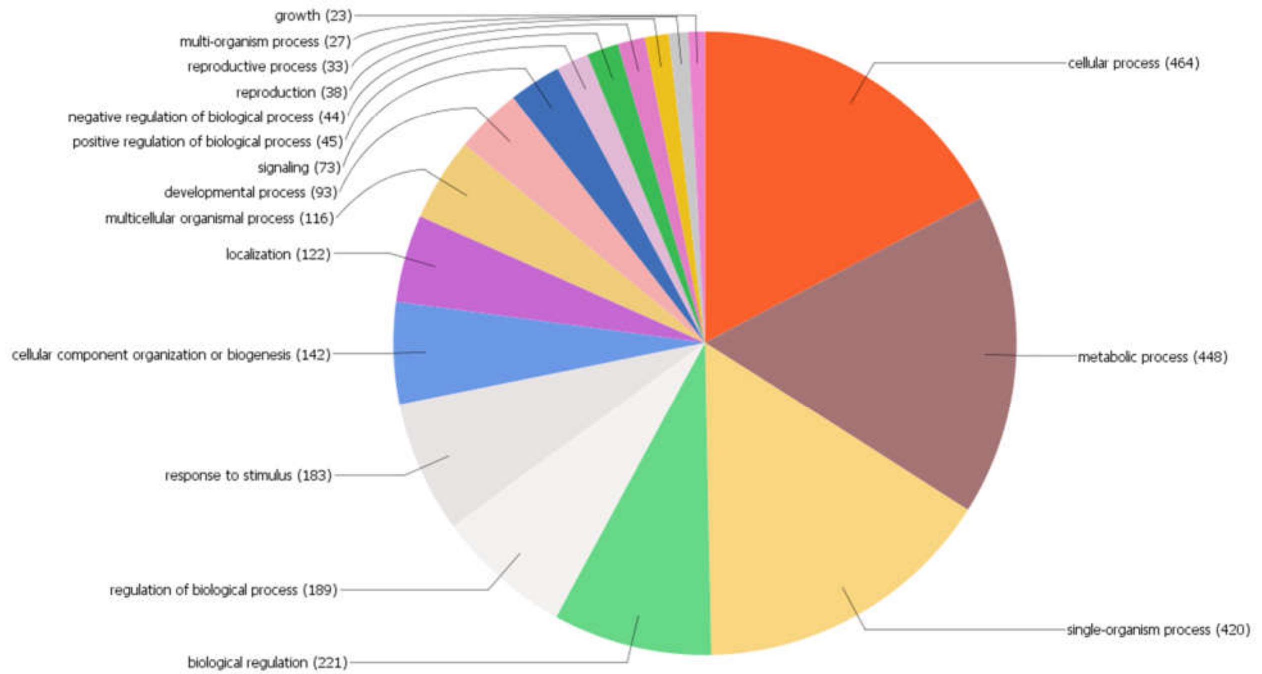

Figure 1.7: Sequence distribution in class 'Biological Process'. The categories of 'Regulation of biological process' (278), 'response to stimulus and signaling' (256), 'Growth and reproduction' (94), and 'developmental process' (93) are predominant.

Table 1.2: GO term enrichment analysis of DEGs. Enrichment analysis of 1187 DEGs compared to PANTHER and GO Consortium database of known Arabidopsis functional ontology resulted in an enrichment of 144 GO terms. Ontology (far right) refers to the classification of the GO term sorted according to Biological process (P), Molecular function (F) and Cellular component (C).

GO ID	Description	Number in Reference list	Number in Input List	P-value	Ontology
(GO:0030154)	cell differentiation	44	42	0.0000312	P
(GO:0000910)	cytokinesis	57	54	0.000000796	P
(GO:0009719)	response to endogenous stimulus	108	99	3.21E-12	P
(GO:0006778)	porphyrin-containing compound metabolic process	69	58	0.0000127	P
(GO:0048731)	system development	54	44	0.000936	P
(GO:0008643)	carbohydrate transport	142	112	8.06E-10	P
(GO:0006109)	regulation of carbohydrate metabolic process	45	35	0.0193	P
(GO:0009605)	response to external stimulus	55	42	0.00648	P
(GO:0006869)	lipid transport	58	44	0.00501	P
(GO:0006631)	fatty acid metabolic process	149	111	2.56E-08	P
(GO:0006644)	phospholipid metabolic process	128	95	0.000000606	P
(GO:0006732)	coenzyme metabolic process	140	103	0.000000215	P
(GO:0006766)	vitamin metabolic process	66	48	0.00627	P
(GO:0006897)	endocytosis	157	112	0.000000225	P
(GO:0009110)	vitamin biosynthetic process	62	44	0.0218	P
(GO:0006790)	sulfur compound metabolic process	235	162	7.57E-10	P
(GO:0019219)	regulation of nucleobase-containing compound metabolic process	201	137	6.46E-08	P
(GO:0006898)	receptor-mediated endocytosis	72	49	0.0249	P
(GO:0008203)	cholesterol metabolic process	140	95	0.0000343	P
(GO:0032501)	multicellular organismal process	108	73	0.000923	P
(GO:0044707)	single-multicellular organism process	108	73	0.000923	P
(GO:0009628)	response to abiotic stimulus	101	67	0.00382	P
(GO:0008104)	protein localization	216	143	0.000000171	P
(GO:0015931)	nucleobase-containing compound transport	111	73	0.00224	P
(GO:0005977)	glycogen metabolic process	197	127	0.00000597	P
(GO:0006520)	cellular amino acid metabolic process	661	418	7.45E-20	P
(GO:0019748)	secondary metabolic process	224	140	0.00000737	P
(GO:0016192)	vesicle-mediated transport	516	319	1.11E-13	P
(GO:0070271)	protein complex biogenesis	166	101	0.00148	P
(GO:0006461)	protein complex assembly	165	100	0.0019	P
(GO:0006887)	exocytosis	124	75	0.0231	P
(GO:0006820)	anion transport	182	108	0.00208	P
(GO:0009058)	biosynthetic process	1477	875	1.6E-33	P
(GO:0006796)	phosphate-containing compound metabolic process	1237	731	2.39E-27	P
(GO:0035556)	intracellular signal transduction	425	251	1.03E-08	P
(GO:0005976)	polysaccharide metabolic process	362	213	0.00000037	P
(GO:0005975)	carbohydrate metabolic process	757	445	9.27E-16	P

(Table contd.)

GO ID	Description	Number in Reference list	Number in Input List	P-value	Ontology
(GO:0007154)	cell communication	717	420	1.38E-14	P
(GO:0007067)	mitosis	303	176	0.0000196	P
(GO:0006996)	organelle organization	597	345	4.19E-11	P
(GO:0007165)	signal transduction	672	386	3.59E-12	P
(GO:0006807)	nitrogen compound metabolic process	1827	1047	3.84E-35	P
(GO:0006412)	translation	503	288	9.78E-09	P
(GO:0071840)	cellular component organization or biogenesis	1341	766	1.12E-24	P
(GO:0006605)	protein targeting	156	89	0.0404	P
(GO:0006091)	generation of precursor metabolites and energy	347	197	0.0000172	P
(GO:0006629)	lipid metabolic process	631	354	7.31E-10	P
(GO:0006950)	response to stress	718	402	3.37E-11	P
(GO:0050896)	response to stimulus	1280	716	9.99E-21	P
(GO:0009987)	cellular process	6108	3401	1.26E-123	P
(GO:0016043)	cellular component organization	1048	578	2.62E-15	P
(GO:0044085)	cellular component biogenesis	575	315	0.000000123	P
(GO:0006351)	transcription, DNA-dependent	793	433	1.29E-10	P
(GO:0008202)	steroid metabolic process	317	172	0.00134	P
(GO:0008652)	cellular amino acid biosynthetic process	354	192	0.000427	P
(GO:0032502)	developmental process	346	186	0.00101	P
(GO:0006810)	transport	1963	1054	2.06E-25	P
(GO:0022904)	respiratory electron transport chain	230	123	0.0415	P
(GO:0051179)	localization	2044	1091	1.58E-25	P
(GO:0007049)	cell cycle	786	415	2.58E-08	P
(GO:0008152)	metabolic process	7375	3891	1.56E-112	P
(GO:0006366)	transcription from RNA polymerase II promoter	654	343	0.00000192	P
(GO:0050789)	regulation of biological process	888	465	6.04E-09	P
(GO:0065007)	biological regulation	1210	633	1.82E-12	P
(GO:0044238)	primary metabolic process	6057	3164	7.48E-80	P
(GO:0006259)	DNA metabolic process	396	206	0.00238	P
(GO:0042592)	homeostatic process	334	173	0.0135	P
(GO:0009056)	catabolic process	1327	687	9.33E-13	P
(GO:0006139)	nucleobase-containing compound metabolic process	2544	1305	1.79E-24	P
(GO:0015031)	protein transport	1215	618	4.47E-10	P
(GO:0006886)	intracellular protein transport	1197	608	8.45E-10	P
(GO:0006357)	regulation of transcription from RNA polymerase II promoter	428	217	0.00648	P
(GO:0016070)	RNA metabolic process	1525	758	1.12E-10	P
(GO:0019538)	protein metabolic process	2562	1182	5.1E-09	P
(GO:0006464)	cellular protein modification process	1228	559	0.00425	P

(Table contd.)

GO ID	Description	Number in Reference list	Number in Input List	P-value	Ontology
(GO:0005249)	voltage-gated potassium channel activity	30	30	4.81E-04	F
(GO:0005244)	voltage-gated ion channel activity	36	35	1.46E-04	F
(GO:0019209)	kinase activator activity	39	37	1.29E-04	F
(GO:0016209)	antioxidant activity	56	52	1.88E-06	F
(GO:0004601)	peroxidase activity	51	47	1.12E-05	F
(GO:0005216)	ion channel activity	61	55	1.94E-06	F
(GO:0042626)	ATPase activity, coupled to transmembrane movement of substances	140	126	1.99E-15	F
(GO:0015144)	carbohydrate transmembrane transporter activity	113	92	7.47E-09	F
(GO:0008483)	transaminase activity	64	51	2.24E-04	F
(GO:0005484)	SNAP receptor activity	62	49	4.40E-04	F
(GO:0005516)	calmodulin binding	76	59	8.31E-05	F
(GO:0008324)	cation transmembrane transporter activity	261	199	8.10E-17	F
(GO:0015078)	hydrogen ion transmembrane transporter activity	92	70	1.67E-05	F
(GO:0015926)	glucosidase activity	72	54	6.34E-04	F
(GO:0016831)	carboxy-lyase activity	63	47	2.85E-03	F
(GO:0000981)	sequence-specific DNA binding RNA polymerase II transcription factor activity	96	70	7.51E-05	F
(GO:0005509)	calcium ion binding	129	94	1.03E-06	F
(GO:0016829)	lyase activity	295	207	7.32E-14	F
(GO:0003735)	structural constituent of ribosome	427	288	2.83E-17	F
(GO:0016407)	acetyltransferase activity	154	102	2.29E-05	F
(GO:0016746)	transferase activity, transferring acyl groups	279	183	9.01E-10	F
(GO:0016836)	hydro-lyase activity	130	83	1.13E-03	F
(GO:0016853)	isomerase activity	246	155	4.77E-07	F
(GO:0016462)	pyrophosphatase activity	478	300	5.08E-14	F
(GO:0016779)	nucleotidyltransferase activity	163	99	1.22E-03	F
(GO:0005198)	structural molecule activity	800	480	3.05E-19	F
(GO:0016301)	kinase activity	834	500	5.07E-20	F
(GO:0016740)	transferase activity	2167	1278	3.84E-51	F
(GO:0016491)	oxidoreductase activity	1273	745	1.39E-27	F
(GO:0005215)	transporter activity	1562	912	6.43E-34	F
(GO:0022857)	transmembrane transporter activity	1433	834	2.21E-30	F
(GO:0016757)	transferase activity, transferring glycosyl groups	391	221	2.11E-06	F
(GO:0004672)	protein kinase activity	586	327	3.50E-09	F

(Table contd.)

GO ID	Description	Number in Reference list	Number in Input List	P-value	Ontology
(GO:0000989)	transcription factor binding transcription factor activity	192	107	1.57E-02	F
(GO:0000988)	protein binding transcription factor activity	194	107	2.27E-02	F
(GO:0003700)	sequence-specific DNA binding transcription factor activity	594	325	3.22E-08	F
(GO:0005515)	protein binding	1286	703	9.80E-19	F
(GO:0003700)	Sequence-specific DNA binding transcription factor	594	325	3.22E-08	F
(GO:0003677)	DNA binding	847	453	2.31E-10	F
(GO:0005200)	structural constituent of cytoskeleton	339	180	1.58E-03	F
(GO:0003824)	catalytic activity	6645	3501	5.67E-100	F
(GO:0005488)	binding	3770	1900	3.91E-35	F
(GO:0016787)	hydrolase activity	2326	1159	2.09E-18	F
(GO:0016788)	hydrolase activity, acting on ester bonds	689	341	1.85E-04	F
(GO:0003676)	nucleic acid binding	2177	1062	3.07E-14	F
(GO:0003723)	RNA binding	935	427	1.45E-02	F
(GO:0004842)	ubiquitin-protein ligase activity	340	54	5.73E-12	F
(GO:0031201)	SNARE complex	65	52	0.000063	C
(GO:0005768)	endosome	49	39	0.00134	C
(GO:0005794)	Golgi apparatus	171	134	2.95E-12	C
(GO:0016021)	integral to membrane	393	306	1.02E-27	C
(GO:0005840)	ribosome	266	199	3.34E-16	C
(GO:0030312)	external encapsulating structure	97	68	0.000159	C
(GO:0005618)	cell wall	97	68	0.000159	C
(GO:0005829)	cytosol	546	377	1.28E-24	C
(GO:0030054)	cell junction	101	66	0.00179	C
(GO:0005874)	microtubule	77	49	0.0285	C
(GO:0005886)	plasma membrane	825	493	1.91E-19	C
(GO:0016020)	membrane	1261	741	1.18E-27	C
(GO:0005773)	vacuole	152	89	0.00468	C
(GO:0005783)	endoplasmic reticulum	148	86	0.00767	C
(GO:0043226)	organelle	2646	1524	4.04E-55	C
(GO:0005737)	cytoplasm	2440	1366	9.15E-43	C
(GO:0044464)	cell part	4461	2469	3.7E-80	C
(GO:0005622)	intracellular	4248	2318	4.1E-69	C
(GO:0005654)	nucleoplasm	206	110	0.0236	C
(GO:0030529)	ribonucleoprotein complex	538	287	0.00000188	C
(GO:0005634)	nucleus	1064	549	2.5E-10	C
(GO:0032991)	macromolecular complex	1574	761	3.89E-09	C
(GO:0043234)	protein complex	1110	489	0.0475	C

Table 1.3: Pathway enrichment of the DEGs based on KEGG pathway database

Group	Pathway	Identifier
Metabolism	Fatty acid elongation	ko00062
	Phenylpropanoid biosynthesis	ko00940
	Pyrimidine metabolism	ko00240
	N-Glycan biosynthesis	ko00510, ko00513
	Purine metabolism	ko00230
	Oxidative phosphorylation	ko00190
	Diterpenoid biosynthesis	ko00904
	Sphingolipid metabolism	ko00600
	Terpenoid backbone biosynthesis	ko00900
	Sesquiterpenoid and triterpenoid biosynthesis	ko00909
	Glycine, serine and threonine metabolism	ko00260
	Glyoxylate and dicarboxylate metabolism	ko00630
	Carbon metabolism	ko01200
	Glutathione metabolism	ko00480
	Metabolism of xenobiotics by cytochrome P450	ko00980
	Drug metabolism - cytochrome P450	ko00982
	Platinum drug resistance	ko01524
	Chemical carcinogenesis	ko05204
	Glycerolipid metabolism	ko00561
Genetic information processing	RNA degradation	ko03018
	RNA transport	ko03013
	mRNA surveillance pathway	ko03015
	Base excision repair	ko03410
	Nucleotide excision repair	ko03420 ko00513
	Basal transcription factors	ko03022
	DNA replication	ko03030
	Mismatch repair	ko03430
	Homologous recombination	ko03440
Signaling	FoxO signaling pathway	ko04068
	Sphingolipid signaling pathway	ko04071
	MAPK signaling pathway - plant	ko04016
	mTOR signaling pathway	ko04150
	Plant hormone signal transduction	ko04075
	MAPK signaling pathway - plant	ko04016
	Plant hormone signal transduction	ko04075

Table 1.4: List showing 58 differentially expressed transcription factors represented among the 1187 DEGs.

Sequence name	Sequence description
evm.TU.supercontig_142.60_4-412	bZIP transcription factor family
evm.TU.supercontig_1675.1_1-506	calmodulin-binding transcription activator 2-like
evm.TU.supercontig_213.9_4-412	calmodulin-binding transcription activator 3 isoform X1
evm.TU.supercontig_914.1_1-3956	CCR4-NOT transcription complex subunit 1-like
evm.TU.supercontig_104.52_2-285	E2F transcription factor 3
evm.TU.supercontig_89.35_3-210	ethylene-responsive transcription factor 2-like
evm.TU.supercontig_50.28_2-306	ethylene-responsive transcription factor ERF095-like
evm.TU.supercontig_38.72_4-212	Ethylene-responsive transcription factor ERF114
evm.TU.supercontig_21.243_0-531	General transcription factor 3C polypeptide 3
evm.TU.supercontig_52.111_1-639	heat stress transcription factor A-6b
evm.TU.supercontig_27.45_4-702	heat stress transcription factor B-2a-like
evm.TU.supercontig_224.11_2-803	heat stress transcription factor B-3
evm.TU.supercontig_3486.1_0-181	MADS-box transcription factor 6
evm.TU.supercontig_119.43_0-629	mediator of RNA polymerase II transcription subunit 13 isoform X1
evm.TU.supercontig_1346.1_1-208	mediator of RNA polymerase II transcription subunit 14
evm.TU.supercontig_2756.3_0-1270	mediator of RNA polymerase II transcription subunit 14
evm.TU.supercontig_25.148_0-355	mediator of RNA polymerase II transcription subunit 15a isoform X2
evm.TU.supercontig_75.48_0-299	mediator of RNA polymerase II transcription subunit 17
evm.TU.supercontig_2.207_16-1698	mediator of RNA polymerase II transcription subunit 23 isoform X1
evm.TU.supercontig_12.295_25-1738	myb A
evm.TU.supercontig_123.16_11-225	myb X isoform X1
evm.TU.supercontig_798.1_138-946	myb-related 305-like
evm.TU.supercontig_138.21_2-653	myb-related 308-like
evm.TU.supercontig_566.3_22-268	NAC domain-containing 35-like
evm.TU.supercontig_2586.1_3-972	NAC domain-containing 73-like
evm.TU.supercontig_6.222_4-437	NAC domain-containing 7-like isoform X1
evm.TU.supercontig_106.75_56-629	NAC domain-containing 83-like
evm.TU.supercontig_435.1_140-495	NAC transcription factor 25-like
evm.TU.supercontig_2794.2_124-349	probable transcription factor KAN4 isoform X3
evm.TU.supercontig_11.67_0-461	probable WRKY transcription factor 43
evm.TU.supercontig_169.18_5-367	probable WRKY transcription factor 49
evm.TU.supercontig_86.65_0-260	probable WRKY transcription factor 70 isoform X2
evm.TU.supercontig_1195.3_6-908	probable WRKY transcription factor 71
evm.TU.supercontig_43.76_3-501	probable WRKY transcription factor 75
evm.TU.supercontig_2.26_1-1110	probable WRKY transcription factor 9
evm.TU.supercontig_73.36_0-1450	TATA-binding -associated factor BTAF1 isoform X1
evm.TU.supercontig_1.81_0-440	T-box transcription factor isoform 1
evm.TU.supercontig_4.168_12-661	transcription factor bHLH118-like
evm.TU.supercontig_138.9_57-346	transcription factor DIVARICATA-like
evm.TU.supercontig_2174.1_1-376	transcription factor jumonji domain
evm.TU.supercontig_3.92_0-1367	Transcription factor jumonji domain-containing isoform 2
evm.TU.supercontig_16.21_0-368	transcription factor MUTE

(Table contd.)

Sequence name	Sequence description
evm.TU.supercontig_7.125_100-541	transcription factor MYB44
evm.TU.supercontig_642.1_0-239	transcription factor PAR1-like
evm.TU.supercontig_51.95_40-163	transcription factor TFIID
evm.TU.supercontig_3.56_20-250	transcription factor TRY
evm.TU.supercontig_18.75_34-378	transcription factor UPBEAT1
evm.TU.supercontig_471.1_2-1870	transcription initiation factor TFIID subunit 1 isoform X1
evm.TU.supercontig_1020.1_70-999	transcription initiation factor TFIID subunit 2 isoform X2
evm.TU.supercontig_705.1_96-717	transcription initiation factor TFIID subunit 7-like
evm.TU.supercontig_104.56_2-214	transcription repressor OFP6-like
evm.TU.supercontig_50.48_1-521	transcriptional regulator SUPERMAN-like
evm.TU.supercontig_6.243_86-922	transformation transcription domain-associated
evm.TU.supercontig_6.244_50-3589	transformation transcription domain-associated -like
evm.TU.supercontig_6.244_3743-4984	transformation transcription domain-associated -like
evm.TU.supercontig_21.93_0-310	transforming growth factor-beta receptor-associated 1
evm.TU.supercontig_51.22_1-2363	translational activator GCN1
evm.TU.supercontig_1.158_1-722	truncated transcription factor CAULIFLOWER A-like

Table 1.5: List showing the 22 differentially expressed genes encoding Hormone related genes (HRGs) from the 1187 DEGs

Sequence name	Sequence description
evm.TU.supercontig_1525.3_416-567	abscisic acid 8 -hydroxylase 1
evm.TU.supercontig_99.39_148-876	abscisic acid 8 -hydroxylase 2
evm.TU.supercontig_21.181_13-855	abscisic acid 8 -hydroxylase 3-like
evm.TU.supercontig_6.143_1-1868	auxilin 1
evm.TU.supercontig_14.238_0-2790	auxin transport BIG
evm.TU.supercontig_3603.1_1-353	auxin transport BIG
evm.TU.supercontig_37.207_1-279	auxin-induced 15A-like
evm.TU.supercontig_6.232_2-304	auxin-induced 15A-like
evm.TU.supercontig_37.55_7-344	auxin-induced 6B-like
evm.TU.supercontig_37.208_50-237	auxin-induced X15-like
evm.TU.supercontig_144.5_1-285	auxin-responsive SAUR32-like
evm.TU.supercontig_99.56_2-367	auxin-responsive SAUR32-like
evm.TU.supercontig_20.149_40-357	auxin-responsive SAUR68-like
evm.TU.supercontig_279.6_2-247	cytokinin riboside 5 -monophosphate phosphoribohydrolase LOG1
evm.TU.supercontig_320.1_6-645	cytokinin riboside 5 -monophosphate phosphoribohydrolase LOG5
evm.TU.supercontig_823.1_10-1135	gibberellin 20 oxidase 1-B-like
evm.TU.supercontig_111.7_0-213	gibberellin-regulated 4-like
evm.TU.supercontig_20.146_4-413	indole-3-acetic acid-induced ARG7-like
evm.TU.supercontig_99.41_1-946	jasmonate O-methyltransferase-like
evm.TU.supercontig_65.135_365-481	probable auxin efflux carrier component 1c
evm.TU.supercontig_6.73_4-1799	probable indole-3-acetic acid-amido synthetase
evm.TU.supercontig_19.55_31-548	two-component response regulator ARR17

Table 1.6: List showing the 35 differentially expressed genes coding for chromatin related proteins among the 1187 DEGs

Sequence name	Sequence description
evm.TU.supercontig_88.2_0-244	chromatin assembly factor 1 subunit FAS1-like
evm.TU.supercontig_3234.2_1-223	chromatin assembly factor 1 subunit FAS2
evm.TU.supercontig_51.12_14-657	CHROMATIN REMODELING 20
evm.TU.supercontig_1026.2_0-1191	CHROMATIN REMODELING 35-like isoform X1
evm.TU.supercontig_84.19_22-979	CHROMATIN REMODELING 5
evm.TU.supercontig_146.9_44-234	chromatin structure-remodeling complex BSH
evm.TU.supercontig_16.101_313-1507	chromatin structure-remodeling complex SYD isoform X1
evm.TU.supercontig_16.95_1-3015	chromatin structure-remodeling complex SYD isoform X1
evm.TU.supercontig_6080.1_0-1723	chromatin structure-remodeling complex SYD isoform X1
evm.TU.supercontig_59.113_92-841	chromosome transmission fidelity 18 homolog
evm.TU.supercontig_2632.1_0-1243	DNA annealing helicase and endonuclease ZRANB3
evm.TU.supercontig_1221.1_1-429	DNA gyrase subunit chloroplastic mitochondrial isoform X1
evm.TU.supercontig_170.6_1-204	DNA ligase 1
evm.TU.supercontig_50.138_0-524	DNA mismatch repair MSH6
evm.TU.supercontig_5.338_3-783	DNA polymerase beta isoform X1
evm.TU.supercontig_32.112_1-1624	DNA polymerase epsilon catalytic subunit A-like
evm.TU.supercontig_30.92_4-788	DNA polymerase eta isoform X2
evm.TU.supercontig_30.95_10-429	DNA polymerase eta isoform X2
evm.TU.supercontig_48.64_4-208	DNA polymerase zeta catalytic subunit isoform X1
evm.TU.supercontig_2894.1_0-154	DNA polymerase zeta processivity subunit
evm.TU.supercontig_2949.1_258-417	DNA repair
evm.TU.supercontig_140.31_3-995	DNA repair UVH3 isoform X1
evm.TU.supercontig_10.19_0-1869	DNA replication ATP-dependent helicase nuclease DNA2 isoform X2
evm.TU.supercontig_1735.1_5-187	DNA topoisomerase 4 subunit B
evm.TU.supercontig_74.63_59-471	DNA topoisomerase 4 subunit B (DUF810)
evm.TU.supercontig_14.252_4-299	DNA-directed RNA polymerase chloroplastic mitochondrial
evm.TU.supercontig_92.7_5-1267	DNA-directed RNA polymerase I subunit 1
evm.TU.supercontig_17.33_4-856	DNA-directed RNA polymerase III subunit RPC2
evm.TU.supercontig_271.2_315-2639	DNA-directed RNA polymerase V subunit 1
evm.TU.supercontig_271.4_0-226	DNA-directed RNA polymerase V subunit 1
evm.TU.supercontig_3599.1_0-211	DNAJ heat shock N-terminal domain-containing
evm.TU.supercontig_52.147_691-834	DNAJ heat shock N-terminal domain-containing isoform 2
evm.TU.supercontig_23.130_11-586	dnaJ homolog subfamily C GRV2-like isoform X1
evm.TU.supercontig_23.134_1-3229	dnaJ homolog subfamily C GRV2-like isoform X1
evm.TU.supercontig_953.3_3-404	dnaJ P58IPK homolog

Table 1.7: List showing the 21 differentially expressed genes coding transporter proteins (TP) among the 1187 DEGs

Sequence name	Sequence description
evm.TU.supercontig_224.5_0-801	ABC transporter A family member 1 isoform X1
evm.TU.supercontig_224.7_6-343	ABC transporter A family member 1 isoform X1
evm.TU.supercontig_3.35_5657-6000	ABC transporter C family member 10-like
evm.TU.supercontig_10.181_287-1945	boron transporter 4-like
evm.TU.supercontig_2.233_22-909	calcium uniporter mitochondrial-like
evm.TU.supercontig_79.62_1-2658	intracellular transport USO1
evm.TU.supercontig_850.1_14-1926	intracellular transport USO1-like
evm.TU.supercontig_52.19_1-2086	oligopeptide transporter 1-like
evm.TU.supercontig_80.121_21-1613	organic cation carnitine transporter 4
evm.TU.supercontig_1552.1_0-1772	phospholipid-transporting ATPase 1-like
evm.TU.supercontig_1088.2_1-231	sugar transporter ERD6-like 3
evm.TU.supercontig_1088.3_1-749	sugar transporter ERD6-like 3
evm.TU.supercontig_29.159_0-1571	transport SEC16B homolog
evm.TU.supercontig_1109.2_12-1117	UDP-galactose UDP-glucose transporter 2-like
evm.TU.supercontig_21.183_1-679	vacuolar iron transporter homolog 4-like
evm.TU.supercontig_34.191_279-949	tyrosine-specific transport isoform X3
evm.TU.supercontig_122.41_10-244	probable GABA transporter 2
evm.TU.supercontig_170.63_1-722	probable manganese-transporting ATPase PDR2
evm.TU.supercontig_1208.1_564-1340	probable polyol transporter 6
evm.TU.supercontig_53.38_701-1649	probable potassium transporter 17
evm.TU.supercontig_5.297_2-163	phospholipid-transporting ATPase 3

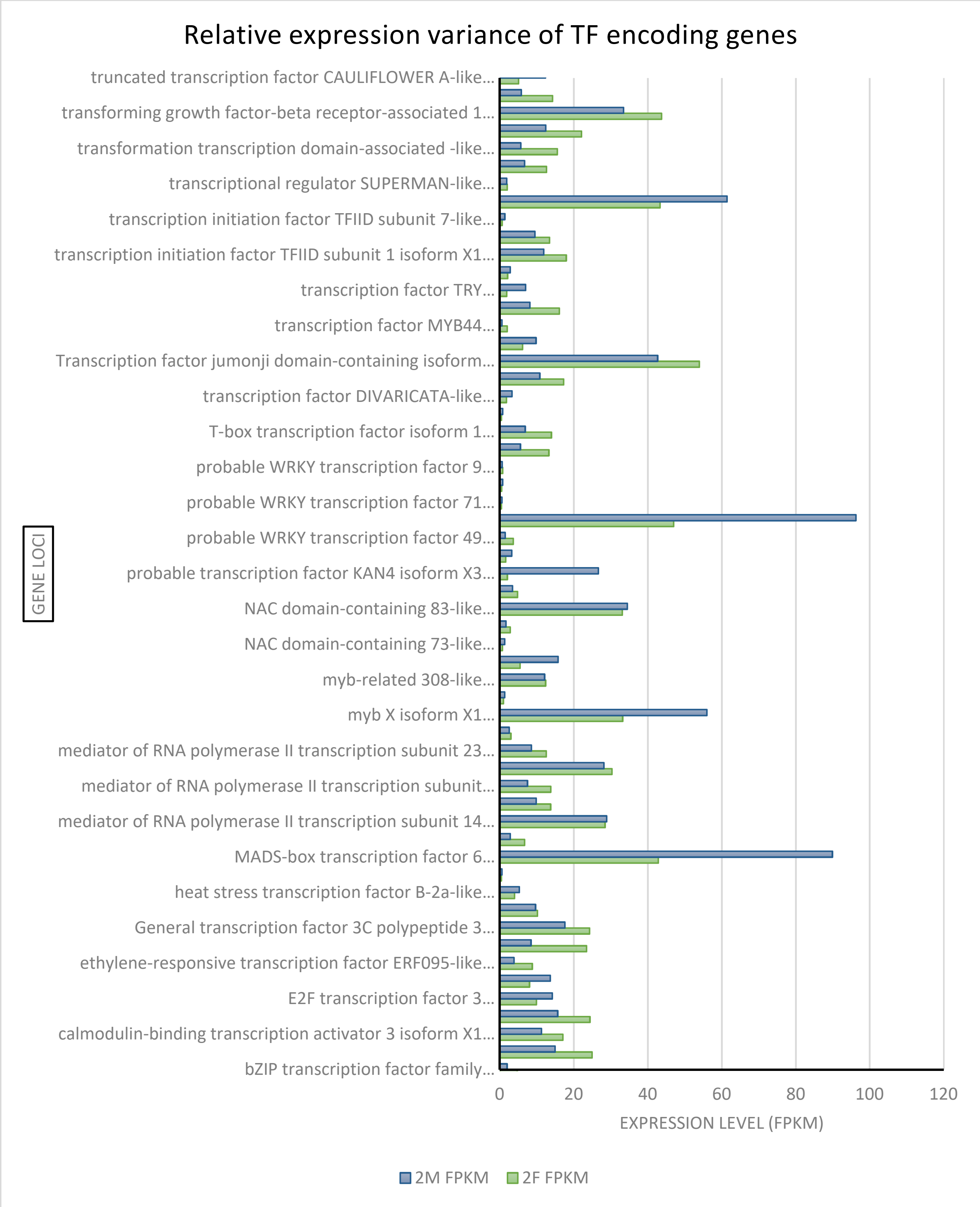

Figure 1.8: Figure showing relative expression among the TF coding genes between Female and Male SAM samples at 2 weeks before flowering. Blue bars represent male expression and green bars represent female expression.

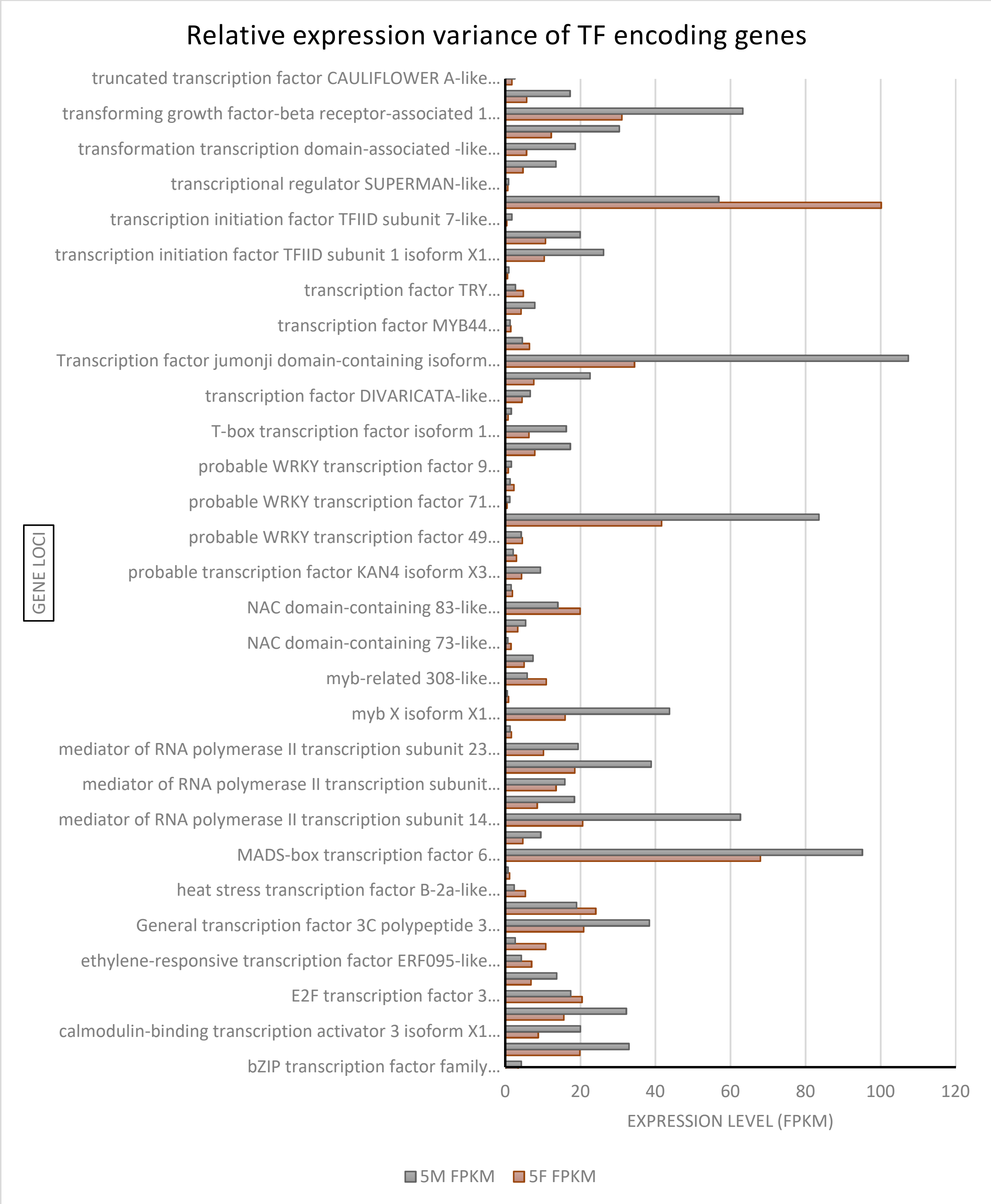

Figure 1.9: Relative expression among TF coding genes between male and female SAM samples at the 5 week before flowering stage. Brown bars represent female and gray bars represent male samples.

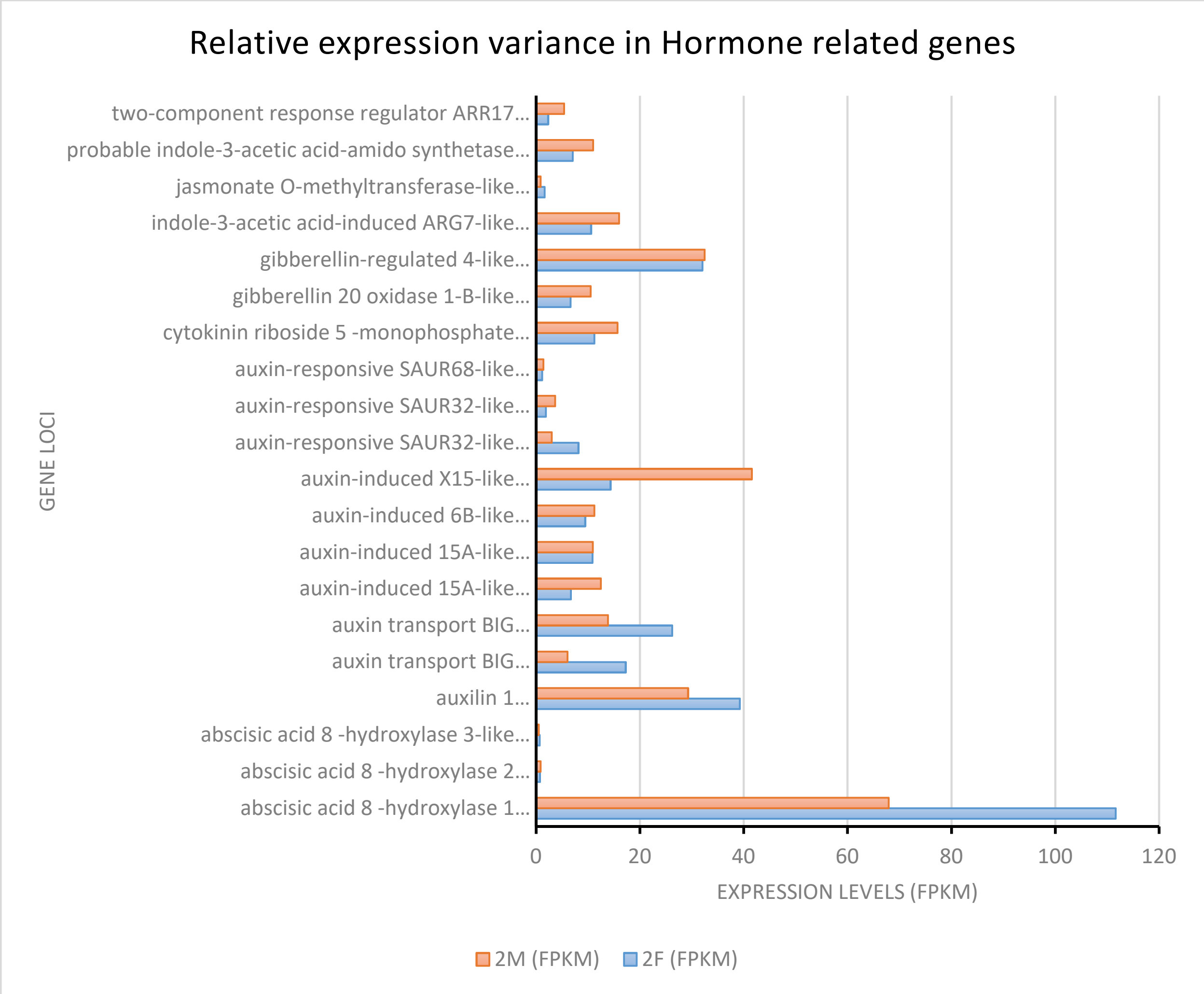

Figure 1.10: Relative expression variance among hormone related genes between male and female SAM samples at the 2 week before flowering stage. Red bars represent male samples and blue bars represent female samples.

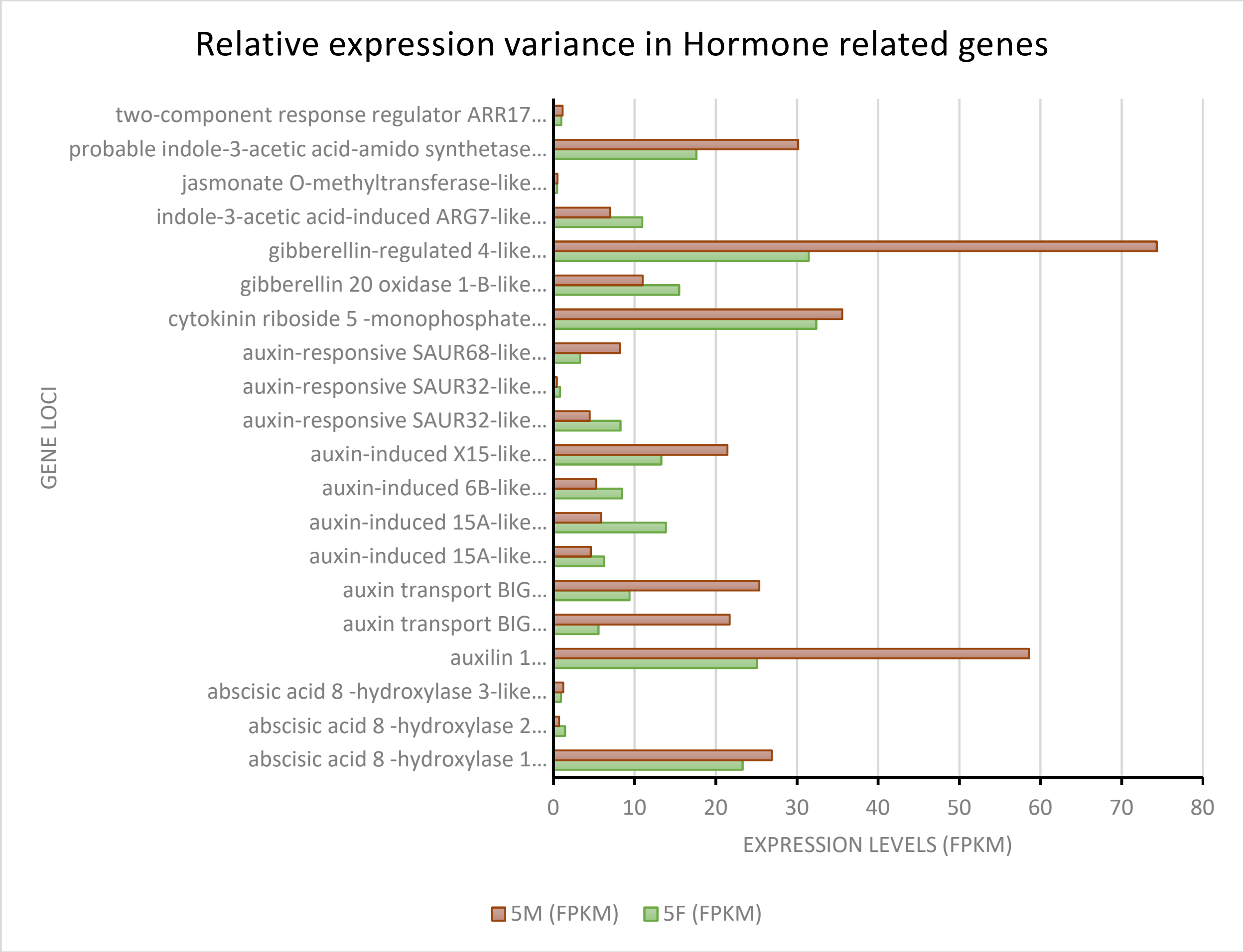

Figure 1.11: Relative expression variance among the hormone related genes between male and female SAM samples at the 5 week before flowering stage. Brown bars represent male samples and green bars represent female samples.

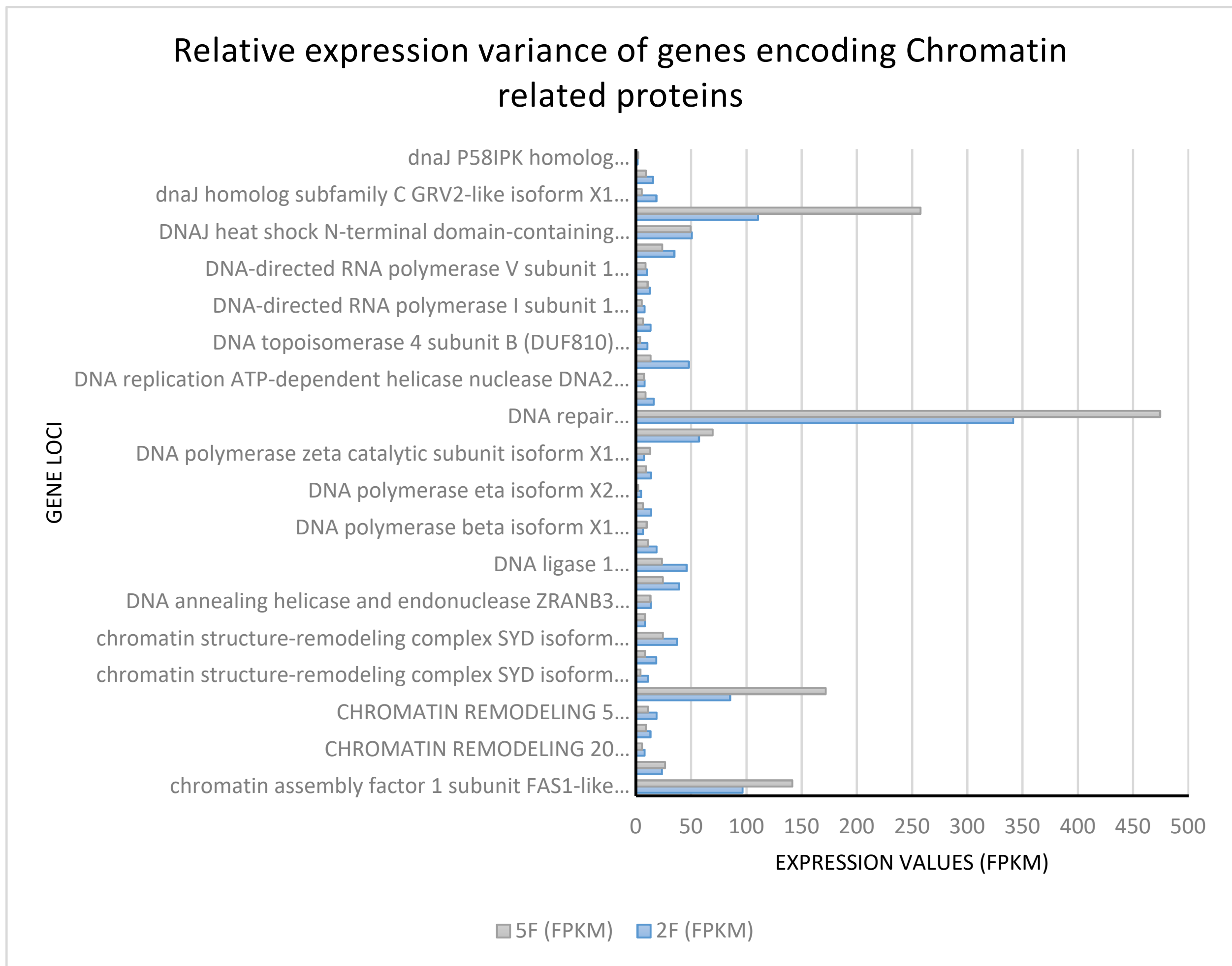

Figure 1.12: Expression variance among the genes encoding chromatin maintenance related proteins between female SAM samples at 2 and 5 weeks before flowering stages. Blue bars represent 2 week samples and grey bars represent 5 week samples.

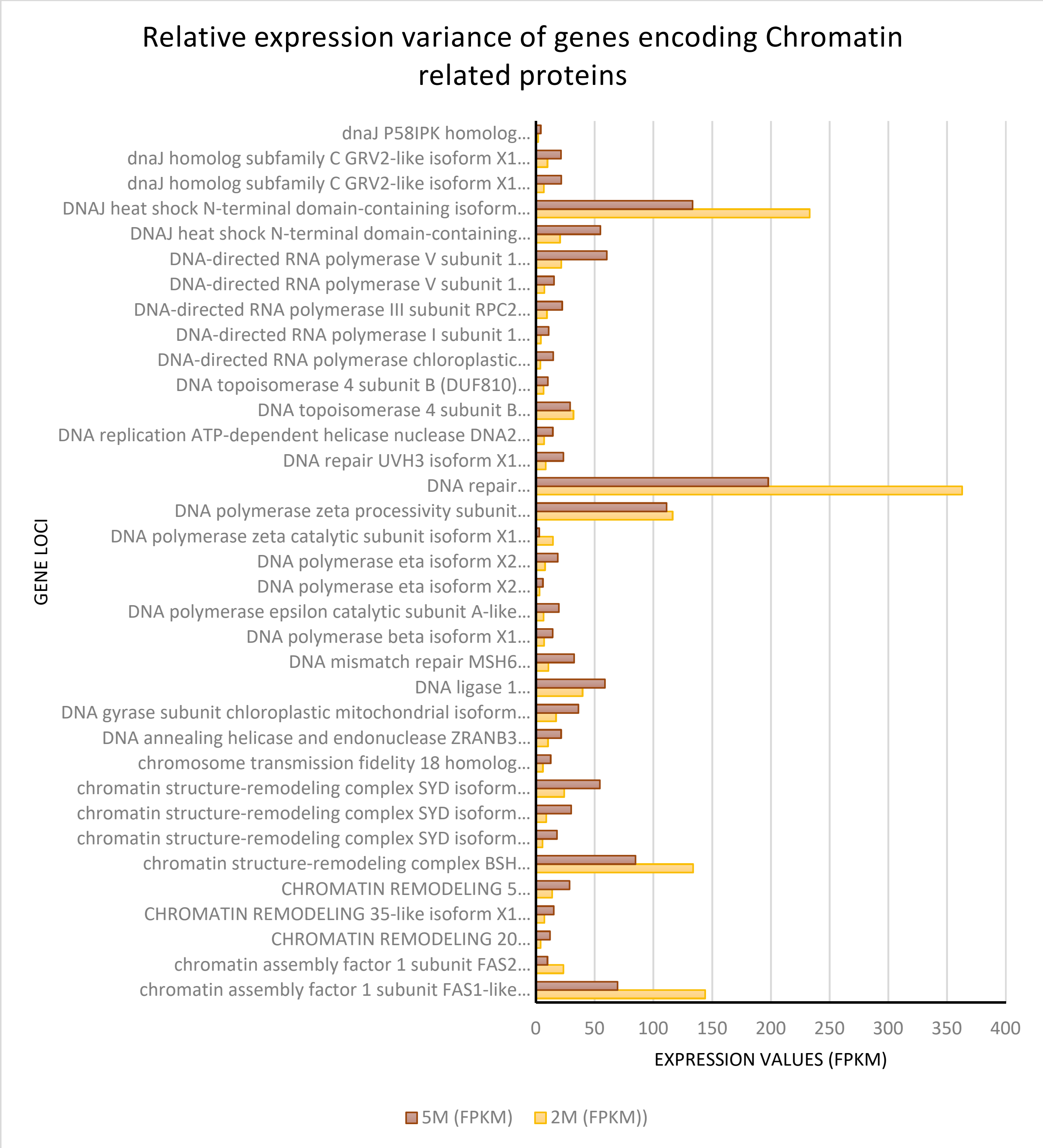

Figure 1.13: Expression variance among genes encoding chromatin maintenance related proteins within male SAM samples at the 2 and 5 week before flowering stage. Yellow bars represent 2 week samples and maroon bars represent 5 week samples.

Chapter 3: Generation and analysis of female sterile mutants in Arabidopsis

Abstract

Female sterile mutants in Arabidopsis are underrepresented compared to male sterile mutants in mutant databases. Most available mutants in this category display female reproductive defects in conjunction with defects in other reproductive organs as well as non-reproductive pleiotropic defects. Most 'female sterile only' mutants have perturbations in pathways regulating ovule development, pollen recognition, and transmitting tract development. EMS was tested for half lethal dose to optimize the frequency to generate female sterile mutants. After screening 2386 Arabidopsis M2 mutant lines, 3 individual mutant lines, FS21, FS23, and FS322, were found defective in the female reproductive organs leading to female sterility. FS21 showed highly reduced stigmatic papillae and lack of fertilization. FS23 had similar morphology as shown in FS21 and displayed random abortion of siliques. FS322 displayed deformed stigma in the shape of an upturned saucer. Identifying the causal genes involved will help to solve one of the most challenging and interesting biological problems in reproductive biology in land plants and help us in improving our understanding of the evolution of sex determination and sex chromosomes in land plants.

Introduction

Arabidopsis thaliana is used as a model organism for studies in various aspects of plant biology research owing to the extensive database of genes and their mutant phenotypes as well as the small size of the sequenced genome of 125Mb from the Columbia (Col-0) accession (The

Arabidopsis Genome Initiative 2000; (Somerville & Koornneef 2002). Arabidopsis mutants cover all aspects of plant development, physiology and metabolism and are excellent starting points for fundamental research. However, previously documented mutant phenotypes for reproductive processes extensively cover mainly the 'male sterile' aspect of development. The 'female sterile' aspect is under-represented in Seed Genes Project, the manually curated comprehensive collection of 2400 genes with loss-of-function mutant phenotypes (Lloyd & Meinke 2012), and also in the recently constructed RIKEN database RARGE-II, which is an integrated database of mutant phenotypes in Arabidopsis (Akiyama, 2014). These databases returned a list of 152 reproductive mutants, among which only 14 were in the 'abnormal flowers' and 'abnormal gynoecium' mutant categories. Most of the other mutant phenotypes occur in conjunction with male reproductive organ (androecium) defects and embryo defects, and a host of other pleiotropic effects of the mutation. In addition to these, most female sterile mutants described are in pathways regulating ovule development, pollen recognition by stigma, and transmitting tract development. Our efforts are to study the developmental processes affecting proper initiation and development of the gynoecia (comprising 2 fused carpels), through generation and study of mutants defective only in the development of female organs (gynoecium) and gametophyte while other floral organs are unaffected.

The ABCE model

The classic ABC model of plant floral development was a major milestone in understanding the genetics of flower development. This model postulated how three regulatory gene functions A, B and C work in a combinatorial fashion to give organ identity in a whorl-specific manner (Coen & Meyerowitz 1991). With the exception of the A function gene APETALA2, which

is a founding member of the AP2/ETHYLENE RESPONSE FACTOR transcription factor family (Zik & Irish 2003), all of the other Arabidopsis 'ABC' genes code for MIKCc-type MADS-domain transcription factors (Krizek & Fletcher 2005; Irish 2006). The ABC model was expanded in recent times with the addition of the class E floral-homeotic genes which also encode MADS-domain transcription and exhibit flower specific expression. The partially redundant class E genes, the SEPALLATA genes function redundantly with the class A, B and C genes to confer sepal, petal, stamen and carpel identity (Pelaz et al. 2000; Zahn et al. 2005). The modified ABCE model now takes precedence in providing a basis for the development of the flower (Zahn et al. 2006) through intricate higher-order and highly redundant genetic functions via interactions with a host of other gene products and cofactors, such as UNUSUAL FLORAL ORGANS (UFO), LEAFY (LFY), WUSCHEL (WUS), LEUNIG (LUG), CRABS CLAW (CRC), SEUSS (SEU), SUPERMAN (SUP), KNUCKLES (KNU) and miR172 (reviewed in Krizek & Fletcher 2005) to name a few, in a spatial-temporal manner, starting from the signals and events required for the initiation of floral primordia, to the separation of the floral organs into 'whorls' and including the continued presence of these gene products in the later stages of development of the floral organs. In addition to these well-known genes that act both upstream and downstream of the ABCE model, the databases mentioned also reveal a host of other genes, both with known functions and unknown, that are implicated in the development process of the female organs. Some of these genes are YUCCA1, YUCCA4, PIN, PINOID (working in the auxin hormone biosynthesis and response pathways)(Youfa Cheng et al. 2006; Cheng et al. 2007; Xing et al. 2013; Bennett et al. 1995; Benjamins et al. 2001; Furutani et al. 2004; Christensen et al. 2000; Lampugnani et al. 2013b), ACC synthase (working in the ethylene biosynthesis pathway)(De Martinis & Mariani 1999; Boualem et al. 2008; Boualem et al.

2009; Sherif et al. 2009), COP1-interacting protein-related (Wei & Deng 2003; Stewart et al. 2016), while some of the genes with unknown functions include ribosomal proteins (L39), Leucine-rich-repeat transmembrane proteins, zinc finger proteins, AP2/B3-like transcription factor family protein (Krogan et al. 2012; Zik & Irish 2003).

The class C gene AGAMOUS (AG) that specifies carpel identity and confers floral determinacy, also bears cadastral functions. The expression domain of AG itself is maintained by AP2 and a co-repressor complex of LUG and SEU (ÓMaoiléidigh et al. 2013; O'Maoileidigh et al. 2014). Other genes, such as the HUA and HEN (HUA ENHANCER) functioning in RNA metabolism, have been found to regulate AG at the post-transcriptional level (Chen & Meyerowitz 1999), (Cheng & Chen 2004). Studies on the downstream targets of AG have found most of the genes known to be involved in carpel and stamen development (Gómez-Mena et al. 2005). In addition to this, a microarray-based global analysis of comparative expression profiles of the floral-homeotic mutants identified 260 putative carpel-specific genes, most of which lack the CArG box required for MADS protein binding suggesting that these floral genes are potentially indirectly regulated by the floral homeotic gene products (Wellmer et al. 2004). The relatively small number of canonical genes in this regulatory network also indicates that this regulation might be mostly carried out by regulatory elements that have not yet been correlated with the genetics of the organogenesis process, both direct and indirect, to fine tune the proper anatomical structures necessary for successful reproduction.

Unisexuality in the plant kingdom has evolved independently multiple times, and in different ways in different lineages with different mechanisms, from a hermaphrodite ancestor over the course of millions of years of evolution. The ABCE model of floral development elegantly

explains the basic underlying developmental framework of the individual whorls across species. However, it stops short of universally explaining the occurrence of selective organ abortion, as in the case of unisexual flowers. This is possibly indicative of the presence of a more refined Gene Regulatory Network (GRN) controlling organ development after specification of the organ primordia, and acting upstream and downstream of ABCE activity (Ainsworth et al. 1997; Wellmer et al. 2004; Stilio et al. 2011). The first step towards sexual dimorphism is the abortion of either one of the sexual organs, by either a gain of function mutation, or a loss of function mutation, leading to the development of unisexual flowers (Ming et al. 2011). This can be affected by mutations in the immense number of genes controlling the large number of specialized genetics functions required at various developmental stages as well as mutations in the many regulatory genes (Wellmer et al. 2004; Zhang et al. 2005).

Male and female sterile mutations however, have different rates and probabilities of being fixed in a population owing to the abundant and mobile nature of pollen and sedentary recipient nature of gynoecia. The relative nature of the male and female sterile mutations (dominant or recessive to each other) can give rise to a variety of scenarios leading to different outcomes depending on their locations on the chromosomes. Stable dioecy and sex chromosomes only arises when the sex determining genes are closely linked on the same chromosome and exhibit dominance over each other (Charlesworth & Charlesworth 1978), (Charlesworth 1996) followed by suppression of recombination between the two genes (Ming et al. 2011). In other words, the two mutations have to be dominant over recessive or null alleles, and closely linked on the same chromosome.

Hormones in floral development

Plant growth hormones are also directly and indirectly involved in sex expression with various underlying genetic mechanisms. In recent studies, ethylene in the determination of sex in *Cucumis melon* (Boualem et al. 2008), (Boualem et al. 2009), (Martin et al. 2009) and brassinosteroid in maize (Hartwig et al. 2011) has been found to affect the development of specific reproductive organs. However, being effected by growth hormones which are involved in a host of other biological functions also involves pleiotropic effects in the plants. Even so, hormonal interactions can drive a response pathway to induce a certain amount of plasticity in sex determination which can ultimately be modulated by environmental factors. In Arabidopsis, the growth hormones auxin and cytokinin play important roles in the correct specification and patterning of the gynoecium structure and suppression of stamen development. Auxin biosynthesis (Y Cheng et al. 2006), auxin transport (Cheng et al. 2008) as well as auxin response (Ellis et al. 2005) mutants have been observed with floral defects. YUCCA and PIN mutants further cement the role of auxin and its transport in this context. Cytokinin also regulates floral meristem size through regulation of the meristem maintenance gene CLV1 and WUS (Lindsay et al. 2006). GA is involved in promoting stamen and anther development (Cheng et al. 2004) and it works through regulation of the floral meristem control gene LFY. Jasmonate signaling is required for stamen and pollen maturation (Park et al. 2002), (Ito et al. 2007). Brassinosteriod mutants show a dwarfed phenotype and reduced male fertility (Ye et al. 2010). Ethylene, also regarded as a feminizing hormone, plays roles in floral development, fruit ripening and senescence and has been found to be expressed in the stigma, style and the ovary but not in the pollen and anther (De Martinis & Mariani 1999); however, no specific hormone can be identified as a feminizing or

masculinizing agent only given that the hormones are important parts of the stress stimuli response mechanisms so very important for the sedentary lifestyle of plants (Golenberg & West 2013).

In the plant kingdom, hermaphroditism is the norm, with only about 6% of flowering plants (angiosperms) being dioecious (15,600 dioecious angiosperms in 987 genera and 175 families) and 5% monoecious. The distribution of dioecious species is also patchy and uneven (Renner 2014). The widely accepted view is that bisexual flowers are an ancestral, or early trait, from which, unisexuality evolves by the process of random mutations affecting the carpel or stamen abortion (Ming et al. 2011). Taken together, unisexuality evolved multiple times, and with a variety of mechanisms. However, the underlying genes and gene networks of these processes have a potential of having some extent of commonality by the way of having tissue specific roles that are selectively activated or de-activated. Theoretically, it should therefore be possible to generate unisexual flowers in the laboratory by disrupting gene functions, thereby uncoupling the developmental processes governing male and female organs. The stage of developmental arrest, in this case, should also identify early and late acting components of these processes.

Exploring the organ specific developmental gene network will therefore give an insight into the separation of reproductive functionality in flowering plants and help us contribute to the growing knowledge about the genetics of the reproductive developmental pathways in plants. We hope to be able to eventually dissect and identify key players in the sex-determination gene networks in land plants.

Mutagenesis

Female sterile mutants are less common, relative to male sterile mutants. This lack of numbers has hindered the study and understanding of effectors and regulators of the gene networks (both upstream and downstream) concerned with the female reproductive organ development. Given the understanding of the GRNs and the complexity and redundancy of these processes, many of the established methodology are not applicable to exploratory research of these GRNs as they mostly target one gene or genomic region at a time. In addition to the classical techniques such as chemical, radiation and insertional mutagenesis, currently available techniques and experimental approaches including but not limited to the application of high-throughput genomics along with Translating ribosome affinity purification (TRAP); (Heiman et al. 2008), (Jiao & Meyerowitz 2010), Isolation of nuclei tagged in specific cell types (INTACT; (Deal & Henikoff 2010)), a combination of Fluorescence Activated Cell Sorting (FACS) and Laser Capture Microscopy (LCM) (Wuest et al. 2010), (Liu et al. 2011), ChIP-Seq and proteomics will enable us to garner information to fill the gaps in our knowledge about the stage-specific development of the floral organs.

To circumvent the problem of redundancy and to target phenotypes, we generated a collection of Arabidopsis mutants via chemical mutagenesis using ethyl-methyl-sulfonate (EMS). EMS induces point mutations in the nature of G/C to A/T transitions, which could be detected as single nucleotide polymorphisms (SNPs) in the population. The advantage of this process is that EMS induces random mutations, allowing us to simultaneously target multiple genes in a genome-wide fashion. We then screened the mutants as separate families for specific phenotypes defective in reproductive processes leading to disruption or reduction of seed set.

Through this approach, we hoped to be able to screen for mutants defective only in female gynoecium development. We also sought to maintain a seed pool so that we could trace back the mutant of interest to its originating family, and thereby identify the genes and their roles in development.

Methods

EMS Mutagenesis

Ethylmethylsulfonate (EMS) induces single nucleotide polymorphisms (SNP) into the genome by the alkylation of guanine residues leading to a G:C to A:T base transition. We carried out pilot experiments to deduce the range of EMS concentrations and treatment time variations suited to our aims. We found that EMS concentrations of 40 mM for a treatment time of 14 hours gave the preferred 50% lethality rate associated with approximately 1 SNP in 300 kb of genomic DNA (Table 2.1).

Prior to EMS treatment, wildtype Col-0 Arabidopsis seeds were stratified by soaking them in 100 mM phosphate buffer (pH 7.5) at 4 C for 4 days. The buffer was then aspirated and replaced with 40 mM EMS solution in phosphate buffer for a duration of 14 hours by gentle rotation. Seeds were then washed 5 times with equal volume of 100 mM Na-thiosulphate to remove traces of EMS, followed by 5 washes with equal volume of ddH2O and planted immediately.

Growing and screening of treated plants

EMS treated seeds were planted in a 3:1 ratio of potting mix: vermiculite and grown in growth room conditions of 20-22 C and 16h/8h daylight period with 125 $\mu molm^{-2}s^{-1}$ of incident light intensity. Seeds harvested from each M1 plant were maintained as distinct M2 families. This strategy of harvesting allows us to recover sterile mutants as heterozygous siblings. Additionally, this strategy almost guarantees that two mutants with similar phenotypes from different M1 plants will have been a result of independent mutation events (Maple & Møller 2007). Primary screening was done by scoring defects in floral development, reduction in fertility (reduced seed set), and distortion in segregation ratios. Secondary screening was done by dissection of selected mutants to look for defects in gynoecium development. Twenty-four seeds from each M2 line was planted to maximize probability of observing obvious visible defects and to make note of the segregation ratio of the observed defects.

To test for female sterility versus male sterility, suspected mutant plants were used as pollen donors (male parent) to fertilize emasculated flowers of wildtype Col-0 ecotype (isogenic unmutagenized progenitor). Simultaneously, WT Col-0 plants were used as pollen donors to fertilize emasculated mutants. This step has the added advantage of generating our backcrossed mutant F1 population.

Results and Discussion

Through our study, we independently generated 2386 individual mutant lines (M2 lines) from the mutagenized progenitor seed stock of *Arabidopsis thaliana* Col-0 ecotype by EMS

induced mutagenesis. As expected from an EMS induced mutagenesis screen, we observed a random distribution of mutants with aberrations in almost every possible physiological and developmental aspect of the plant. We identified a wide variety of morphological, physiological and reproductive phenotypes ranging from chlorophyll biosynthesis (variegated leaves, pale green leaves, albino lethal plants), altered architecture (dwarf mutants) (Figure 2.1), altered leaf phyllotaxy, altered leaf morphology (serrated margins, altered leaf shape size and number), complete lack of rosette leaves, altered trichome development (increase and decrease in trichomes), flowering time mutants (both early and late flowering), reduced fertility, altered floral architecture (complete lack of petals, sepalloid petals) (Figure 2.2), reproductive defects, non-viable seed-set, miniature flowers and even one instance of an aberrant phenotype with 3 cotyledons. Some of the more interesting mutants in our collection that are relevant to this study show dramatically changed floral and morphological structures and characteristics, i.e. ranging from complete lack of petals to petaloid sepals to fused stamens, severely deformed gynoecia, reduced fertility, a mutant with flowers and buds clustered at apical meristem (reminiscent of *ap1-cal* mutants) (Figure 2.1.H) and miniature flowers.

In addition to these mutant phenotypes, we also observed mutants that have aborted siliques, or siliques that failed to mature and elongate. This appeared to be random along the same inflorescence that contained mature and fertile siliques. Mutants showing these phenotypes appeared multiple times in our collection of M2 families as separate instances and could be indicative of a) independent mutations which are possibly allelic, or b) ‘leaky’ phenotypes expected from EMS mutagenesis experiments. One of our mutants of interest, M2

line 23, seems to belong to this particular class of mutants, while displaying additional mutant traits in general plant architecture and size.

Previous studies yielded and documented very few female reproductive organ mutants while other mutant phenotypes such as male reproductive organ defective, embryo defective, reduced fertility and various other morphological defects have been reported with relatively much higher frequency. The rarity of female mutants may be a consequence of the fact that they do not set seeds, or set seed with very low frequency, making it is easy to miss this class of mutants from pooled seeds as generally practiced in high throughput screen. The only way to maintain and explore mutants of this type is through heterozygous mutant lines which can then be traced back by maintaining the mutant lines as distinct individual families, necessitating the use of a strategy whereby we harvest seeds from each M1 plant (plants grown from EMS-treated seeds) separately and generate a collection of M2 families (second generation derived from M1 plants).

The other reason that can be hypothesized for the low occurrence of the female sterile mutations is that in random-mating populations, a plant without functional androecium can still reproduce sexually with pollen from other plants and achieve full seed set because pollen is relatively abundant and typically mobile. On the other hand, a female sterile plant that only makes functional pollen, must compete with other pollen for fertilization for the mutation to persist. In this case, unless there is a significant increase in pollen production or survival leading to a much higher reproductive fitness (greater than 1.00), female sterility has a very small chance of being established in the population (Charlesworth & Charlesworth 1978). This applies an

evolutionary constraint, thereby increasing genetic redundancy in the controlling gene networks (Ming et al. 2011).

Mutant Lines

The primary goal of this study was to identify female sterile mutations. Screening of the 2386 M2 lines for reduced seed set, abortion and defects in gynoecia allowed us to identify 3 individual female sterile mutant lines. Male sterile mutants range in the categories of fused stamens, increased and decreased number of stamens, indehiscent anthers, infertile pollen and changes in pollen color and shape and were more numerous.

Female sterile mutants that we observed are in the categories of ovule abortion, gametophyte defects, randomly distributed whole silique abortion in the same individual plant, and severe defects in gynoecium structure and shape. Since we are mainly focused of gynoecium structural and morphological defects, aberrations in any gynoecium structure, particularly in the early stages of development were our primary interest, rather than female gametophytic mutants. In this context, we have been able to identify morphological defects with implications in fertility. The 3 identified mutant M2 lines have resulted in a complete lack of post fertilization silique elongation and maturation even after anthesis. In addition, in each of these 3 M2 lines FS21, FS23, and FS322 (Figure 2.3, 2.4 and 2.5) mutant phenotypes occur at very low frequency (approximately 1 in 16 plants) suggesting that the mutant phenotype is recessive or harbors a loss-of-function mutation. Closer microscopic examination of the dissected flowers of these lines

revealed that the stamens bear morphologically normal anthers with dehiscing pollen, except in M2 line 322, which bears defective and non-dehiscing anthers.

To test whether these mutant M2 plants produce viable pollen and if the lack of fertilization is due to self-incompatibility, we made reciprocal crosses with WT Col-0 plants of Arabidopsis as both pollen parent and pollen recipient. These crosses did not result in the expected silique elongation and maturation when WT Col-0 was used as the pollen parent. However, when the mutant was used as a pollen donor, the crosses gave viable seeds with varying degree of germination and survival of the progeny generation. This indicates that the mutants are likely female sterile with functioning male counterparts. Additionally, these reciprocal crosses will also help to reduce the non-causal SNPs in their genetic background.

M2 line FS21

When line 21 was used as the male parent with a WT as the female parent both fertilization and seed set were observed. However, a reduced degree of germination was observed in the progeny. In contrast, the reciprocal cross (Line 21 as female parent and WT as male parent) did not display signs of fertilization or seed set (silique elongation and maturation). This indicated that mutant line 21 is likely female sterile while producing viable male pollen. Closer inspection of the floral parts of line 21 revealed subtle defects in stigma structure when compared to WT stigma (Figure 2.3.B and G). The stigma of the stage 13 flowers (Smyth et al. 1990), 1 day after anthesis, of line 21 shows highly reduced stigmatic papillae while other parts

are morphologically normal. In comparison, WT stigma show elongated stigmatic papillae at this stage.

Morphologically, this mutant shows an altered plant growth habit compared to WT Col-0. The rosette and cauline leaves are reduced in size and display a slightly altered phyllotaxy. In addition, the mutant exhibits a much higher frequency of axillary branches and inflorescences along with a reduced overall plant height. Inflorescence meristem of this mutant also contains reduced number of flowers and floral buds compared to the WT (8-10 flowers and buds compared to 20-25 in the WT). This is however offset by the increase in lateral branching and increase in lateral flowers compared to the WT phenotype.

The structure of the flower in the mutant is not very different from the WT condition. The mutant flowers are slightly smaller than the WT flowers and more rounded in shape. The arrangement and number of floral organs within the mutant flower is similar to that of the WT, with no other visible defects (apart from the stigma).

The backcrossed progeny resulting from a cross using the mutant as a pollen donor and WT as the female parent does show a reduced seed set as well as reduced germination frequency, compared to that of a WT progeny in our lab growth conditions.

M2 line FS23

Mutant M2 line 23 (Fig. 2.4) also displays a phenotype similar to the one shown by mutant M2 line 21, i.e. with lack of fertilization in the ovules on the mutant plant, short plant stature and

smaller leaves. Initially it was difficult to generate a backcrossed population form this mutant line, leading us to expect a complete loss of fertility in both male and female systems. However, we were able to use the pollen from the mutant to fertilize WT ovules. Some individuals from the same M2 line do fertilize the WT ovules readily and self-fertilize while having other phenotypic features such as short stature and smaller than WT flowers. This may indicate 2 separate point mutations that create a total loss of fertility not related to the aberration in the growth stature of the plant. This could also indicate the segregation of the traits as well as the heterozygote and homozygote dominant (WT like) genotypes within the M2 line. The pollen of the mutants from this line however, do not show any obvious visible defects or lack of dehiscence.

The morphology of the mutants from M2 line 23 are also notable because it displays characteristics observed in the M2 line 21. However, there are some distinct differences too. Segregating mutants of line 23 display random abortion of siliques on the same inflorescence that contains properly elongating mature and fertile siliques (Fig. 2.4). This could be a pleiotropic effect of a disrupted function, or a 'leaky' phenotype that is dependent on additional factors apart from exclusively genetics interactions. Mutants in this family also display significantly smaller leaves, flowers, and stature. The mutants don't show a reduction in the number of leaves, flowers, and inflorescences compared to the WT Col-0 phenotype. These characters make it similar, yet different from the mutant line 21. Given the design of our mutagenesis experiment, each family or M2 line arises from an independent mutation event, and therefore, mutants 23 and 21 might represent allelic mutations of the same gene, or mutations in regulatory elements acting as a higher order complex in the same developmental pathway.

M2 Line FS322

M2 line FS322, displays some interesting phenotypes. The individuals grown from the M2 line show a segregation in severity of the mutant phenotype, with some plants displaying deformed and non-functional anthers, while some others harbor functional and dehiscing anthers with viable pollen- indicating that this could be a recessive mutant and could only be maintained as a heterozygote. This necessitated the use of siblings of the suspected homozygous recessive mutant to generate a backcrossed mapping population. The backcrossed (BC) progeny generated using siblings of mutant FS322 as a male parent and WT Col-0 as female parent shows segregation of phenotypes in a ratio that is distorted from the typical expected Mendelian ratios (Expected 7:1, observed ~1:15) (Page & Grossniklaus 2002). Individual plants from the BC progeny segregated by severity of the mutant condition as well as in the variation in phenotypes. Some plants display phenotype similar to the PINOID (PID) and PINFORMED (PIN) mutant defective in auxin transport (Bennett et al. 1995), (Christensen et al. 2000), (Benjamins et al. 2001), (Furutani et al. 2004), (Y Cheng et al. 2006). The inflorescence meristems in such cases terminated with a pin-like appendage with purple coloration (Fig. 2.5 A and C) and reduced number of flowers (Figure 2.6 I), and the deformed stigma taking the shape of an upturned saucer (Fig. 2.5 E and H). These plants also had a reduced number of cauline leaves, reduced number of inflorescences, and reduced lateral branching (Fig. 2.5 I).

At the same time, other plants from the same BC progeny (constructed using a heterozygous sibling) display only the inflorescence meristem terminating in a pin-like appendage, and a highly deformed gynoecium, while the other deviations from the WT phenotype were not visible. These plants however, still do not undergo self-fertilization and

silique elongation and maturation, but undergo post fertilization silique elongation and maturation when used as the pollen donor in crosses with the WT Col-0 plants with highly reduced seed set (3-4 mature seeds in each silique). These phenotypes show similarities with the YUCCA (YUC) mutant defective in auxin biosynthesis, particularly *yuc1-1yuc4-1* double mutants, i.e., complete absence of ovary valve with enlarged apical stigma in the shape of an upturned saucer (Cheng et al. 2007). The stigma itself does not have much in terms of papillae on the top surface, rather, the papillae are mostly concentrated on the sides (Figure 2.5 E and G, compare to J).

The *yuc1-1yuc4-1* phenotypes are also visible on other individual plants within the same M2 line without either the pin or pid mutant phenotypes. It should also be noted that none of the mutant phenotypes of this M2 line display changes to petal number and organization, as is to be expected from the mutant phenotypes of PINFORMED or PINOID. However, the petals are much thinner and ribbon like and smaller than that of a WT phenotype.

Discussion

The main purpose of this study was to generate and characterize reproductive mutants deficient in the female floral organ developmental processes. The flower in an angiosperm is an incredibly complex structure, which is ontogenetically a highly modified leaf. It is expected that certain fundamental and basic developmental programs would tend to be similar between leaves and flowers. At the same time however, the extent of modification and the sheer range and variety in structure and organization of the flower is highly suggestive of the canonical 'ABCE'

model based on homeotic genes. Contrary to expectations, gene perturbation studies in the ABCE model have not yet resulted in complete transformation of a leaf into a floral organ or that of a floral organ to a leaf in its native configuration. Efforts to understand this paradox led to the identification of a lot of additional genetic and environmental players that fine tune the canonical model. The simplicity and effectiveness of the ABCE model although, is unable to explain the differences in floral structure and sexuality that arise without either the modification of this model and/or the existence of sex chromosomes in many angiosperm species. The mutants developed during this study would help improve our understanding and add to the knowledgebase of developmental biology.

The generation of mutants specific to the female organ abortion, followed by a detailed process of mutant identification and characterization will be the key to the dissection of the gene regulatory network specific to the determination and development of the female organs and identify genes other than the homeotic genes, possibly upstream or downstream acting genes, cis- or trans-acting regulatory elements involved in the process. Having a repertoire of genes in these categories will eventually help us to be able to predict the sex determination network and in the future, be able to engineer plants to the goal of having hermaphrodite populations, bypassing or avoiding the huge losses incurred by the agriculture community in dealing with the 1:1 segregation of male: female plants in dioecious crops. Hermaphrodite plants in populations have the advantage of being able to simultaneously bear fruit and still allow outcrossing (to maintain genetic diversity) when the mutations or sex-determining genes are on non-recombining chromosomes, and therefore minimizing the possibility of reversion to the state of dioecy.

Tables and Figures

Table 2.1: Germination and survival rates of EMS treated seeds at different treatment conditions.

Treatment	Total no. of seeds planted	Germination rate	Survival rate (after 30 days)
40mM, 14hrs	400	219 (54%)	196 (49%)
40mM, 18hrs	400	180 (45%)	162 (40.5%)
40mM, 22hrs	400	160 (40%)	144 (36%)

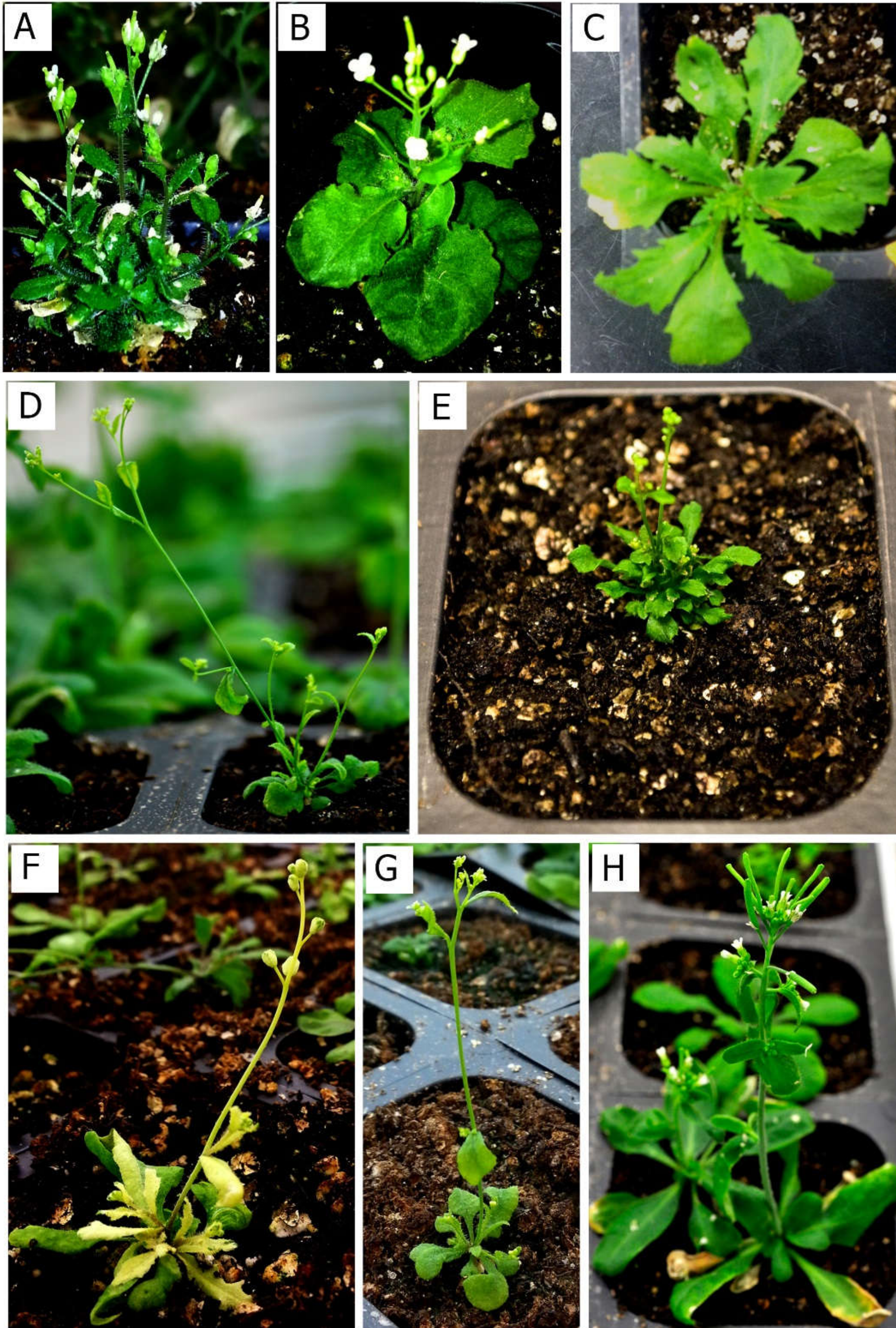

Figure 2.1: Variation in mutant phenotypes. (A) Dwarf with increased trichomes, (B) Severe dwarf and early flowering along with altered leaf phyllotaxy, (C) Serrated leaf margins, (D) Early flowering, (E) Dwarf and late flowering, (F) Variegated leaves, (G) Early flowering, (H) Altered inflorescence meristem.

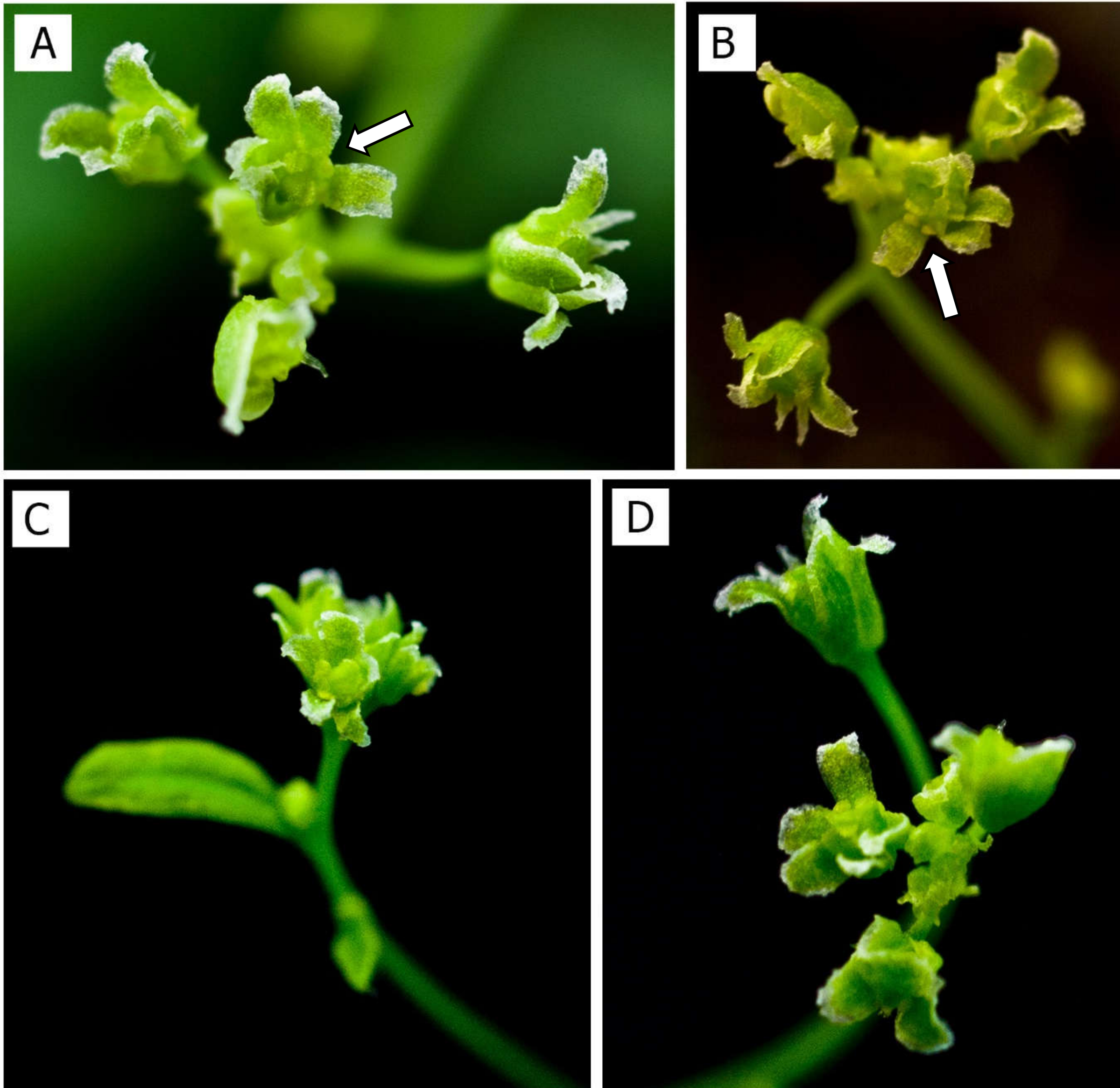

Figure 2.2: Mutant flowers showing altered floral structure. (A) and (B) Flowers showing complete lack of petals, (C) and (D) Flowers showing petaloid sepals. In addition, these flowers also lack the normal number of anthers (arrow).

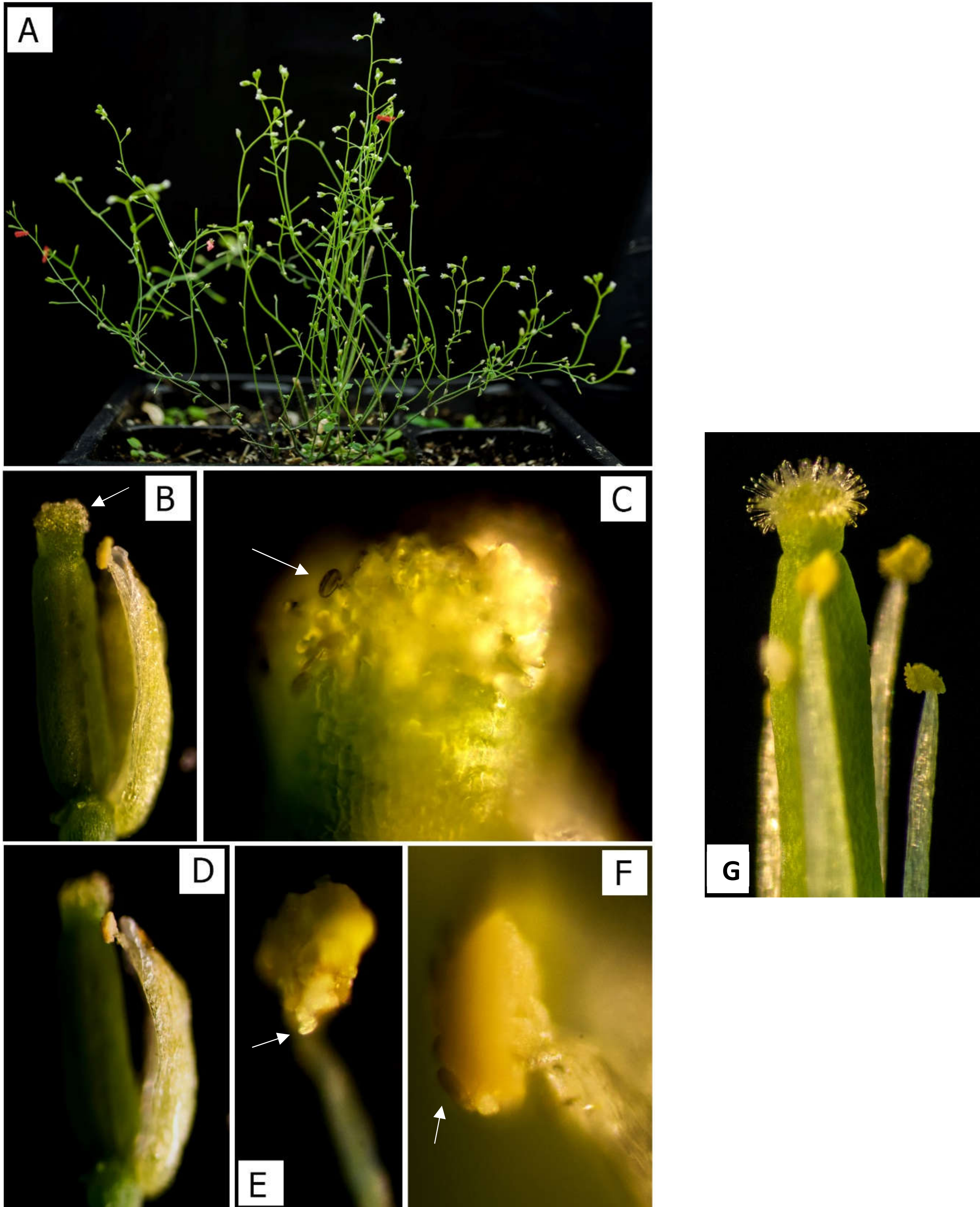

Figure 2.3: Mutant M2 line FS21: (A) Growth habit and architecture of mutant plant, (B) Micro-photograph of mutant gynoecium (note reduced stigmatic papillae shown by arrow), (C) Magnified view of stigma with pollen (arrow), (D) Micro-photograph of mutant anther, (E) and (F) Magnified view of dehiscent anther with pollen (arrow).

(G) Side panel shows Wildtype stigma and anther for comparison (petals have been removed)

Figure 2.4. Mutant M2 line FS23: (A) Growth habit and architecture of mutant plant, (B & D) Siliques displaying lack of maturation and elongation (arrows), (C) Flower morphology showing subtle changes to structure, (E & F) Stigma displaying reduced papillae (arrow).

(G) Side panel shows Wildtype stigma and anther for comparison (petals have been removed)

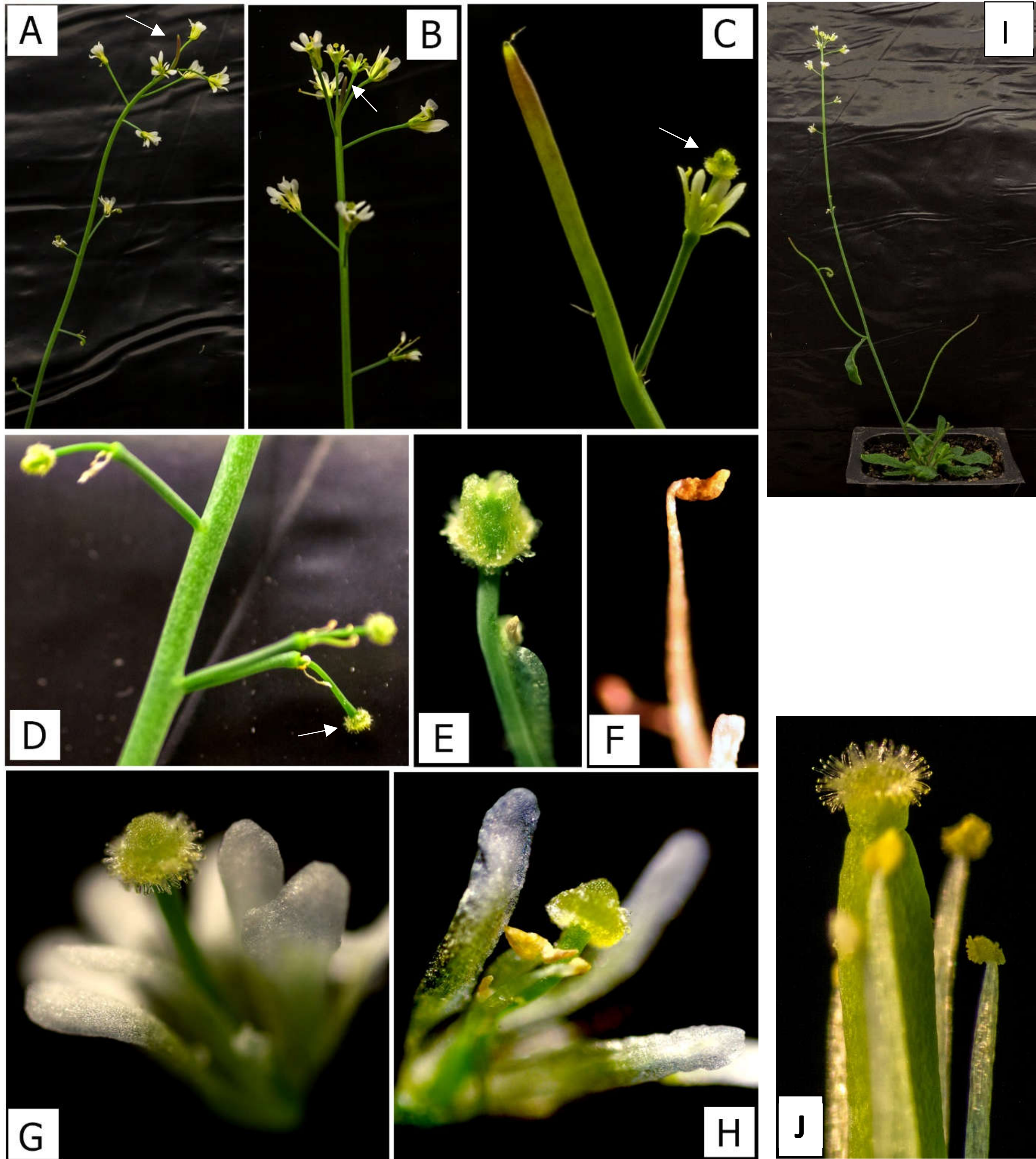

Figure 2.5. Mutant M2 line FS322: (A) and (B) Structure of inflorescence (note pin-like apical meristem shown by arrow), (C) Tip of apical meristem and open flower showing defective stigma (arrow), (D) Open flowers showing severely deformed gynoecia (arrow), (E)Magnified view of gynoecia and stigma with characteristic 'dome-shaped' stigma, (F) Magnified view showing non-dehiscent and deformed anther, (G) and (H) Complete flower showing characteristic stigma, petal and anther deformity, (I) Growth habit and architecture of mutant plant.

(J) Side panel shows Wildtype stigma and anther for comparison (petals have been removed)

Chapter 4: Identification of genomic regions associated with female sterility mutant FS322

Abstract

Female sterility has not been systematically investigated because of the difficulty to maintain such mutants in heterozygous state, low throughput, and extra generation with recorded pedigree required for screening female sterile mutants. The deficiency of female sterile genes hindered the establishment of sex determination gene network. The developmental program of reproductive organs is highly coordinated and involve multiple genes with overlapping functions. Furthermore, these genes can be located anywhere in the genome. Here we used EMS mutagenesis with the technique of whole genome sequencing to identify chromosomal regions related to the phenotype of the female sterile mutant FS322. The single nucleotide polymorphisms (SNPs) induced by EMS clustered on the long arms of chromosome 1 and 3, with many of the SNPs causing non-synonymous and STOP codon changes in genes known to have roles in floral development. We also identified genes with unknown functions and with primary roles other than in reproductive morphology. These candidate genes provide the basis for further exploration of the causal gene of FS322 and elucidation of floral organogenesis; and will improve our understanding of reproductive development in plants.

Introduction

Forward genetic screens utilizing map-based cloning approaches have been instrumental in revealing the identity and functions of genes in plants as well as many other organisms. In case of plants, Arabidopsis has served as the model organism for dicots for over 3 decades owing to its ease of cultivation and manipulation in a laboratory, and small genome size of 125Mb. The sequenced genome of Arabidopsis from the Columbia (Col-0) accession has been extensively used to understand plant development, physiology and metabolism (Arabidopsis Genome Initiative, 2000; reviewed in The Arabidopsis Book, 2008). Much of our knowledge and understanding of gene identity and functions have come from gene perturbation studies and molecular analyses of mutant phenotypes resulting from these studies. However, generation of these mutants and the subsequent mapping of the involved genes using traditional approaches are time consuming and laborious. With the advent of next generation sequencing (NGS) technologies, combining the traditional methods with high throughput sequencing has accelerated the identification of the underlying causal mutations for a wide range of mutant phenotypes.

Initiation and development of reproductive organs in plants are highly complex and coordinated biological processes. The ABCE model elegantly explains the initiation and development of floral organs orchestrated in a combinatorial fashion (Coen & Meyerowitz 1991; Zahn et al. 2005; Zahn et al. 2006). Many of the components of this model, including the canonical genes with the A, B, C, and E functions were identified by the analysis of mutants using a forward genetic approach. Increasingly, the study of the gene regulatory networks (GRNs) involved in this developmental process have shown a high level of genetic redundancy. The proper development

of the reproductive organs also involve the interactions of the A, B, C and E class of genes with a host of other gene products and cofactors, such as UNUSUAL FLORAL ORGANS (UFO), LEAFY (LFY), WUSCHEL (WUS), LEUNIG (LUG), CRABS CLAW (CRC), SEUSS (SEU), SUPERMAN (SUP), KNUCKLES (KNU) and microRNAs miR160 and miR172 to name a few (reviewed in Krizek & Fletcher 2005). These higher order interactions take place in a spatial-temporal manner, starting from the signals and events required for the initiation of floral primordia, to the separation of the floral organs into 'whorls' and including the continued presence of these gene products in the later stages of development of the floral organs.

In addition to these well-known genes that act both upstream and downstream of the ABCE model, the databases such as RARGE II and Seed Genes Project also reveal a host of other genes, both with known functions and unknown, that are implicated in the development process of the female organs (Akiyama et al. 2014; Lloyd & Meinke 2012). Some of these genes are YUCCA1, YUCCA4, PIN, PINOID (working in the auxin hormone biosynthesis and response pathways), ACC synthase (working in the ethylene biosynthesis pathway), and COP1-interacting protein-related, while some of the genes with unknown functions include ribosomal proteins (L39), leucine-rich-repeat transmembrane proteins, and zinc finger proteins, AP2/B3-like transcription factor family protein (Youfa Cheng et al. 2006; Cheng et al. 2007; Xing et al. 2013; Bennett et al. 1995; Benjamins et al. 2001; Furutani et al. 2004; Christensen et al. 2000; Lampugnani et al. 2013; De Martinis & Mariani 1999; Boualem et al. 2008; Boualem et al. 2009; Sherif et al. 2009; Wei & Deng 2003; Stewart et al. 2016; Krogan et al. 2012; Zik & Irish 2003). This redundancy can be explained by the necessity of faithful transfer of genes and reproduction of functional progeny to ensure survival of the species. As such, the phenotypic identification of

mutant phenotypes can be hindered by the necessity to simultaneously perturb multiple genes. Variation in severity of phenotypes further complicate and slow down the mapping process. Single nucleotide polymorphisms (SNPs) caused by radiation or chemical methods are therefore the most well suited tools for maximum probability of observing mutants defective in flower development.

SNPs generated by chemical mutagenesis using ethyl-methyl sulfonate creates G:C to A:T transition mutations that are randomly distributed across the genome. However, EMS creates a high mutation load, and limits the use of direct whole genome re-sequencing of an individual mutant. A combination of bulked segregant analysis and genome re-sequencing provides a method of reducing the number of point mutations and enables fast identification of the putative genomic region harboring the genes responsible for the mutant phenotype. Here we re-sequence the female sterile mutant FS322 and its corresponding wild type bulked siblings to identify the causal mutation(s) and the genomic region containing the genes associated with the mutation.

Materials and methods

Plant material

A previously identified female sterile mutant FS322 generated from the Columbia 0 (Col-0) background utilizing EMS mediate mutagenesis was used in this study. FS322 was backcrossed with its isogenic un-mutagenized progenitor line (Col-0), and the resulting F_1 progeny was allowed to self-fertilize to give rise to the backcrossed F2 progeny (BC_1F_2). The segregating F_2 population was then screened for the mutant phenotype and bulked into a mutant pool and a

WT-like pool. All plants were grown from seeds stratified by soaking them in 100 mM phosphate buffer (pH 7.5) for 4 days at 4C. Plants were grown in 3:1 mixture of LC1 potting soil and vermiculite under growth room conditions of 16 hours light/8 hours dark cycles at 22-24C and light intensity of 125 μmol m^{-2} s^{-1}.

DNA isolation, library preparation and sequencing

Approximately 700 BC_1F_2 were grown, and leaf samples of equal quantity were collected from 30 plants scored as exhibiting the mutant phenotype of FS322. In parallel, leaf samples were collected from another 30 plants categorized as displaying the WT-like phenotype. Genomic DNA was extracted individually from each leaf sample using the CTAB method and then quantified using Nanodrop 2000 spectrophotometer (ThermoScientific, USA) and visualized on agarose gel (1% TAE-agarose) by electrophoresis. Equal amounts of genomic DNA from each individual sample was pooled to construct the mutant bulk as well as the WT-like bulk. DNA library was prepared using NEBNext Ultra DNA Library Prep Kit (Illumina, Cat # E3730) according to the manufacturer's instructions. Pooled and indexed libraries were sequenced on the Illumina HiSeq X system, using 150nt paired-end protocol. Quality control was performed using FastQC to remove adapters and low quality sequences, allowing an average retention of 11,709,434 clean reads per library (Q30 = ~90%).

Sequence analysis

We used BWA (Li & Durbin 2009) to independently align the clean read sets of both mutant bulk and the WT-like bulk to the Col-0 reference genome of Arabidopsis (Li & Durbin 2009; Garcia et al. 2016; The Arabidopsis Information Resource 10). The resulting alignments for

both pools were then manipulated using SAMtools and variant calls were made using the 'mpileup' function built into SAMtools (Li & Durbin 2009; Li et al. 2009). The resulting variant files containing the SNPs were then analyzed and annotated with the TAIR 10 and AraPort 11 gene models for Arabidopsis in parallel using the Galaxy web based platform, a standalone software package snpEff and the web based tool suite SNPTrack (Afgan et al. 2016; Lindner et al. 2012; Leshchiner et al. 2012; http://genetics.bwh.harvard.edu/snptrack/). Filtering of SNPs were carried out using the VCFtools program package based on sequence quality, read coverage and depth at the SNP site (Danecek et al. 2011).

Results

FS322 is a female sterile mutant that was identified from a previously described EMS mutagenesis study that we conducted. This mutant displays interesting phenotypes as well as segregation of the severity of the mutant phenotype. Some plants display phenotypes similar to the PINOID (PID) and PINFORMED (PIN) mutant defective in auxin transport (Bennett et al. 1995; Christensen et al. 2000; Benjamins et al. 2001; Furutani et al. 2004; Y Cheng et al. 2006). The inflorescence meristems in such cases terminated with a pin-like appendage with purple coloration and reduced number of flowers (Figure 3.1). Others show similarities with the YUCCA (YUC) mutant defective in auxin biosynthesis, particularly *yuc1-1yuc4-1* double mutants, i.e., complete absence of ovary valve with enlarged apical stigma in the shape of an upturned saucer (Cheng et al. 2007). In this case, the stigma, itself does not have much in terms of papillae on the

top surface, rather, the papillae are mostly concentrated on the sides (Figure 3.1.E and G, compare to J).

The *yuc1-1yuc4-1* phenotypes are also visible on other individual plants within the same M2 line without either the *pin* or *pid* mutant phenotypes. It should also be noted that none of the mutant phenotypes of this M2 line display changes to petal number and organization, as is to be expected from the mutant phenotypes of PINFORMED or PINOID. However, the petals are much thinner and ribbon like and smaller than that of a WT phenotype. These plants also had a reduced number of cauline leaves, reduced number of inflorescences, and reduced lateral branching. Some of the mutants have deformed and non-functional anthers, while others harbor functional and dehiscing anthers with viable pollen- indicating that this could be a recessive mutant and could only be maintained as a heterozygote. We therefore used a mutant individual that made viable and dehiscing pollen but did not set seed when allowed to self-pollinate, as the male parent and used a wild-type Col-0 plant as the female parent for the generation of the backcrossed (BC) mapping population. The BC progeny showed segregation of phenotypes (Table 3.1) in a ratio that is distorted from the typical expected Mendelian ratios of 3:1 (expected in BC population for recessive gene) and 7:1 (expected in M2 for chimeric mutant plant). This situation, as discussed earlier, is in agreement with the genetically equivalent cell number (GECN) of Arabidopsis seeds (GECN=2), where EMS mutagenized only one of the two cells (Page & Grossniklaus 2002).

The BC_1F_2 population was generated by crossing the FS322 as the male parent with the Col-0 line as unmutagenized progenitor female parent and then allowing the resulting F_1 plants to self-pollinate. Leaf samples were collected from 30 individuals of mutant-like and WT-like

phenotype and the extracted DNA was pooled into a mutant and WT pool. A total of 11,969,277 and 11,449,591 high quality reads (Q30 ~90%) with an average of approximately 25X coverage depth were generated for the mutant and WT pools respectively. The reference genome for Arabidopsis line Col-0 (TAIR-10) was then used to make variant calls for the mutant bulk and the WT bulk independently. The resulting SNPs from the WT bulk were then subtracted from the SNPs in the mutant bulk, and then filtered by coverage depth (threshold minimum of 10X), quality (>10, <100) and canonical G/C to A/T substitutions (EMS induced) to reduce the ubiquitous low background of false positive resulting from systemic error and background noise. In this study, we focused exclusively on SNPs because insertions and deletions are not associated with EMS mutagenesis. This resulted in a total of 852 SNPs across the genome, an average of 1 SNP in ~150kb.

Annotation and filtering of the SNPs based on the TAIR-10 gene models for non-synonymous SNPs in coding regions and for STOP site mutations further reduced the SNP count to 312. Of these, 35 SNPs resulted in a premature STOP codon being gained, 1 STOP codon being lost, 1 START codon being lost, and 21 splice site variants, while the remaining 254 SNPs were non-synonymous mutation in the coding regions. 12 of these SNPs (AT1G01120, AT1G15310, AT1G17090, AT1G17600, AT1G17910, AT1G24430, AT1G64100, AT2G17660, AT4G22485, AT1G51190, AT1G09620, AT1G64060) were filtered based on allele frequencies (AF > 0.5, < 1.0) and localized on the long arm of chromosome 1 (Figure 3.2). The STOP codon changes mostly clustered along the long arm of chromosome 3, with some isolated aggregations on other chromosomes as well (Figure 3.3). 18 of the SNPs causing premature STOP codons did not have a known protein function, and were distributed randomly across the genome, with 3 of them

(AT3G11890, AT3G12150, AT3G12835) clustered on a 326kb region on the long arm of chromosome 3 (Figure 3.4).

The 12 non-synonymous SNPs in the mutant pool (AF> 0.5) all have known protein functions, with the exception of AT1G17090 (Table 3.2). The most notable gene on this list is PLETHORA2 (PLT2) which is a member of the AINTEGUMENTA-LIKE (AIL) subclass of AP2/EREB family of transcription factors. The activity of PLT2 is dependent on auxin response transcription factors and plays a role in stem cell maintenance and pattern formation in response to auxin transport (Mudunkothge & Krizek 2012; Jia et al. 2015). SOC3, a gene coding a Toll/interleukin receptor (TIR)-nucleotide binding protein was also identified as having a non-synonymous mutation. SOC3 is involved in defense responses and chilling sensitivity and takes part in temperature mediated activation of cell death (Tan et al. 2007; Zhang et al. 2005). The others on this list are involved in signaling, seed storage, cell wall biosynthesis and wax biosynthesis (Table 3.2). The only gene on this list that gains a STOP codon, incidentally codes for a kinase-related protein with unknown function.

Most of the 36 SNPs that resulted in STOP codon changes that we found, had known functions attributed to the genes in which they appeared (Table 3.3). Included in this list are HUELLENOS (HLL), MEIOSIS DEFECTIVE 1 (MEI1), ESSENTIAL MEIOTIC ENDONUCLEASE 1B (EME 1B), SPATULA (SPT), RIBOSOMAL RNA PROCESSING 5 (RRP5), PRESEQUENCE PROTEASE 1 (PREP1) AND BRASSINOSTEROID-SIGNALING KINASE 2 (BSK2), all of which are known to play roles in development of the female reproductive organs. This list of gene loci also includes the genes AT3G12835, AT1G66680 (AR401) which are expressed in flower tissue but with unknown protein functions. PREP1 is an interesting gene coding a signal peptide involved in enzyme degradation

in mitochondria and chloroplast. It is expressed only in siliques and flower tissues and may be involved in processes that confer female sterility.

Discussion

The main purpose of this study was to identify genomic regions harboring mutations in a female sterile phenotype in Arabidopsis. Given that the developmental program of the female reproductive organs is a highly complex and coordinated process, we expected a significant amount of redundancy in gene function. It is interesting to note that except a few known genes that are members of transcription factor families, most other genes that were identified have primary roles not implicitly related to reproductive development. This suggests that we were successful in targeting genes taking part in processes upstream and downstream of the known floral specification and development genes. The identified SNPs in our study cluster on the long arm of chromosome 1 and chromosome 3, suggesting that the genes responsible for the observed phenotype of FS322 reside in those genomic regions.

The female sterile mutant FS322 displays a phenotype similar to the *pin*, *pid* and *yuc1-4yuc4-1* mutants (Figure 1). However, these genes did not appear in our list of candidate genes indicating that the phenotype is not caused by mutations in the genes involved in the *pin*, *pid* and *yucca* mutants (Bennett et al. 1995; Benjamins et al. 2001; Furutani et al. 2004; Trigueros et al. 2009; Xing et al. 2013). In addition, the mutant phenotype appeared in the M2 and BC_1F_2 population, but not in the BC_2F_2 progeny from additional backcrosses. These observations, when considered along with the GECN of Arabidopsis as discussed before, suggests that the phenotype is a multigenic trait that requires the simultaneous presence of mutations in multiple genes for

the. Scoring of the mutant phenotype in the F2 progenies was also made difficult by the fact that the subtle changes in fertility of FS322, typical of multigenic traits, may have resulted in overlooking or mis-scoring of the mutants.

Mutations in ribosomal proteins give rise to a variety of developmental phenotypes. The molecular basis for such defects are not fully understood. However, female fertility is sensitive to levels of ribosomal proteins and can give rise to variation in phenotype severity, as in the case of Ribosomal Protein L27a (RPL27a) and it's paralogs RPL27aB and ROL27aC in a dosage dependent manner (Zsogon et al. 2014; Devis et al. 2015), Ribosomal protein mutants are also reported to be viable, but with subtle changes in leaf shape, inflorescence defects and flowering (Ito et al. 2000; Pinon et al. 2008; Byrne 2011; Szakonyi & Byrne 2017; Stirnberg et al. 2012). We identified 6 SNPs in genes involved in ribosome biogenesis and are structural components of ribosomes (AT1G17560.1, AT1G64600.1, AT3G11964.1 AT3G25470.1, AT4G08350.1 and AT4G34730.1). Among these, AT1G17560.1 (HUELLENLOS) and AT3G11964.1 Ribosomal RNA processing 5 (RRP5) have been shown to be involved in the female reproductive organogenesis (Skinner et al. 2001; Missbach et al. 2013). AT4G08350.1 (Global transcription factor group A2, GTA2) and AT4G34730.1 (RBFA domain containing protein 1, RBFA1) are necessary for ribosome structure and function, but have no roles ascribed to them in female fertility so far.

The gene loci on chromosome 1 that were found to have non-synonymous or STOP codon changes identified some candidate genes for female sterility in the mutant FS322. Some of the most interesting ones include genes that have known functional domains but no associated molecular and biological functions (based on TAIR10 annotations). These include SNARE associated Golgi proteins, F-Box/RNI-like superfamily proteins, PLANT RING/U-BOX 18 (PUB18),

Receptor like protein 1(RLP1), Leucine Rich Repeat transmembrane protein kinase, GPI transamidase subunit PIG-U and protein-protein interaction regulator family protein. These gene loci are not inclusive of the other identified gene loci with unknown domain structures and functions and known players such as ARF12, SPL10, WD40 domain proteins, BEL-1 like homeodomain, PLETHORA2, CULLIN3.

Chromosome 3 contains a collection of SNPs that affect the functions of many genes known to influence female reproductive development. These include MYB77, Embryo Defective 2423, RRP5, Embryo Sac Development Arrest 30, Citrate synthase 1. Chromosome 3 also contains a significant number of Tetratricopeptide protein genes, Pentatricopeptide protein genes, and Ankyrin repeat protein genes with non-synonymous and STOP mutations. Interestingly, we also found an over-representation of these genes in the female tissue in our previous study of differential gene expression between male and female floral development (unpublished data).

Candidate genes identified through the current study therefore, present a potential for data mining and exploration of their functions specific to female sterility. We were successful in mapping EMS-induced SNPs in the FS322 line to the long arms of chromosome 1 and 3, suggesting that additional genes involved in the proper development of the female reproductive program are located on these chromosomes. Further study, involving larger pools of mutant plants followed by complementation analysis, is required to unequivocally associate genes in those regions with the phenotype and characterize their biological and physiological roles played in maintaining fertility of the female organs.

Additional studies could involve targeted gene editing protocols, such as the CRISPR-Cas9 system, to generate mutation in specific genes, or T-DNA insertion lines for complementation studies, timing specific and tissue specific gene expression localization studies, and in-silico protein-protein interaction studies based on predicted 3-dimensional protein structures, in order to sort potential roles of the genes in female reproductive development.

Tables and Figures

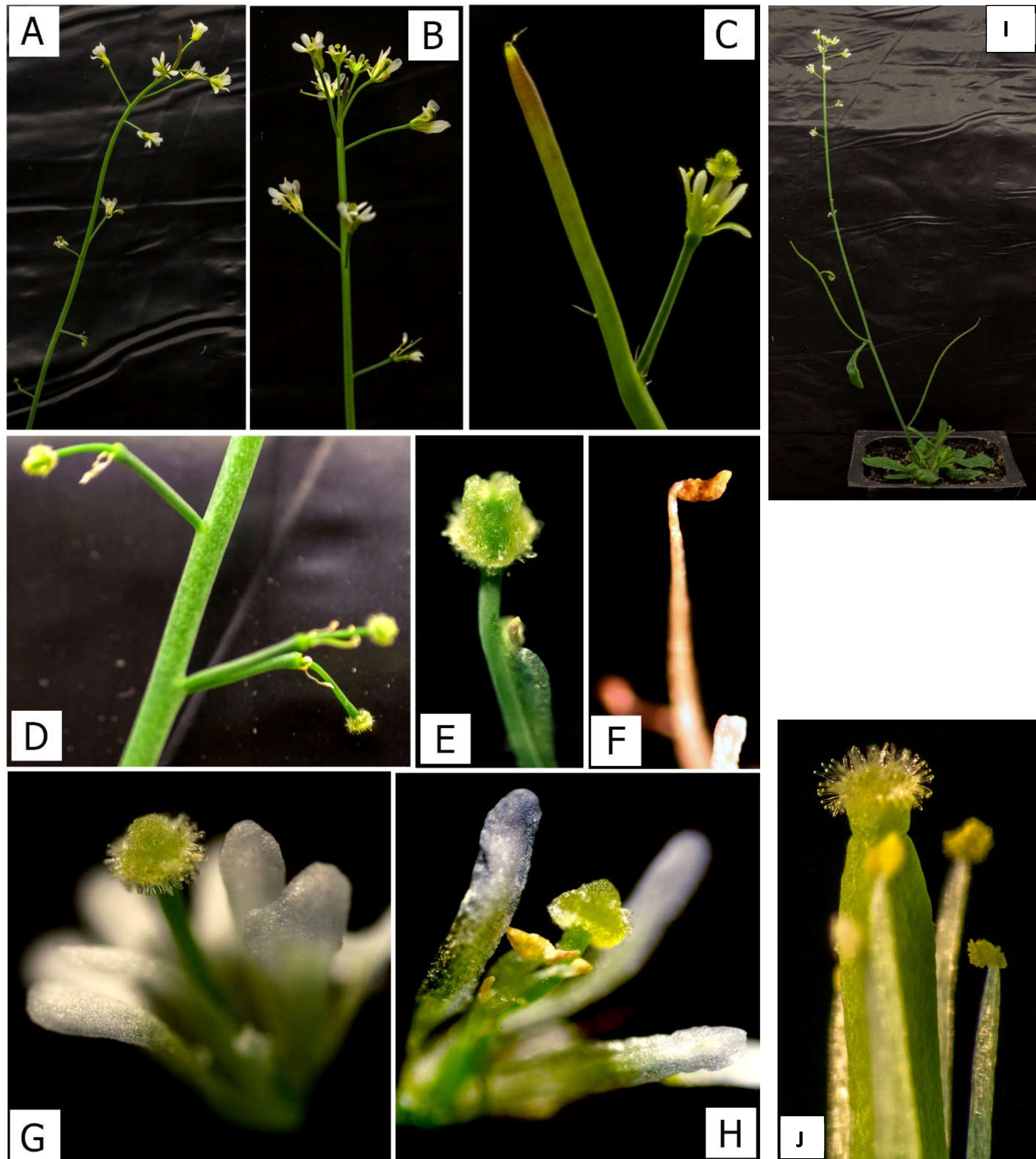

Figure 3.1. Mutant M2 line FS322: (A) and (B) Structure of inflorescence (note pin-like apical meristem shown by arrow), (C) Tip of apical meristem and open flower showing defective stigma (arrow), (D) Open flowers showing severely deformed gynoecia (arrow), (E)Magnified view of gynoecia and stigma with characteristic 'dome-shaped' stigma, (F) Magnified view showing non-dehiscent and deformed anther, (G) and (H) Complete flower showing characteristic stigma, petal and anther deformity, (I) Growth habit and architecture of mutant plant.

(J) Side panel shows Wildtype stigma and anther for comparison (petals have been removed)

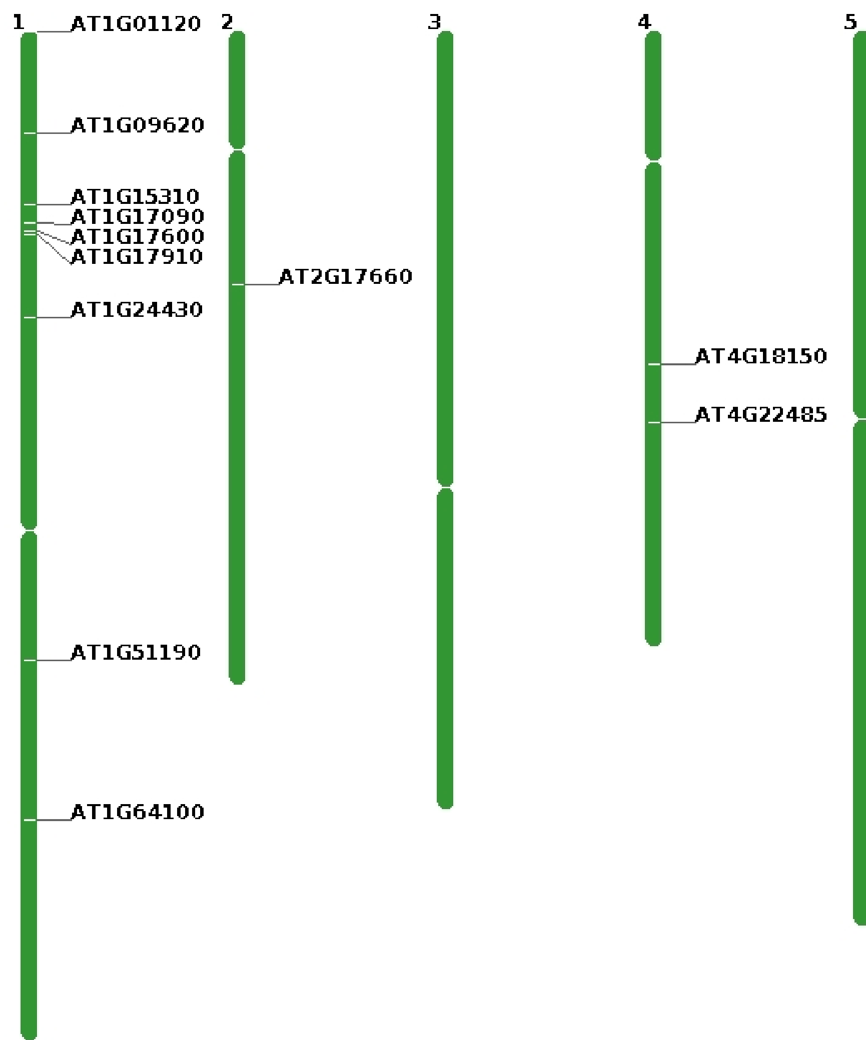

Figure 3.2: Chromosomal distribution and location of SNPs in 12 genes with unknown protein function

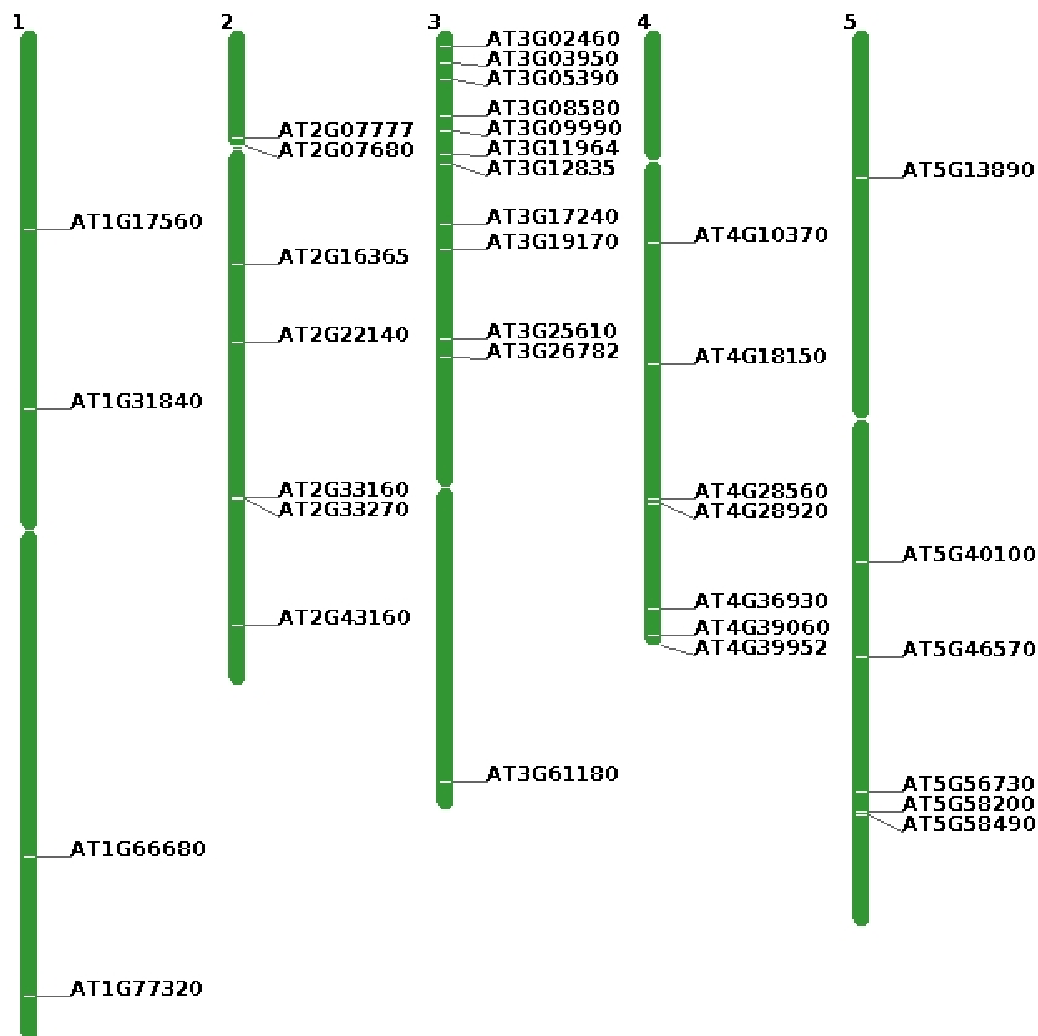

Figure 3.3: Chromosomal distribution and location of SNPs creating STOP codon changes

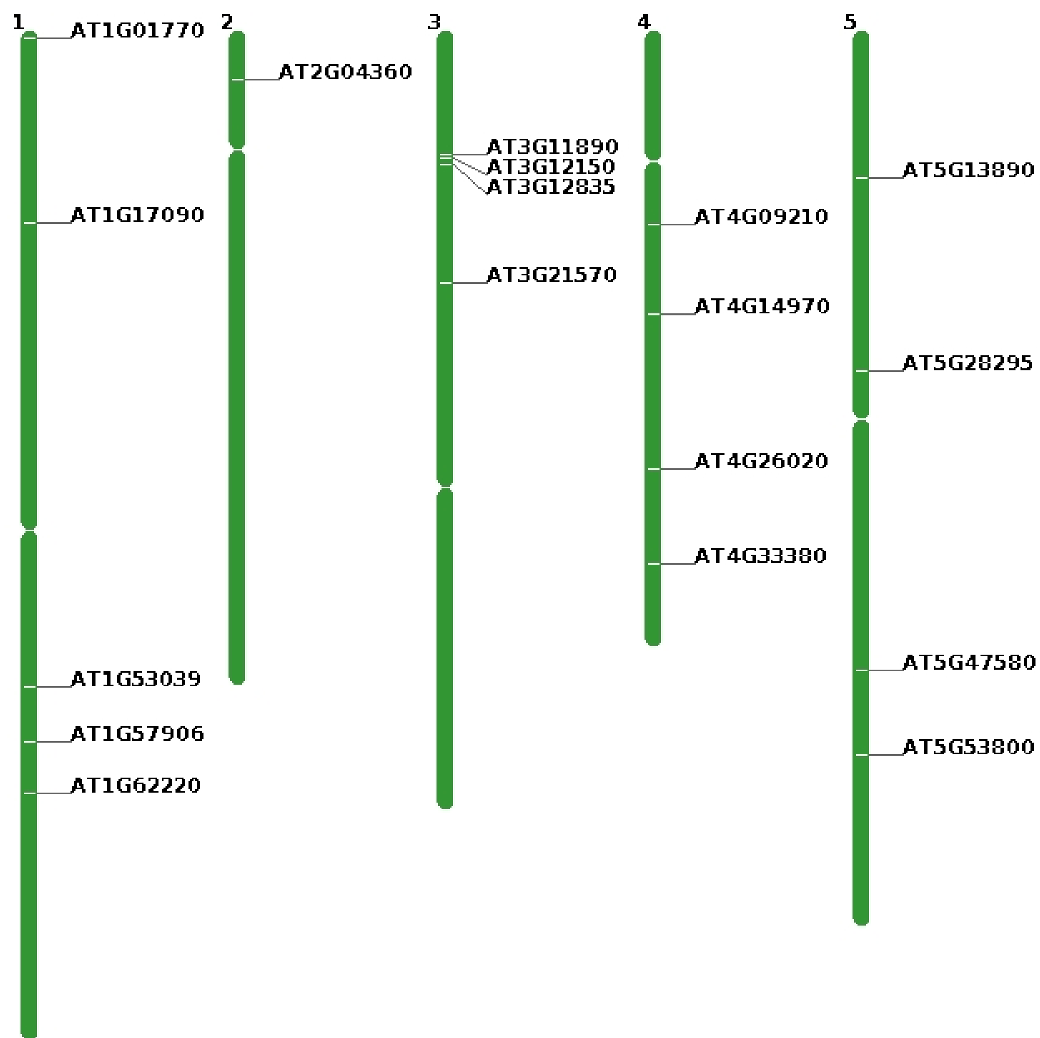

Figure 3.4: Chromosomal distribution and location of SNPs causing STOP codon changes in genes with unknown function

Table 3.1: Segregation ratio of mutant phenotype in backcross generations of the mutants (BC_1F_2). All backcrosses were conducted using WT progenitor as female parent and M2 mutant as male parent.

M2 Mutant Line	No. of seeds planted (approx.)	No. of plants		
		Wild Type phenotype	Mutant phenotype	Segregation ratio
FS 21	200	191	16	1:12.04
FS 23	200	189	14	1:13.51
FS 322	700	686	48	1:14.30

Table 3.2: Table 2. SNPs with known function that display allelic frequency (AF) higher than 0.5

Gene Model Name	Gene Model Description	Primary Gene Symbol	SNP Effect	Amino Acid change
AT1G17090.1	Unknown protein; INVOLVED IN: biological process unknown;	-	NON SYNONYMOUS	A->D
AT1G01120.1	Involved in the critical fatty acid elongation process in wax biosynthesis.	3-KETOACYL-COA SYNTHASE 1 (KCS1)	NON SYNONYMOUS	A->T
AT1G09620.1	ATP binding; FUNCTIONS IN: nucleotide binding, Arabidopsis thaliana protein match is: tRNA synthetase class I (I, L, M and V) family protein (TAIR:AT4G04350.1)	-	NON SYNONYMOUS	A->D
AT1G15310.1	54 kDa protein subunit of SRP that interacts with the signal peptide of secreted proteins	SIGNAL RECOGNITION PARTICLE 54 KDA SUBUNIT (ATHSRP54A)	NON SYNONYMOUS	I->K
AT1G17600.1	SOC3 is a TIR-NB-leucine-rich repeat (TNL) protein.	(SOC3)	NON SYNONYMOUS	L->H
AT1G17910.1	Wall-associated kinase family protein; FUNCTIONS IN: kinase activity; BEST Arabidopsis thaliana protein match is: Wall-associated kinase family protein (TAIR:AT1G19390.1)	-	NON SYNONYMOUS	R->L
AT1G24430.1	HXXXD-type acyl-transferase family protein; FUNCTIONS IN: transferase activity, BEST Arabidopsis thaliana protein match is: HXXXD-type acyl-transferase family protein (TAIR:AT3G26040.1)	-	NON SYNONYMOUS	M->I
AT1G51190.1	Encodes a member of the AINTEGUMENTA-like (AIL) subclass of the AP2/EREBP family of transcription factors and is dependent on auxin response transcription factors.	PLETHORA 2 (PLT2)	NON SYNONYMOUS	R->H
AT1G64100.1	pentatricopeptide (PPR) repeat-containing protein; BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT1G12300.1)	-	NON SYNONYMOUS	R->I
AT2G17660.1	RPM1-interacting protein 4 (RIN4) family protein, functions in defense response; BEST Arabidopsis thaliana protein match is: RPM1-interacting protein 4 (RIN4) family protein (TAIR:AT4G35655.1)	-	NON SYNONYMOUS	R->W
AT4G22485.1	Encodes a Protease inhibitor/seed storage/LTP family protein	-	NON SYNONYMOUS	Q->K

Table 3.3: SNPs with allelic frequency > 0.5 and causing STOP codon changes

Gene Model Name	Gene Model Description	Primary Gene Symbol	SNP Effect	Amino acid change
AT1G66680.1	Unknown function	(AR401)	STOP GAINED	S->Stop
AT1G17560.1	Encodes HUELLENLOS (HLL), HLL is essential for normal ovule development.	HUELLENLOS (HLL)	STOP GAINED	S->Stop
AT1G31840.1	Tetratricopeptide repeat (TPR)-like superfamily protein	-	STOP GAINED	E->Stop
AT1G77320.1	Mutant is defective in meiosis and produces abnormal microspores.	MEIOSIS DEFECTIVE 1 (MEI1)	STOP GAINED	K->Stop
AT2G07680.1	Encodes ABCC13/MRP11, a member of the multidrug resistance associated protein MRP/ABCC subfamily, expression is induced by gibberellic acid.	ATP-BINDING CASSETTE C13 (ABCC13)	STOP GAINED	C->Stop
AT2G07777.1	ATP synthase 9 mitochondrial	-	STOP GAINED	W->Stop
AT2G16365.1	PCH1 binds and stabilizes the active (Pfr) form of phytochrome B and is involved in the formation of photobodies in the nucleus.	PHOTOPERIODIC CONTROL OF HYPOCOTYL 1 (PCH1)	STOP GAINED	K->Stop
AT2G22140.1	Forms a complex with MUS81 that functions as endonuclease in DNA recombination and repair processes.	ESSENTIAL MEIOTIC ENDONUCLEASE 1B (EME1B)	STOP GAINED	R->Stop
AT2G33160.1	Predicted exo-polygalacturonase gene	NIMNA	STOP GAINED	E->Stop
AT2G33270.1	Encodes a member of the thioredoxin family protein. Located in the chloroplast.	ATYPICAL CYS-HIS RICH THIOREDOXIN 3 (ACHT3)	STOP GAINED	C->Stop
AT2G43160.3	ENTH/VHS family protein	-	STOP Lost	Stop->L
AT3G12835.1	Unknown protein	-	STOP GAINED	Q->Stop
AT3G25610.1	Encodes aminophospholipid ATPase10 (ALA10), a P4-type ATPase flippase that internalizes exogenous phospholipids across the plasma membrane.	AMINOPHOSPHOLIPID ATPASE10 (ALA10)	STOP GAINED	S->Stop
AT3G26782.1	Tetratricopeptide repeat (TPR)-like superfamily protein	-	STOP GAINED	Y->Stop
AT3G61180.1	RING/U-box superfamily protein; BEST Arabidopsis thaliana protein match is: Zinc finger, C3HC4 type (RING finger) family protein (TAIR:AT4G11680.1)	-	STOP GAINED	G->Stop
AT3G02460.1	Ypt/Rab-GAP domain of gyp1p superfamily protein; BEST Arabidopsis thaliana protein match is: plant adhesion molecule 1 (TAIR:AT5G15930.1)	-	STOP GAINED	K->Stop
AT3G03950.3	Physically interacts with CIPK1. Located in the nucleus.	EVOLUTIONARILY CONSERVED C-TERMINAL REGION 1 (ECT1)	STOP GAINED	Y->Stop
AT3G05390.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	-	STOP GAINED	R->Stop
AT3G08580.1	mitochondrial ADP/ATP carrier	ADP/ATP CARRIER 1 (AAC1)	STOP GAINED	Q->Stop
AT3G09990.1	Nucleoside transporter family protein; BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT4G05120.1)	-	STOP GAINED	K->Stop

(*Table contd.*)

Gene Model Name	Gene Model Description	Primary Gene Symbol	SNP Effect	Amino acid change
AT3G11964.1	Encodes a nucleolar protein that is a ribosome biogenesis co-factor. Mutants display aberrant RNA processing and female gametophyte development.	RIBOSOMAL RNA PROCESSING 5 (RRP5)	STOP GAINED	K->Stop
AT3G17240.1	lipoamide dehydrogenase precursor	LIPOAMIDE DEHYDROGENASE 2 (mtLPD2)	STOP GAINED	K->Stop
AT3G19170.1	Zinc metalloprotease pitrilysin subfamily A. Signal peptide degrading enzyme targeted to mitochondria and chloroplasts. Expressed only in siliques and flowers	PRESEQUENCE PROTEASE 1 (PREP1)	STOP GAINED	R->Stop
AT4G10370.1	Cysteine/Histidine-rich C1 domain family protein involved in intracellular signaling pathway	-	STOP GAINED	S->Stop
AT4G18150.1	Kinase-related protein of unknown function (DUF1296)		STOP GAINED	R->Stop
AT4G28560.1	Encodes a member of a novel protein family that contains contain a CRIB (for Cdc42/Rac-interactive binding) motif required for their specific interaction with GTP-bound Rop1 (plant-specific Rho GTPase).	ROP-INTERACTIVE CRIB MOTIF-CONTAINING PROTEIN 7 (RIC7)	STOP GAINED	Y->Stop
AT4G28920.1	Protein of unknown function (DUF626)	-	STOP GAINED	L->Stop
AT4G36930.1	Encodes a transcription factor of the bHLH protein family. Mutants have abnormal, unfused carpels and reduced seed dormancy.	SPATULA (SPT)	STOP GAINED	Q->Stop
AT4G39060.1	BEST Arabidopsis thaliana protein match is: Galactose oxidase/kelch repeat superfamily protein (TAIR:AT4G19250.1)	-	STOP GAINED	K->Stop
AT4G39952.1	Pentatricopeptide repeat (PPR) superfamily protein	-	STOP GAINED	R->Stop
AT5G13890.1	Family of unknown function (DUF716)		STOP GAINED	C->Stop
AT5G46570.1	Encodes BR-signaling kinase 2 (BSK2), one of the three homologous BR-signaling kinases (BSK1, BSK2, BSK3,	BRASSINOSTEROID -SIGNALING KINASE 2 (BSK2)	STOP GAINED	K->Stop
AT5G56730.1	Insulinase (Peptidase family M16) protein	-	STOP GAINED	S->Stop
AT5G58200.2	Calcineurin-like metallo-phosphoesterase superfamily protein	-	STOP GAINED	K->Stop
AT5G58490.1	NAD(P)-binding Rossmann-fold superfamily protein	-	STOP GAINED	K->Stop

Chapter 5: Conclusions

The parallel design of our experiments identified a large portion of known genes and the processes they are involved, in various aspects of growth and development. The patterns of expression are consistent with previous studies and logical expectations, and provide support to the validity and accuracy of the RNA-Seq data. We were also able to identify a significant number of unknown gene loci and gene loci with putative and predicted functions. We demonstrated that many biological processes and genes involved in them are shared between developmental programs of male and female sex organs. Many of the DEGs that were upregulated in one sample compared to the other, represent hormone pathways, stem cell maintenance pathways, cell differentiation pathways, and cellular signaling pathways. Further study of the unknown gene loci will be required to identify these unknown loci and understand their physiological and molecular function in relation to the developmental programs of sexual organs. These genes may represent a host of potential targets for manipulation to re-engineer dioecious plant species to the ancestral hermaphroditic state, to improve our understanding of the intricate and complex gene regulatory networks underlying sex specification, and therefore, the evolution of sex chromosomes in land plants.

The generation of mutants specific to the female organ abortion, followed by a detailed process of mutant identification and characterization will be the key to the dissection of the gene regulatory network specific to the determination and development of the female organs, and help identification of genes other than the homeotic genes, possibly upstream or downstream

acting genes, cis- or trans-acting regulatory elements involved in the process. Having a repertoire of genes in these categories will eventually help us to be able to predict the sex determination network, and in the future, be able to engineer plants to the goal of having hermaphrodite populations. This will enable us to bypass or avoid the huge losses incurred by the agriculture community in dealing with the 1:1 segregation of male: female plants in dioecious crops. Hermaphrodite plants in populations have the advantage of being able to simultaneously bear fruit and still allow outcrossing (to maintain genetic diversity) when the mutations or sex-determining genes are on non-recombining chromosomes, and therefore minimizing the possibility of reversion to the state of dioecy.

Candidate genes identified through the current study therefore, present a potential for data mining and exploration of their functions specific to female sterility. We were successful in mapping EMS-induced SNPs in the FS322 line to the long arms of chromosome 1 and 3, suggesting that additional genes involved in the proper development of the female reproductive program are located on these chromosomes. Further study, involving larger pools of mutant plants followed by complementation analysis, is required to unequivocally associate genes in those regions with the phenotype and characterize their biological and physiological roles played in maintaining fertility of the female organs.

Our combined approach in studying floral sex determination has provided evidence towards the involvement of many genes that have not been directly associated with floral development. These genes include genes coding for proteins with the functional motifs of Ankyrin repeats, WD40 repeats, tri- tetra- and penta-tricopeptides. Given the intricate

involvement of the hormone auxin and the ABCE gene model in specifying the structure and functional morphology of the floral organs, and the analogy between leaf and floral development, our findings are indicative of the domain repeat containing proteins playing important roles downstream of both the ABCE genes and auxin hormone. Domain repeats in protein coding genes can serve many functions, such as the transmembrane domains of proteins in signaling cascades, in protein-recognition and protein-protein interactions, as anchor points for multimeric protein complexes, as binding sites for transcription factors, metal ion cofactors, and as chaperones guiding proper protein folding. WD40, Ankyrin and penta-tricopeptide repeat containing proteins have previously been found to be involved in a multitude of developmental processes, as well as defense responses, programmed cell death, protein kinase activity etc., and as such, provide a reference point to future research in deciphering the functional physiological roles played by these genes in floral development.

References

Afgan, E. et al., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(May), p.gkw343. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkw343.

Ainsworth, C., Parker, J. & Buchananwollaston, V., 1997. Sex Determination in Plants. *Current Topics in Developmental Biology*, 38, pp.167–223. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0070215308602471.

Akiyama, K. et al., 2014. RARGE II: An integrated phenotype database of arabidopsis mutant traits using a controlled vocabulary. *Plant and Cell Physiology*, 55(1), pp.1–10.

Arumuganathan, K. & Earle, E.D., 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9(3), pp.208–218.

Ashburner, M. et al., 2000. Gene Ontology: Tool for The Unification of Biology. *Nature Genetics*, 25(1), pp.25–29.

Benjamins, R. et al., 2001. The PINOID protein kinase regulates organ development in Arabidopsis by enhancing polar auxin transport. , 4067, pp.4057–4067.

Bennett, S.R.M. et al., 1995. Morphogenesis in pinoid mutants of Arabidopsis thaliana. , 8, pp.505–520.

Boualem, A. et al., 2009. A Conserved Ethylene Biosynthesis Enzyme Leads to Andromonoecy in Two Cucumis Species. *PLoS ONE*, 4(7), p.e6144. Available at: http://dx.plos.org/10.1371/journal.pone.0006144.

Boualem, A. et al., 2008. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science (New York, N.Y.)*, 321(5890), pp.836–838.

Byrne, M.E., 2011. Involvement of ribosomal protein RPL27a in meristem activity and organ development. , 6(5), pp.712–714.

Cecchetti, V. et al., 2008. Auxin Regulates Arabidopsis Anther Dehiscence, Pollen Maturation, and Filament Elongation. *the Plant Cell Online*, 20(7), pp.1760–1774. Available at: http://www.plantcell.org/cgi/doi/10.1105/tpc.107.057570.

Charlesworth, B., 1996. The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, 6(2), pp.149–162. Available at: http://www.sciencedirect.com/science/article/pii/S0960982202004487.

Charlesworth, B. & Charlesworth, D., 1978. Model for Evolution of Dioecy and Gynodioecy. *American Naturalist*, 112(988), pp.975–997.

Chen, X. & Meyerowitz, E.M., 1999. HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway. *Molecular Cell*, 3(3), pp.349–360.

Cheng, H. et al., 2004. Gibberellin regulates Arabidopsis floral development via suppression of DELLA protein function. *Development*, 131(5), pp.1055–1064. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14973286.

Cheng, Y. et al., 2008. NPY genes and AGC kinases define two key steps in auxin-mediated organogenesis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52), pp.21017–21022.

Cheng, Y. & Chen, X., 2004. Posttranscriptional control of plant development. *Current Opinion in Plant Biology*, 7(1), pp.20–25. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1369526603001365.

Cheng, Y., Dai, X. & Zhao, Y., 2006. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. , pp.1790–1799.

Cheng, Y., Dai, X. & Zhao, Y., 2006. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes Dev*, 20(13), pp.1790–1799. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16818609.

Cheng, Y., Dai, X. & Zhao, Y., 2007. Auxin Synthesized by the YUCCA Flavin Monooxygenases Is Essential for Embryogenesis and Leaf Formation in Arabidopsis. , 19(August), pp.2430–2439.

Cheng, Y. & Zhao, Y., 2007. A Role for Auxin in Flower Development. , 49(1), pp.99–104.

Christensen, S.K. et al., 2000. Regulation of Auxin Response by the Protein Kinase PINOID. , 100, pp.469–478.

Coen, E.S. & Meyerowitz, E.M., 1991. The war of the whorls: genetic interactions controlling flower development. *Nature*, 353(6339), pp.31–37.

Conesa, A. & Stefan, G., 2012. Blast2GO Teaching Exercises.

Danecek, P. et al., 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15), pp.2156–2158.

Dellaporta, S.L. & Calderon-urrea, A., 1994. The Sex Determining Process in Maize. *Science*, 266(December), pp.1501–1505.

Devis, D. et al., 2015. Dosage Sensitivity of RPL9 and Concerted Evolution of Ribosomal Protein Genes in Plants. , 6(December), pp.1–12.

Ellis, C.M. et al., 2005. AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2regulate senescence and floral organ abscission in Arabidopsisthaliana. *Development*, 132(20), pp.4563–4574. Available at: http://dev.biologists.org/cgi/doi/10.1242/dev.02012.

Friml, J. et al., 2003. Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. *Nature*, 426(6963), pp.147–153.

Fujioka, S. et al., 1988. Qualitative and Quantitative Analyses of Gibberellins in. *Plant physiology*, (88), pp.1367–1372.

Furutani, M. et al., 2004. PIN-FORMED1 and PINOID regulate boundary formation and cotyledon development in Arabidopsis embryogenesis. *Development (Cambridge, England)*, 131(20), pp.5021–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15371311.

Garcia, V. et al., 2016. Rapid identification of causal mutations in tomato EMS populations via mapping-by-sequencing. , 11(12), pp.2401–2418.

Golenberg, E.M. & West, N.W., 2013. Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. *American Journal of Botany*, 100(6), pp.1022–1037.

Gschwend, A.R. et al., 2012. From the Cover: Rapid divergence and expansion of the X chromosome in papaya. *Proceedings of the National Academy of Sciences*, 109(34), pp.13716–13721.

Han, J. et al., 2014. The effects of gibberellic acid on sex expression and secondary sexual characteristics in papaya. *HortScience*, 49(3), pp.378–383.

Hartwig, T. et al., 2011. Brassinosteroid control of sex determination in maize. *Proceedings of the National Academy of Sciences*, 108(49), pp.19814–19819.

Hawkins, C. & Liu, Z., 2014. A model for an early role of auxin in Arabidopsis gynoecium morphogenesis. *Frontiers in Plant Science*, 5(July), pp.1–12. Available at: http://journal.frontiersin.org/article/10.3389/fpls.2014.00327/abstract.

Hu, J. et al., 2008. Potential sites of bioactive gibberellin production during reproductive growth in Arabidopsis. *The Plant cell*, 20(2), pp.320–36. Available at: http://www.plantcell.org/content/20/2/320.short.

Irish, V.F., 2006. Duplication, Diversification, and Comparative Genetics of Angiosperm MADS-Box Genes. *Advances in Botanical Research*, 44(6), pp.129–161.

Ito, T. et al., 2007. The Homeotic Protein AGAMOUS Controls Late Stamen Development by Regulating a Jasmonate Biosynthetic Gene in Arabidopsis. *the Plant Cell Online*, 19(11), pp.3516–3529. Available at: http://www.plantcell.org/cgi/doi/10.1105/tpc.107.055467.

Ito, T., Kim, G. & Shinozaki, K., 2000. Disruption of an Arabidopsis cytoplasmic ribosomal protein

S13-homologous gene by transposon-mediated mutagenesis causes aberrant growth and development. , 22.

Jia, Y. et al., 2015. The Arabidopsis thaliana elongator complex subunit 2 epigenetically affects root development. , 66(15), pp.4631–4642.

Krizek, B. a & Fletcher, J.C., 2005. Molecular mechanisms of flower development: an armchair guide. *Nature reviews. Genetics*, 6(9), pp.688–698.

Krogan, N.T., Hogan, K. & Long, J.A., 2012. APETALA2 negatively regulates multiple floral organ identity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. , 4190, pp.4180–4190.

Lampugnani, E.R., Kilinc, A. & Smyth, D.R., 2013a. Auxin controls petal initiation in Arabidopsis. , 194, pp.185–194.

Lampugnani, E.R., Kilinc, A. & Smyth, D.R., 2013b. Auxin controls petal initiation in Arabidopsis. *Development (Cambridge, England)*, 140(1), pp.185–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23175631.

Leshchiner, I. et al., 2012. Mutation mapping and identification by whole-genome sequencing. *Genome Research*, 22(8), pp.1541–1548.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), pp.1754–1760.

Li, J. et al., 2016. *RABBIT EARS* regulates the transcription of *TCP4* during petal development in Arabidopsis. *Journal of Experimental Botany*, 67(22), p.erw419. Available at: http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/erw419.

Li, L. et al., 2004. The Tomato Homolog of CORONATINE-INSENSITIVE1 Is Required for the Maternal Control of Seed Maturation , Jasmonate-Signaled Defense Responses , and Glandular Trichome Development. *The Plant cell*, 16(January), pp.126–143. Available at: www.plantcell.org/cgi/doi/10.1105/tpc.017954.

Lindner, H. et al., 2012. SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing. *Genetics*, 191(4), pp.1381–1386. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3416015&tool=pmcentrez&rendertype=abstract.

Lindsay, D.L., Sawhney, V.K. & Bonham-Smith, P.C., 2006. Cytokinin-induced changes in CLAVATA1 and WUSCHEL expression temporally coincide with altered floral development in Arabidopsis. *Plant Science*, 170(6), pp.1111–1117. Available at: http://linkinghub.elsevier.com/retrieve/pii/S016894520600032X.

Liu, X. et al., 2010. The role of floral organs in carpels, an Arabidopsis loss-of-function mutation in MicroRNA160a, in organogenesis and the mechanism regulating its expression. *Plant Journal*, 62(3), pp.416–428.

Lloyd, J. & Meinke, D., 2012. A Comprehensive Dataset of Genes with a Loss-of-Function Mutant Phenotype in Arabidopsis. *Plant Physiology*, 158(3), pp.1115–1129.

Mandaokar, A. et al., 2006. Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling. *Plant Journal*, 46(6), pp.984–1008.

Maple, J. & Møller, S.G., 2007. Mutagenesis in Arabidopsis. *Methods in molecular biology (Clifton, N.J.)*, 362, pp.197–206.

Martin, A. et al., 2009. A transposon-induced epigenetic change leads to sex determination in melon. *Nature*, 461(7267), pp.1135–1138. Available at: http://www.nature.com/doifinder/10.1038/nature08498.

De Martinis, D. & Mariani, C., 1999. Silencing Gene Expression of the Ethylene-Forming Enzyme Results in a Reversible Inhibition of Ovule Development in Transgenic Tobacco Plants. *The Plant Cell*, 11(6), pp.1061–1072. Available at: http://www.plantcell.org/content/11/6/1061.abstract.

Mattsson, J., Ckurshumova, W. & Berleth, T., 2003. Auxin Signaling in Arabidopsis Leaf Vascular Development 1. , 131(March), pp.1327–1339.

Mina, S.S.Æ. & Hitoshi, T.Æ., 2009. Auxin – cytokinin interactions in the control of shoot branching. , (October 2008), pp.429–435.

Ming, R. et al., 2008. The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). , 452(April).

Ming, R., Bendahmane, A. & Renner, S.S., 2011. Sex Chromosomes in Land Plants. *Annual Review of Plant Biology*, 62(1), pp.485–514. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-arplant-042110-103914 [Accessed October 1, 2015].

Missbach, S. et al., 2013. 40S Ribosome Biogenesis Co-Factors Are Essential for Gametophyte and Embryo Development. , 8(1), pp.1–19.

Mudunkothge, J.S. & Krizek, B.A., 2012. Three Arabidopsis AIL/PLT genes act in combination to regulate shoot apical meristem function. *Plant Journal*, 71(1), pp.108–121.

Nagpal, P. et al., 2005. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development (Cambridge, England)*, 132(18), pp.4107–4118. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16107481.

O'Maoileidigh, D.S., Graciet, E. & Wellmer, F., 2014. *Genetic Control of Arabidopsis Flower Development*, Elsevier. Available at: http://linkinghub.elsevier.com/retrieve/pii/B9780124171626000067.

Ó'Maoiléidigh, D.S., Graciet, E. & Wellmer, F., 2014. Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytologist*, 201(1), pp.16–30. Available at: http://doi.wiley.com/10.1111/nph.12444 [Accessed May 13, 2015].

Okada, K. et al., 1991. Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. *The Plant cell*, 3(7), pp.677–684. Available at: http://www.plantcell.org/content/3/7/677.full.pdf+html.

ÓMaoiléidigh, D.S. et al., 2013. Control of reproductive floral organ identity specification in Arabidopsis by the C function regulator AGAMOUS. *The Plant cell*, 25(7), pp.2482–503. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3753378&tool=pmcentrez&rendertype=abstract.

Open, D.M.C.W. et al., 2014. Ribosomal Protein RPL27a Promotes Female Gametophyte Development in a. , 165(July), pp.1133–1143.

Page, D.R. & Grossniklaus, U., 2002. The art and design of genetic screens: Arabidopsis thaliana. *Nature reviews. Genetics*, 3(2), pp.124–136.

Park, J.H. et al., 2002. A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in Arabidopsis due to a block in jasmonic acid biosynthesis. *Plant Journal*, 31(1), pp.1–12.

Pelaz, S. et al., 2000. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, 405(6783), pp.200–203.

Pinon, V. et al., 2008. Three PIGGYBACK genes that specifically influence leaf patterning encode ribosomal proteins. , 1324, pp.1315–1324.

Renner, S.S., 2014. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*, 101(10), pp.1588–1596. Available at: http://www.amjbot.org/cgi/doi/10.3732/ajb.1400196.

Sessions, R. a & Zambryski, P.C., 1995. Arabidopsis gynoecium structure in the wild and in ettin mutants. *Development (Cambridge, England)*, 121(5), pp.1519–1532.

Sherif, S. et al., 2009. Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. , 60(3), pp.907–922.

Signaling, A. et al., 2013. Maize LAZY1 Mediates Shoot Gravitropism and In fl orescence Development through Regulating Auxin. , 163(November), pp.1306–1322.

Skinner, D.J. et al., 2001. The Arabidopsis HUELLENLOS Gene , Which Is Essential for Normal Ovule Development , Encodes a Mitochondrial Ribosomal Protein. , 13(December), pp.2719–2730.

Smyth, D.R., Bowman, J.L. & Meyerowitz, E.M., 1990. Early flower development in Arabidopsis. *The Plant cell*, 2(8), pp.755–767.

Somerville, C. & Koornneef, M., 2002. A fortunate choice: the history of Arabidopsis as a model plant. *Nature reviews. Genetics*, 3(11), pp.883–889.

Stewart, D., Graciet, E. & Wellmer, F., 2016. Molecular and regulatory mechanisms controlling floral organ development. *FEBS Journal*, 283, p.n/a-n/a. Available at: http://doi.wiley.com/10.1111/febs.13640%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26725470.

Stilio, S. Di et al., 2011. Multiple developmental processes underlie sex differentiation in angiosperms. , 27(9).

Stintzi, A. & Browse, J., 2000. The Arabidopsis male-sterile mutant, opr3, lacks the 12-oxophytodienoic acid reductase required for jasmonate synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 97(19), pp.10625–10630. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10973494%5Cnhttp://www.pnas.org/content/97/19/10625.full.pdf.

Stirnberg, P. et al., 2012. Mutation of the cytosolic ribosomal protein-encoding RPS10B gene affects shoot meristematic function in Arabidopsis. , pp.1–20.

Szakonyi, D. & Byrne, M.E., 2017. a n d e s i o s c i e n c e o n o t d i s t r i b u t e. , 2324(March).

Tan, X. et al., 2007. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. , 20, pp.1–20.

Thomas, P.D. et al., 2003. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, 13(9), pp.2129–2141. Available at: http://www.genome.org/cgi/doi/10.1101/gr.772403.

Trapnell, C. et al., 2012. Articles Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31(1), pp.46–53. Available at: http://dx.doi.org/10.1038/nbt.2450.

Trapnell, C. et al., 2014. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.

Trigueros, M. et al., 2009. The NGATHA genes direct style development in the Arabidopsis gynoecium. *The Plant cell*, 21(5), pp.1394–1409.

Wang, J. et al., 2012. Sequencing papaya X and Y h chromosomes reveals molecular basis of

incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences*, 109(34), pp.13710–13715.

Wei, N. & Deng, X., 2003. The COP9 signalosome. *Annu Rev Cell Dev Biol*, 19, pp.261–286. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14570571%5Cnpapers3://publication/doi/10.1146/annurev.cellbio.19.111301.112449.

Wellmer, F. et al., 2004. Genome-wide analysis of spatial gene expression in Arabidopsis flowers. *The Plant cell*, 16(5), pp.1314–26. Available at: http://www.plantcell.org/content/16/5/1314.short [Accessed October 1, 2015].

Wikström, N., Savolainen, V. & Chase, M.W., 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings. Biological sciences / The Royal Society*, 268(1482), pp.2211–20. Available at: http://rspb.royalsocietypublishing.org/content/268/1482/2211.

Wu, M.F., Tian, Q. & Reed, J.W., 2006. Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development*, 133(21), pp.4211–4218. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17021043.

Xing, S. et al., 2013. SPL8 and miR156-targeted SPL genes redundantly regulate Arabidopsis gynoecium differential patterning. *Plant Journal*, 75(4), pp.566–577.

Ye, Q. et al., 2010. Brassinosteroids control male fertility by regulating the expression of key genes involved in Arabidopsis anther and pollen development. *Proceedings of the National Academy of Sciences*, 107(13), pp.6100–6105. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0912333107.

Zahn, L.M. et al., 2005. The evolution of the SEPALLATA subfamily of MADS-box genes: A preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics*, 169(4), pp.2209–2223.

Zahn, L.M., Feng, B. & Ma, H., 2006. Beyond the ABC-Model: Regulation of Floral Homeotic Genes. *Advances in Botanical Research*, 44(6), pp.163–207.

Zhang, X. et al., 2005. Genome-wide expression profiling and identification of gene activities during early flower development in Arabidopsis. *Plant molecular biology*, 58(3), pp.401–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16021403.

Zik, M. & Irish, V.F., 2003. Global Identification of Target Genes Regulated by APETALA3 and PISTILLATA Floral Homeotic Gene Action. , 15(January), pp.207–222.