Comparative Promoter Analysis of APOLLO for Identification of Regulatory Networks Involved in Apomixis

A Thesis Submitted to the College of
Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Plant Sciences
University of Saskatchewan
Saskatoon

By
Tyler Lefley
Winter 2020

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9

Canada

Abstract

The genus *Boechera* is a complex case study of reproductive plant biology stemming from its ability to produce seed sexually and asexually (apomixis). Previous work accomplished by the Sharbel Laboratory discovered an exclusive differentially expressed gene during early megasporogenesis with a 20nt indel differentiating two alleles of a gene called APOLLO. APOLLO is a gene encoding a DEDDH 5'-3' exonuclease with extensive copy number variation of both alleles throughout the genus, but what makes APOLLO such an important candidate is that only one allele (sex allele) is found in sexual *Boechera* while apomictic *Boechera* have both alleles (sex and apo allele). To understand the cause of differential expression between the sex and apo allele of APOLLO, the promoter has been called into question. Unlike genes that present a conserved exon-intron structure, promoters are non-coding DNA elements with no distinct organizational patterns in relation to the motifs that attract transcriptional machinery, and therefore the understanding of promoter evolution is unresolved.

During gene duplication events it is assumed that the promoter is a single unit with the gene, but not all duplications are entitled to that degree of certainty because there is plenty of evidence for the formation of chimeric and fusion genes. What happens when a gene is duplicated but its promoter is left behind and how does it affect gene expression? This is in fact the case of the APOLLO apo allele. By sequencing 1kb upstream of a conserved area within the first exon of APOLLO a compelling case has been put together showing that a novel promoter is highly conserved, apoallele specific, and preliminarily responsible for the differential expression of APOLLO in pre-meiotic ovules of sexual and apomictic *Boechera*

Acknowledgements

It has been quite a journey these past few years and I have many people to be thankful for in my development as a plant scientist. None of this would have even started if it were not for the encouragement from Dr. Graham Scoles. Graham was the central node for connecting me with members of his barley laboratory, Dr. Helen Booker, Dr. Axel Diedrichsen, and ultimately my current supervisor Dr. Tim Sharbel during the last year of my bachelor's degree. Without the resources he made available for my growing interest into apomixis, I would not have met these amazing people.

Tim has been nothing short of incredible over my past four years working for him. As I have gained more experience interacting with other scientists / researchers, I noticed how rare it is in academia to have a boss as supportive and instrumental in giving advice that is applicable to maintaining personal goals and the work-life balance. GIFS has been a developing organization over the past few years and with a continuously growing staff comes increased expertise to aid in the design and troubleshooting of experiments. I would like to recognize the many post-doctoral researchers and lab technicians; Martin Mau, Xingliang Ma, Joanne Ernest, Dorota Paczesniak, Marco Pellino, Tiina Liiving, Rick Goertzen, and Angie Li for their knowledgeable contributions through problem solving/ technical support.

Lastly, I have to thank my family for their support. My friends were all pretty shocked when I enrolled in graduate school but my parents were the ones that knew I could do it. Now it feels even better saying that I completed the program.

Table of Contents

Permission to Use i

Abstract ii

Acknowledgements iii

Table of Contents iv

List of Tables vi

List of Figures ix

List of Abbreviations xi

1 Introduction 1

1.1 Apomixis 4

1.2 Apomixis and Agriculture – A Tool for Fixing Heterosis 8

2 Literature Review 10

2.1 Boechera - A Model System for the Evolution and Functional Genetics of Apomixis 10

2.2 Apomictic Candidate Genes 15

3 Materials and Methods 18

3.1 Apomixis Candidate Gene Identification 18

3.2 Microarray Analysis 18

3.3 DNA Sequence Annotation 19

4 Results 22

4.1 Compilation of Apomixis Candidate Genes 22

4.2 Microarray Analysis 31

4.3 Identification of APOLLO 32

4.4 Presence and Absence of Probes in Sexual and Apomictic Genomes 42

4.5 Copy Number Variation of Secondary Candidate Genes 46

4.6 Is Gene Regulation Influenced by Copy Number Variation? 47

4.7 Preliminary Identification of a Gene only Found in Boechera 47

5 Discussion 49

5.1 Current Methodology for Investigating Apomixis 49

5.2 Apomixis is Multidimensional 50

5.3 Emergence of BABY BOOM as a Significant Apomictic Factor for Asexuality in Field Crops 52

5.4 AT2G02090, a Locus Important to DNA Repair and Recombination Identified in Sexual *B. stricta* was not Identified in Apomictic *B. divaricarpa* 54

5.5 Is Downregulation of Ribosomal Protein Genes Linked to Upregulated Expression of a Novel Nucleolar Protein Family? 56

5.6 Downregulation of Plant Defensin Proteins in Apomictic Boechera 56

5.7 Copy Number Variation for Secondary Candidate Genes between Apomictic and Sexual Whole Genome Assemblies 57

6 Conclusion 59

Transition 60

7 *In Silico* Promoter Analysis 61
7.1 Introduction 61
8 Literature Review 63
8.1 Non-Coding DNA Evolution 63
8.2 Promoter Evolution 65
8.3 The Meselson Effect and Gene Duplication Influence on Promoter evolution 67
9 Materials and Methods 69
9.1 Accession Selection 69
9.2 DNA Extractions and Quality Testing 70
9.3 Polymerase Chain Reaction (PCR) 71
9.4 Cloning and Sequencing 73
9.5 Promoter Analysis 76
10 Results 77
10.1 Assessment of genetic diversity between the apo and sex allele promoters 78
10.2 *In Silico* Promoter Analysis 84
10.3 The Isolated Apo Allele Promoter is Novel 87
11 Discussion 88
11.1 Is there more than one Novel Promoter of APOLLO? 88
11.2 Promoter Motifs of Interest 89
12 Conclusion 92
13 Thesis Conclusion 93
References 94

List of Tables

Table 2.1 Descriptions of seed development processes occurring in each example illustrated in Figure 2.1 13

Table 3.1 Databases used for annotation 21

Table 4.1 Spreadsheet of documented candidate genes for apomixis with their corresponding species of interest and gene ID's in other important crop plants for phylogenic interest. Literature background for all the genes listed here are in Table 4.2. In each gene's corresponding literature, *Arabidopsis thaliana* is highly referenced making it possible for retrieval of the protein ID through TAIR. The protein ID's in this table are the means of identifying orthologous gene ID's in other plants via InParanoid8..... 23

Table 4.2 Apomixis candidate genes identified through published literature including to which stage of apomictic development they are relevant. The majority of genes identified come from sexual backgrounds, but it is indicated if a gene is associated with an apomictic case study. 24

Table 4.3 Microarray probe sequences of secondary candidate genes that were statistically isolated to be differentially expressed between sexual and apomictic megaspore mother cells. 31

Table 4.4 Differential expression statistics for top 20 candidate probes. LogFC is the fold change statistic with positive values indicating upregulation and negative being downregulated in apomictic ovules. AveExpr is the mean log2 expression of a probe across all 18 arrays. t is the value derived from a moderated t-test statistic. The adjusted P value is the P value adjusted to control the false discovery rate across multiple samples via background correction. 32

Table 4.5 BLASTn of candidate probes to a sexual *B. stricta* WGS assembly..... 33

Table 4.6 BLASTn of candidate probes to apomictic *B. divaricarpa* WGS assembly 34

Table 4.7 BLASTn results of candidate *B. stricta* CDS sequences submitted to TAIR10 CDS database..... 35

Table 4.8 BLASTp results of candidate *B. stricta* protein sequences submitted to TAIR10 protein database..... 35

Table 4.9 *Arabidopsis thaliana* loci and their corresponding transcript, protein, and CDS candidate sequences from *B. stricta*..... 36

Table 4.10 Identified *Arabidopsis orthologues* of *Boechera* cDNA containing microarray probe alignment(s). 37

Table 4.11 BLASTn of candidate probes to *B. stricta* transcript database..... 40

Table 4.12 BLASTn results of candidate *B. stricta* transcripts submitted to TAIR10 transcript database 41

Table 4.13 BLASTn of candidate probes to *B. stricta* CDS database 43

Table 4.14 tBLASTx of candidate probes to *B. stricta* protein database 43

Table 4.15 BLASTn results of candidate sequences from sexual *B. stricta* WGS submitted to NCBI nucleotide database with the taxonomic filter limiting results to Virdiplantae 44

Table 4.16 BLASTn results from candidate sequences from apomictic *B. divaricarpa* WGS submitted to NCBI nucleotide database with the taxonomic filter limiting results to Virdiplantae 45

Table 9.1. *Boechera* germplasm used in study. IPK_ID = Accession ID used at IPK Gatersleben with the first two letters being initials of the collector. MOR = Mode of Reproduction..... 69

Table 9.2 Primers from two different publications had been previously experimentally verified for use in amplification of the *Boechera* ITS region (Mummenhoff et al. 1997 provided primer 18F and White et al. 1990 for primer 25R). 70

Table 9.3 PCR protocol for amplification of ITS DNA using primers in Table 9.2 71

Table 9.4 BAC sequences containing APOLLO were initially isolated in Corral et al. 2013 and characterized to contain either the sex or the apo allele. Sequences can be retrieved using the NCBI accession number. 71

Table 9.5 Primers designed from the BAC sequences in Table 9.4 to amplify the promoter of the apo and sex allele of APOLLO. TSP1R is from Corral et al. 2013 and is a universal reverse primer located within the first exon of APOLLO that is used to screen for the 5' UTR indel that is the distinguishing feature between apo and sex alleles. 71

Table 9.6 Expected size of PCR product for each primer pair. Estimates are based on BAC sequences containing APOLLO in Table 9.4 72

Table 9.7 PCR touchdown protocol for TSP1R and pAPOLLO Universal F-1kb primers using KOD HotStart Polymerase (TOYOBO®) which has proofreading capability qualifying it for use with amplicons that need to be sequenced. 72

Table 9.8 PCR protocol used for amplifying the apo allele promoter using Phusion Hot Start II High Fidelity PCR Master Mix. 72

Table 9.9 Colony PCR protocol used for screening DH5α transformed liquid cultures. 2x Taq Master Mix from Frogga Bio™ was used with pJET 1.2 primers in Table 9.10 to amplify all cloned promoter inserts (apo and sex). 73

Table 9.10 pJET primers used for sequencing and colony PCR. Primers come with the pJET 1.2 blunt end-cloning vector provided by ThermoFisher Scientific™. 74

Table 10.1 Analysis of all sequences per accession measuring genetic diversity of the APOLLO sex allele promoter. HD = a measure of haplotype diversity; k = the average number on nucleotide differences; pi = nucleotide diversity per site; pi(JC) = pi with Jukes-Cantor correction. 79

Table 10.2 . SNP Table of apo allele promoters. Location is based on its position upstream of the start codon of APOLLO (right to left). The total number of sequences aligned from the three amplified accessions (CJ73, ES805, and TS290) is 50, which is also the denominator in calculating the variant frequency..... 80

Table 10.3 SNP table based on the alignment of 120 sex allele promoter sequences from all accessions in Table 9.1. Location refers to the genetic position upstream of the APOLLO start codon (right to left). 81

Table 10.4 Analysis of transcription factor binding site (TFBS) in the APOLLO sex allele using MatInspector. A consensus sequence was created using a 50% threshold to eliminate ambiguous nucleotides from the alignment of all 47 identified sex allele haplotypes. Matrix similarities were filtered for results only containing P<0.05 due to the small size of TFBS'. Sequence orientation is left to right with the APOLLO start codon located at 1058bp..... 85

Table 10.5 Analysis of TFBS in the APOLLO apo allele using MatInspector. A consensus sequence from all apo alleles being aligned was created with a 50% threshold therefore eliminating ambiguous nucleotides, which are problematic for TF binding domain identification. Sequence orientation is left to right with the APOLLO start codon located at 1103bp. Matrix similarity was filtered for P<0.05 85

Table 10.6 Transcription factor overrepresentation analysis of the apo allele promoter of APOLLO 86

Table 10.7 Using MatInspector, the promoter of AT1G74390 was analyzed for TFBS with a matrix similarity score filter of p<0.05 87

List of Figures

Figure 1.1 Cellular structure of a mature polygonum type embryo and pollen cell post megasporogenesis and microsporogenesis. Double fertilization relates to both the egg cell and central cell being fertilized by the two sperm cells contained within a single pollen. The 2:3 ratio of embryo: endorpserm is indicative of a diploid zygote. 1

Figure 1.2 The megaspore mother cell develops into the polygonum type embryo sac (8) through a number of steps: the megaspore mother cell must undergo meiotic reduction (1, 2, and 3) resulting in four haploid cells (3); out of those four cells, only one survives (4) and continues through three rounds of mitosis producing eight cells (5, 6, and 7). These seven cells are organized to create an architecture consisting of three antipodals (orange), two central cells (red), and two synergids (green) that flank the egg cell (blue). 5

Figure 1.3 Pollen microspores originate from the pollen sac located within the anther that is supported by a filament. Microsporogenesis begins with a single celled microsporocyte (1) that undergoes meiosis producing a tetrad of reduced microspores (3). As microgametogenesis proceeds a large vacuole forms, pushing the nucleus against the cell wall before mitosis I (4 and 5), which compartmentalizes the generative nucleus from the vegetative nucleus (6). Additionally, during formation of bicellular pollen, the vacuole decreases in size and smaller vesicles are observed (7). During mitosis II of microgametogenesis the generative nucleus doubles (8), resulting in a mature male gametophyte that is capable of double fertilization (embryo and central cell of the female gametophyte)..... 6

Figure 1.4 This figure compares the different pathways of seed development in gametophytic and sporophytic apomixis. Apospory and diplospory are both classified as gametophytic apomixis because the mature embryo in both instances is the same, but achieved through different pathways. Diplospory sees the failure of meiosis leading to no reduction of the megaspore. The unreduced megaspore then becomes source material for the initiation of embryo sac development. In aposporic apomixis, the megaspore mother cell undergoes meiotic reduction but all four reduced cells degenerate and are replaced by a single aposporous initial cell. The aposporous initial cell originates from the nucellus or integument tissue that surrounds the megaspore mother cell. This aposporous initial cell develops mitotically into a mature embryo with the same ploidy as the parent. Sporophytic apomixis differs from gametophytic apomixis in that the megaspore mother cell follows normal sexual development resulting in a reduced mature embryo. What makes this apomictic is the parthenogenic development of additional ovaries from nucellus or integument tissues that can accompany the sexually derived embryo. The somatic ovules are genetic clones of the mother plant. 7

Figure 2.1 Seed formation in *Boechera* is characterized by embryos and endosperm of varying ploidies based on i) fertilization of the egg and ii) central cells. 1. Both central cell and egg are fertilized by haploid sperm cells; 2. No fertilization of the unreduced egg cell occurs with autonomous endosperm formation; 3. Egg cell is unfertilized but haploid sperm cell contributes to the central cell; 4. Both haploid sperm cells contribute to create a triploid embryo and a 5n central cell; 5. Unreduced diploid or reduced tetraploid sperm cells double fertilize unreduced diploid

embryo; 6. Only the central cell is fertilized by a diploid sperm cell; 7 and 8. Triploids are fertile and have triploid pollen, with both autonomous and pseudogamous endosperm formation..... 12

Figure 9.1 Using the ITS PCR protocol in Table 9.3, spacer DNA was amplified from 1µl of DNA solution from all accessions in Table 21 using 2x Taq FroggaMix polymerase and primers in Table 9.2. 10µl of PCR product was run on 0.8% agarose gel for 40 minutes at 100kv with a Bioline HyperLadder™ 1kb. 70

Figure 9.2 Colony PCR results for both the sex allele (A) and apo allele (B) on 0.8% agarose gel with a Bioline HyperLadder™ 1 kb. Faint bands above the 1kb positive results (i) is the plasmid backbone (approx. 2kb) and the smaller products (ii) are off target amplification that were sequenced and found to be the result of the TSP1R acting as both the forward and reverse primer for this small piece of DNA. The sex allele colonies in (A) are from accession JL100 and apo allele colonies (B) are from accession TS290. 74

Figure 9.3 The Sanger sequencing method used in this study dye terminator sequencing which uses fluorescent dyes to label each of the four nucleic acids. Sanger sequencing is a PCR reaction requiring a 5'-3' directional primer to anchor as the starting point on the DNA template. Sanger sequencing consistently provides sequencing reads upwards of >600bp in one direction until quality drops off and the chromatograph becomes unusable (A and B). Because the Apollo promoter is approximately 1.1kb in size, sequencing had to be done in both directions to create an overlap region in the middle that would act as a bridge (C). The alignment of both sequencing reads exemplified in A and C completely covers the 1.1kb region with high quality chromatograph peaks (D). 75

Figure 10.1 APOLLO 5' UTR alignment 44bp upstream of the start codon for exon 1 of all sequenced clones for accession TS290 (*B. retrofracta*). Within the black rectangle is the conserved 20bp indel that characterizes apo alleles from sex alleles as highlighted in Corral et al. 2013.... 77
Figure 10.2 Using the TAIR genome browser to capture the size of the promoter for AT1G74390 (APOLLO's orthologue) it became evident that the 200bp region (between green arrows) must promote transcription bi-directionally because transcription of AT1G7440 occurs in the opposite direction. 78

Figure 10.3 Sex promoter SNP alignment showing the placement of all SNP's in Table 10.3 across all 120 cloned sequences. Red bar on the right represents the start codon for APOLLO. At 80bp, the gap representing an indel is another poly repeat region causing sequencing errors. All black vertical lines represent SNP's. 82

Figure 10.4 Apo promoter SNP alignment visualizing the SNP's in Table 10.2 across all the sequenced clones. The red bar on the right represents the start codon of APOLLO. White gaps are locations of indels, but as referred to in Fig. 10.3 they are the locations of long poly repeats causing sequencing difficulties. Black vertical lines are the SNP's. 83

List of Abbreviations

APOLLO – Apomixis Linked Locus

ASGR – Asexual Genomic Region

BAC – Bacteria Artificial Chromosome

BBM – Baby Boom

BLAST – Basic Local Alignment Search Tool

cDNA – Complimentary DNA

CDS – Coding Sequence

CNV – Copy Number Variation

CRISPR – Clustered Regularly Insterspaced Palindromic Repeats

DNA – Deoxyribonucleic Acid

ITS – Internal Transcribed Spacers

MiMe – Mitosis instead of Meiosis

MMC – Megaspore Mother Cell

MOR – Mode of Reproduction

mRNA – Messenger RNA

NCBI – National Center for Biotechnology Information

ncDNA – Non Coding DNA

OSD1 – Ommission of Second Division 1

PCR – Polymerase Chain Reaction

PDF – Plant Defensin Family

RNA – Ribonucleic Acid

SNP – Single Nucleotide Polymorphism

TAIR – The Arabidopsis Information Resource

TALENS – Transcription Activator-Like Effector Nuclease

TFBS – Transcription Factor Binding Site

UPGRADE2 – Unreduced Pollen Grain Development

UTR – Untranslated Region

WGS – Whole Genome Sequence

ZFN – Zinc Finger Nuclease

1 Introduction

Reproduction occurs through sexual and asexual means, and there are trade-offs to either mode of reproduction which have implications for fitness and survival. Evolution is blind in that living things survive or die, and what matters most is the ability of an individual to successfully produce offspring (i.e. fitness). In plants, sexual reproduction is accomplished through the fusion of male and female gametes (or mating types) and their meiotically-derived genetic variation (Figure 1.1), while asexual reproduction is accomplished through budding, vegetative propagation, fragmentation, and *apomixis* (the asexual reproduction of seeds; *parthenogenesis* in animals).

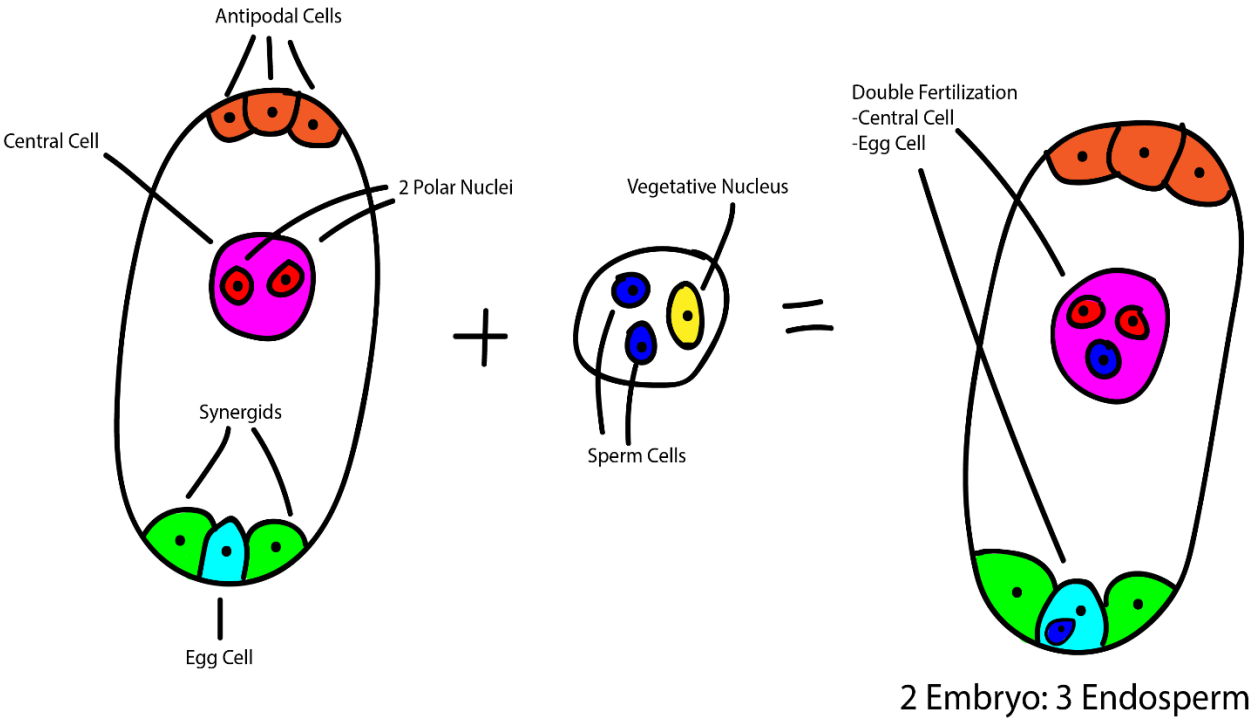

Figure 1.1 Cellular structure of a mature polygonum type embryo and pollen cell post megasporogenesis and microsporogenesis. Double fertilization relates to both the egg cell and central cell being fertilized by the two sperm cells contained within a single pollen. The 2:3 ratio of embryo: endorpserm is indicative of a diploid zygote.

The evolution of sexuality and asexuality each possess their own advantages and disadvantages in terms of reproduction and survival.

Sex is advantageous in that it creates genetic variation upon which natural selection can act, leading to adaptation to biotic or abiotic stresses (Otto, 2009). Mutations occur randomly in a genome and have implications on its fitness. Most genetic mutations can be classified as neutral, slightly-deleterious or deleterious (Keightley and Otto, 2006), and as such impart no positive effect to an organism. However, a small percentage of mutations have positive effects on fitness. Sex and meiotic recombination enable adaptation whereby positive mutations can be recombined into the same individual, leading to significant fitness increase (Bürger, 1999). In the same light, deleterious mutations can segregate into single offspring, with the associated fitness loss (i.e. death) leading to removal of these mutations from a population (Lynch et al. 1995). Alternatively, asexuality (i.e. no meiosis, recombination and syngamy) is predicted to slow adaptation (de Visser at al. 2005).

Sex may be seen as advantageous with respect to purging the weak and favouring the continuation of the fittest individuals, but there are also disadvantages associated with sex. Sex requires a male and a female gamete to fuse, and unlike hermaphroditic organisms which can produce both male and female gametes, most sexual organisms have separate sexes (Policansky, 1982). In order for sex to occur one must find a partner, a process which requires time and energy and, in some cases, can create risk (Hendrichs et al. 1991). Furthermore, finding a partner often entails competition from other members of the same sex (Buunk and Massar, 2014). Plants are inadequate examples for courtship competition due to their sessile nature, but in the animal kingdom the dominant observation involves male competition to earn reproduction with a female (i.e. female choice; Székely et al. 2004). The two-fold cost of sex relates to the cost of producing

males because they cannot give birth, and secondly, how it takes two individuals to create offspring (Otto, 2009). Therefore, asexuals hold a competitive edge in being able to colonize since every individual is capable of producing offspring.

Globally some apomictic plants occupy ranges distinct from their sexual counterparts in what is known as geographic parthenogenesis (Cosendai et al. 2013; Hörandl, 2006). Sex is theoretically disadvantageous compared to asexuality for colonizing novel areas (e.g. post glacial recolonization) because of the need to find and mate with a member of the opposite sex (Hörandl, 2009). In contrast, asexuality is advantageous in the rapid colonization of new or recovering habitats, considering phenomena such as hurricanes, fires, or volcanoes, in addition to post-glacial recolonization of Europe and North America (Lundmark and Saura, 2006).

The Red Queen hypothesis describes antagonistic coevolution between species (e.g. host-parasites) which is thought to favour sexual reproduction (Agrawal, 2006). By creating genetically variable offspring, the imposed selection on particular "weak" genotypes by pathogens would lead to continuous evolution of the host through sex (i.e. selecting for resistant genotypes), securing its existence (Peters and Lively, 1999). Pathogens must similarly change over generations to adapt to genetically evolving host genotypes. What happens when there is no sex? Asexual organisms are at a disadvantage when faced with a pathogen, since the pathogen is expected to adapt to the genetically invariable asexual genotypes which can thus not escape, leading to a decrease in their fitness or extinction (Morran et al. 2011).

While the Red Queen Hypothesis emphasizes the inability to change and adapt to biotic pressures, asexual organisms are also at a disadvantage on the genomic level through mutation accumulation. Muller's Ratchet is an evolutionary hypothesis used to describe the accumulation and fixation of mutations in an asexual lineage due to the lack of recombination and sex (Muller,

1964). The results of mutation accumulation are rarely positive, and associated negative effects on the genomic level are loss of gene function, impaired gene function, and transposable element accumulation (Disteche, 2012). For these reasons asexuality (apomixis in plants, parthenogenesis in animals) is seen as an evolutionary dead end (Hand and Koltunow, 2014). The assumptions that asexual organisms cannot adapt to new selection pressures and furthermore that they will accumulate mutations through time (i.e. Muller's ratchet) provide argumentation (Lovell et al., 2017) for the evolution and maintenance of sex throughout Eukaryotes.

1.1 **Apomixis**

Apomictic plants produce seeds whose embryos (offspring) are genetically identical to the parent plant. Apomixis is derived from sexual reproduction (Lovell et al. 2013)(Figure 1.2; 1.3) whereby the sexual reproductive cycle is modified via two pathways, and referred to as gametophytic or sporophytic (Figure 1.4). Gametophytic apomixis is characterized by the origin of the female gamete (megasporogenesis) from a meiotically unreduced megaspore. In diploid plants, normal megasporogenesis would result in the reduction of the diploid megaspore into four haploid cells through meiosis, but in gametophytic apomixis meiosis is modified through two different pathways: apospory or diplospory (Koltunow, 1993). In apospory the megaspore mother cell degenerates and is replaced by a nucellar cell (i.e. change in cell fate) that continues through megasporogenesis after completely avoiding meiotic reduction. Unlike apospory, diplospory sees the megaspore mother cell go through a modified meiosis (apomeiosis) that results in an egg cell having the same ploidy as the mother plant (Grimanelli et al. 2001; Nogler, 1984). In all cases, the apomeiotically-derived (meiotically-unreduced) egg cell develops into an embryo without fertilization (i.e. parthenogenetic development).

Sporophytic apomixis is similar to gametophytic apospory in that a cell from another tissue type becomes the egg cell, but in sporophytic apomixis instead of just one cell replacing the megaspore there is the possibility of many somatic ovule cells surrounding the meiotically-reduced egg cell that transition to an embryonic cell fate. Once fertilization of the sexually derived embryo has occurred, the previously somatic ovule cells (now somatic embryos) can share the endosperm, which gives rise to seeds containing more than one embryo (polyembryony) (Koltunow, 1993).

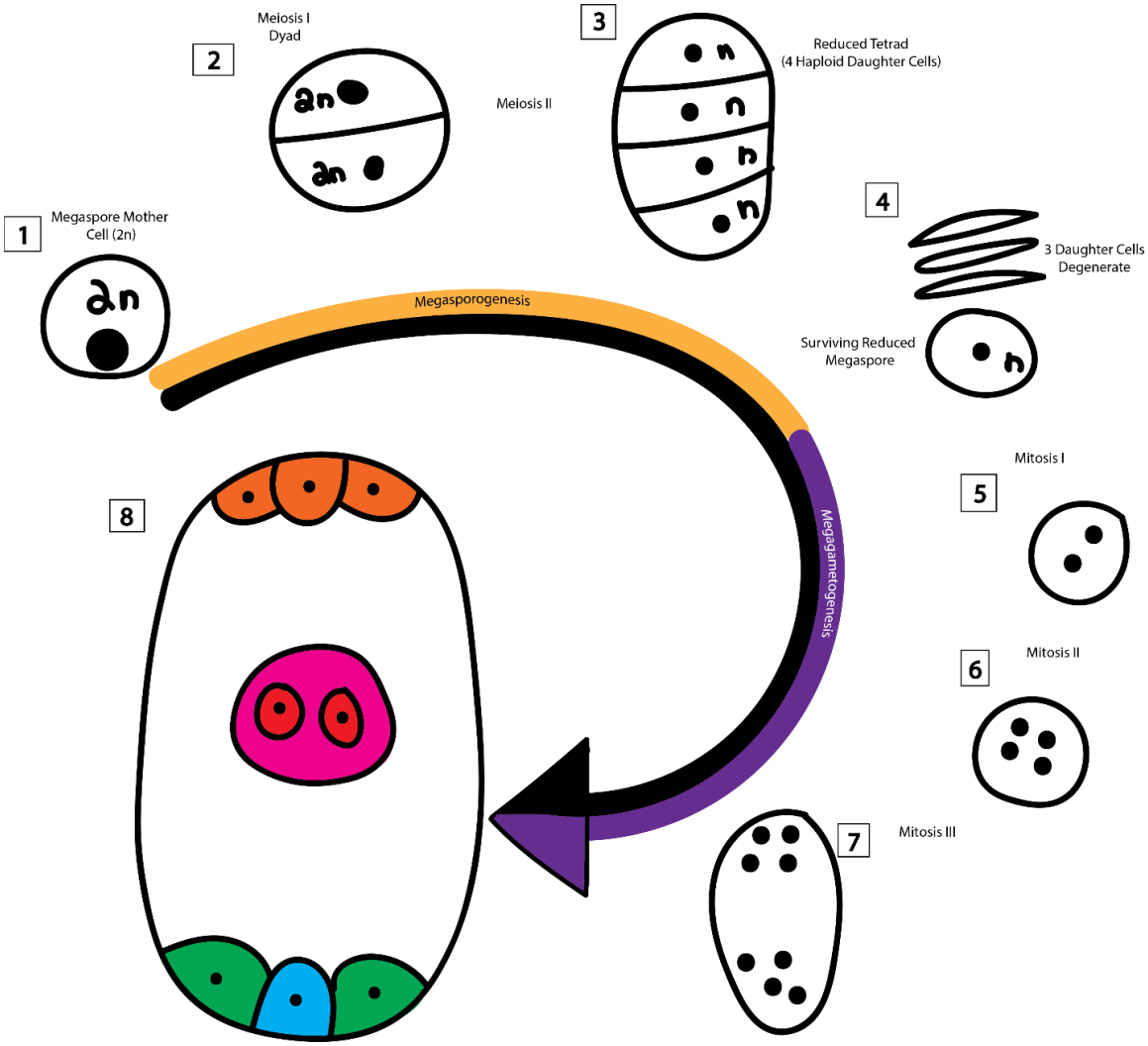

Figure 1.2 The megaspore mother cell develops into the polygonum type embryo sac (8) through a number of steps: the megaspore mother cell must undergo meiotic reduction (1, 2, and 3) resulting in four haploid cells (3); out of those four cells, only one survives (4) and continues through three rounds of mitosis producing eight cells (5, 6, and 7). These seven cells are organized to create an architecture consisting of three antipodals (orange), two central cells (red), and two synergids (green) that flank the egg cell (blue).

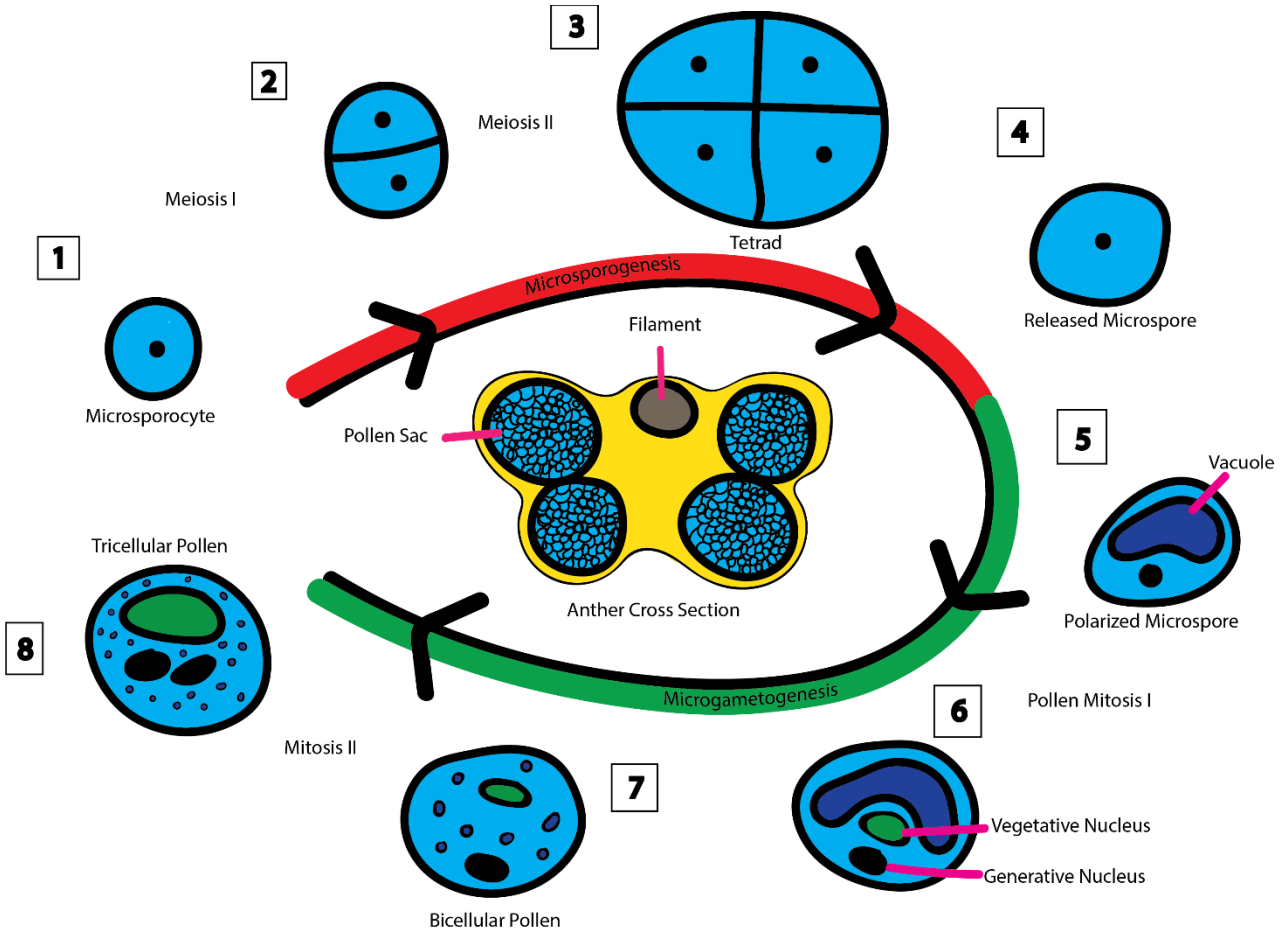

Figure 1.3 Pollen microspores originate from the pollen sac located within the anther that is supported by a filament. Microsporogenesis begins with a single celled microsporocyte (1) that undergoes meiosis producing a tetrad of reduced microspores (3). As microgametogenesis proceeds a large vacuole forms, pushing the nucleus against the cell wall before mitosis I (4 and 5), which compartmentalizes the generative nucleus from the vegetative nucleus (6). Additionally, during formation of bicellular pollen, the vacuole decreases in size and smaller vesicles are observed (7). During mitosis II of microgametogenesis the generative nucleus doubles (8), resulting in a mature male gametophyte that is capable of double fertilization (embryo and central cell of the female gametophyte).

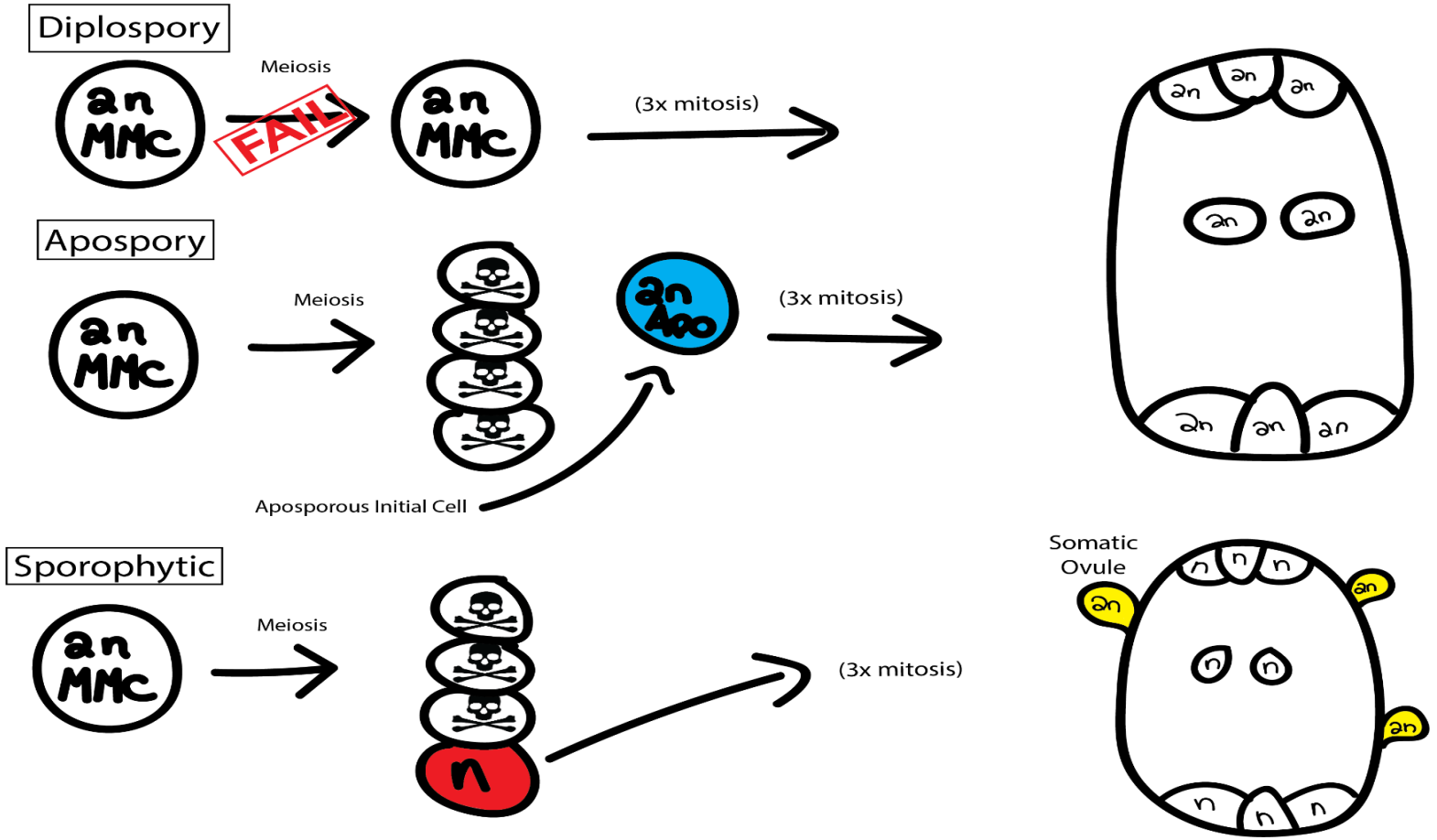

Figure 1.4 This figure compares the different pathways of seed development in gametophytic and sporophytic apomixis. Apospory and diplospory are both classified as gametophytic apomixis because the mature embryo in both instances is the same, but achieved through different pathways. Diplospory sees the failure of meiosis leading to no reduction of the megaspore. The unreduced megaspore then becomes source material for the initiation of embryo sac development. In aposporic apomixis, the megaspore mother cell undergoes meiotic reduction but all four reduced cells degenerate and are replaced by a single aposporous initial cell. The aposporous initial cell originates from the nucellus or integument tissue that surrounds the megaspore mother cell. This aposporous initial cell develops mitotically into a mature embryo with the same ploidy as the parent. Sporophytic apomixis differs from gametophytic apomixis in that the megaspore mother cell follows normal sexual development resulting in a reduced mature embryo. What makes this apomictic is the parthenogenic development of additional ovaries from nucellus or integument tissues that can accompany the sexually derived embryo. The somatic ovules are genetic clones of the mother plant.

1.2 **Apomixis and Agriculture – A Tool for Fixing Heterosis**

Though sex has created vast amounts of variation in the plant kingdom, when it comes to modern agriculture and plant breeding it is undesirable to have variation in large-scale monoculture cropping systems. Farmers require uniformity in relation to what is growing in their fields for decisions to be made; for example, regarding chemical applications and harvest timing. A plant breeder's job is to select desirable traits from available germplasm, and incorporate them into new cultivars/varieties containing improved genetics, leading to higher yields and resistance to biotic and abiotic stresses. Heterosis, a complex trait commonly used to increase yield, is generated by inbreeding selected parental lines for several generations and crossing them to produce offspring that are significantly superior compared to their inbred parents (Charlesworth and Willis, 2009). The entire process of inbreeding selected lines is lengthy and expensive (e.g. emasculation during the production of hybrid maize), and can vary depending on a particular species' lifecycle. In order to create hybrid-breeding systems it is additionally desirable to have male sterile lines that ensure only pollen from another plant is used to fertilize the female gamete (Li et al, 2007). Male sterility can be achieved genetically or through emasculation methods, which further require new lines be developed for desirable hybrids and/or skilled labor for emasculation.

Alternative to conventional breeding practices, the introduction of apomixis into crops would allow for years of research and development to be bypassed, as complex (i.e. heterozygous) genomes could be genetically fixed and fast tracked into current commercial breeding program workflows. The ability to incorporate clonal seed production into the breeding process would greatly speed up the development of new varieties (e.g. single generation hybrids), and for farmers in developing countries the ability to reseed material that has been specifically bred for their local environment would be one of several advantages (Spillane et al. 2004). Farmers would have access

to a larger spectrum of varieties resulting from faster variety development, but there are legal implications, as farmers would have the potential to keep F1 hybrid seeds, which from a business point of view sounds disastrous for seed companies. However, seed companies would be able to benefit from reduced research and development costs due to having the ability to generate new varieties more quickly. Overall, apomixis technology presents a challenge to policy makers and industry, as novel business models will need to be developed in order for all parties to benefit from this disruptive technology.

2 Literature Review

2.1 Boechera - A Model System for the Evolution and Functional Genetics of Apomixis

Boechera is a North American genus consisting of 110 species (Mandákova et al. 2015), and is in the *Brassicaceae* family that also contains economically important plants such as camelina, canola, cabbage, and bok choy. Within the *Brassicaceae*, *Boechera* is the only genus to exhibit asexual reproduction through seeds (apomixis). *Boechera* reproduces via diplosporous apomixis, a pathway composed of three steps, the formation of a meiotically unreduced egg cell (apomeiosis), parthenogenetic development of this egg cell without fertilization, and formation of endosperm with (pseudogamy) or without (autonomous) fertilization of the central cell. Apomixis is extremely rare in crucifers (Mandakova and Lysak, 2008), and because of economically important relatives, the mechanisms that lead to apomixis in *Boechera* hold immense potential to revolutionize plant breeding, as it takes another step forward to meet future global food security.

Apomixis has been observed across *Boechera* (Schranz et al. 2005, Koch et al. 2003), in both the lab and natural habitats. Within natural populations, *Boechera* plants vary in reproductive pathway (sexual and facultative apomictic) and ploidy level (diploid, triploid, tetraploid; Sharbel et al. 2009). Facultative apomixis, whereby a single plant produces both sexual and asexual offspring, is a frequent observation, further obscuring the genetic factors underlying its control (Aliyu et al. 2010). The biological diversity of apomictic *Boechera* has enabled comparisons of the effects of ploidy and reproductive mode on niche occupation, and the ability to adapt to changing environments (Mau et al. 2015). Gene flow (hybridization) between apomictic and sexual *Boechera* occurs via apomictic pollen (Mau et al. 2013) to produce plants that differ in ploidy and reproductive pathway (Figure 2.1; Table 2.1), leading to a complex pedigree comprised of different species (Rushworth et al. 2011).

Sexual diploid organisms are the unadorned model for reproductive biology; two sets of chromosomes with one set each coming from the male and female parent. Biologically this creates a 1:1 ratio of genetic contribution from each parent to their offspring. Sexual seed production in plants adds an additional layer of complexity because seeds have an embryo and endosperm which are produced after the egg and central cells of the ovule undergo "double fertilization". Mature pollen grains contain two sperm nuclei and one vegetative nucleus; the two sperm nuclei being the male gametes that fertilize the egg to create an embryo and central cell to form the endosperm. Overall the developing seed has an embryo: endosperm ploidy ratio of 2 (1 maternal + 1 paternal genome):3 (2 maternal + 1 paternal genome) which is characteristic of sexually reproducing plants. The 2 maternal : 1 paternal genome ratio of developing endosperm is highly constrained in sexual plants (Scott et al. 1998), and crosses involving different ploidies which lead to different endosperm ratios results in seeds with irregular development or complete abortion (Alleman and Doctor, 2000). This observation is at the center of what is known to evolutionary biologists as parental conflict theory through genomic imprinting (Köhler and Weinhofer-Molisch, 2010; Costa et al. 2012).

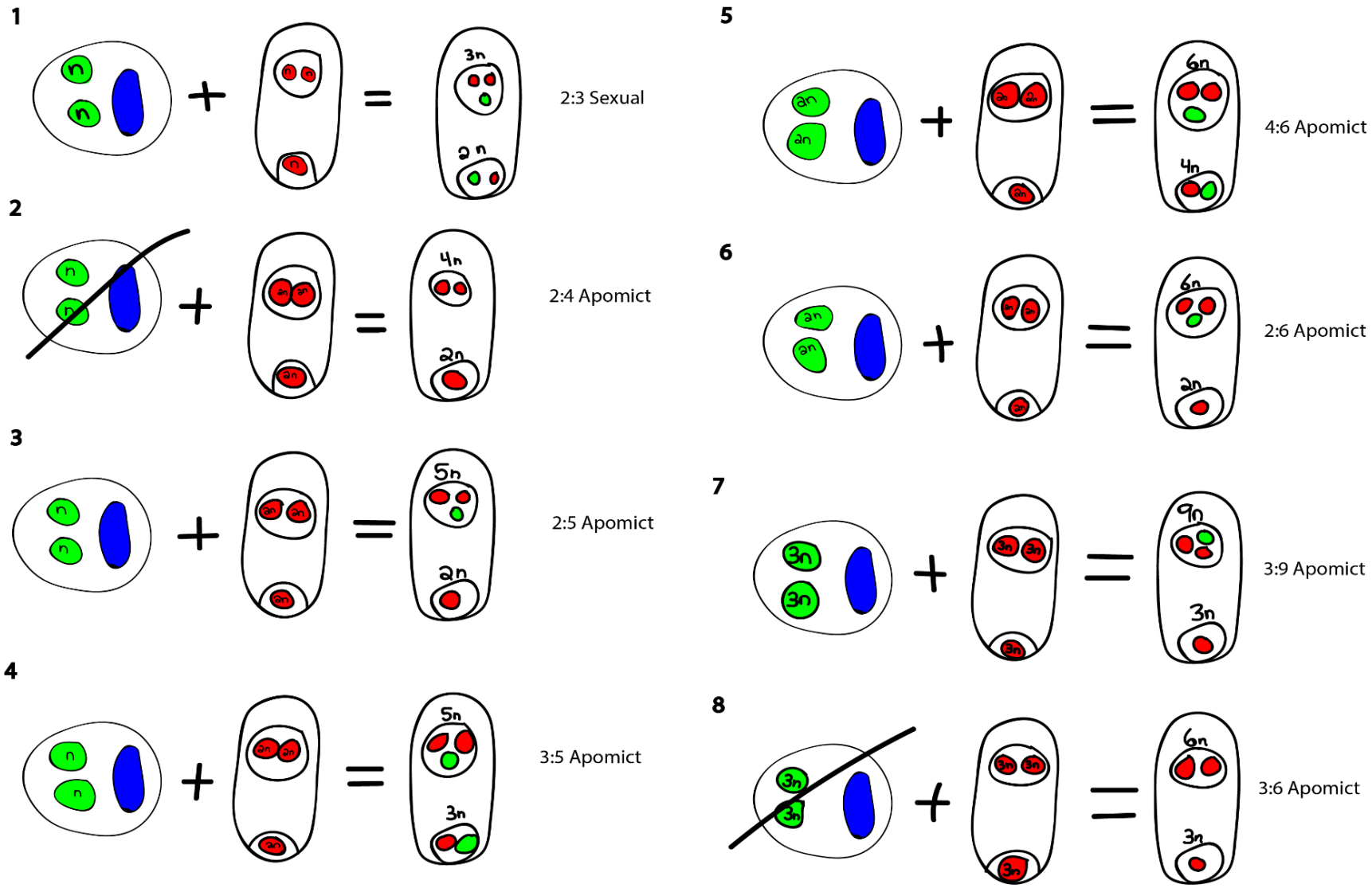

Figure 2.1 Seed formation in *Boechera* is characterized by embryos and endosperm of varying ploidies based on i) fertilization of the egg and ii) central cells. 1. Both central cell and egg are fertilized by haploid sperm cells; 2. No fertilization of the unreduced egg cell occurs with autonomous endosperm formation; 3. Egg cell is unfertilized but haploid sperm cell contributes to the central cell; 4. Both haploid sperm cells contribute to create a triploid embryo and a 5n central cell; 5. Unreduced diploid or reduced tetraploid sperm cells double fertilize unreduced diploid embryo; 6. Only the central cell is fertilized by a diploid sperm cell; 7 and 8. Triploids are fertile and have triploid pollen, with both autonomous and pseudogamous endosperm formation.

Table 2.1 Descriptions of seed development processes occurring in each example illustrated in Figure 2.1

Number	Apomeiosis or Meiosis	Parthenogenesis or Fertilized Egg Cell	Fertilized/ Autonomous Central Cell
1	Meiosis	Fertilized	Fertilized
2	Apomeiosis	Parthenogenesis	Autonomous
3	Apomeiosis	Parthenogenesis	Fertilized
4	Apomeiosis	Fertilized	Fertilized
5	Apomeiosis	Fertilized	Fertilized
6	Apomeiosis	Parthenogenesis	Fertilized
7	Apomeiosis	Parthenogenesis	Fertilized
8	Apomeiosis	Parthenogenesis	Autonomous

In the realm of sexual reproduction, dioecious plant species evolved both chromosomal and autosomal sex determination systems (Barret, 2002; Charlesworth, 2002). However, in plants, dioecy is uncommon and the kingdom is dominated by monoecious plants having both male and female reproductive organs. This poses the question; where is the benefit in maternal and paternal control of the offspring? Chromosome sets coming from a male or female have their own interests and therefore compete for resources (Kinoshita et al. 2008). With plants which either outcross or both self and outcross, it is possible for different males to fertilize the many flowers of a female. With all female egg cells being equally related to the mother, the mother wants to equally distribute her available resources amongst her offspring (Costa et al. 2012; Charnov, 1979). The resource being mentioned is the amount of endosperm, the nourishing tissue that will be used by the developing embryo during germination. With different males fertilizing available egg cells, one would expect selection for genes to "entice" the female for more resource allocation to one's own offspring, thus leading to conflict. The evolutionary resolution to this conflict has been for the mother to double her genomes in the developing endosperm to maintain control of her interests

(Haig and Westoby, 1989). Evolution has given rise to female genome excess in sexually derived triploid endosperm, and interploidy crosses have been used to test hypotheses surrounding this ratio (Pennington et al. 2008). For example, using the model plant *Arabidopsis thaliana*, Scott et al. (1998) conducted interploidy crosses that offset normal sexual endosperm genome ratio, and found that excess paternal contribution resulted in bigger seeds while excess female contribution lead to smaller seeds, as expected under parental conflict theory.

Seed size is a valuable agronomic trait, and hence studying interploidy crosses for commercial purposes could lead to mechanisms to increase crop seed size (i.e. yield). However, in the case of apomixis, which also holds significant plant breeding commercialization potential, (predicted) variability in endosperm balance numbers and genomic imprinting have been demonstrated (Quarin. 1999). In apomictic *Boechera* species, the use of flow cytometry has shown that apomictic diploid plants can have embryo: endosperm ratios of 2:4. 2:5 and 2:6 (Rojek et al. 2018). The seeds produced from apomictic *Boechera* vary in size but no obvious association between sexual or apomictic accessions (Guenter, 2017). With *Boechera* being an example of variability in endosperm contributions of an apomictic species it comes to show that despite being apomictic, it is still possible for genomic contribution to occur because the endosperm can be sexually (e.g. pseudogamy) or asexually derived. For plant breeders, sexually formed endosperm is undesirable and hence the desired ratio of an apomictic diploid plant for agronomic use is 2:4 (autonomous endosperm development of two unreduced central cells), as this would mean asexual production of both important seed tissues and no requirement for crosses to be performed.

2.2 **Apomictic Candidate Genes**

Many candidate genes for apomixis have been identified over the past forty years in both apomictic and sexually reproducing plants**.** One train of logic would present the argument that apomictic candidate genes should be strictly associated with apomictic species, as is supported by examples such as UPGRADE2 in *Boechera* (Mau et al. 2013). Many apomictic organisms are facultative, and thus produce both asexual and sexual offspring (obligate apomicts are extremely rare (Mau et al. 2015)), accentuating the importance of comparing genes involved in megasporogenesis and embryogenesis between sexual organisms. Genes involved in megasporogenesis and embryogenesis can be involved in several other processes given gene expression, protein interactions, and regulatory effects (Santos and Aragão, 2009), therefore being able to identify the catalyst(s) initiating the necessary events leading to apomictic seed production is exceptionally complex.

Genetic technologies have significantly advanced since the development of Sanger sequencing, and when the polymerase chain reaction (PCR) transformed biotechnology in the 1980's. From being excited about replicating and sequencing DNA a few decades ago to the present, where we are able to identify functional interactions and consequences of mutations on a genome level, many doors have been opened for studying the functional genetics of apomixis. Model organisms have been established (e.g. *Arabidopsis thaliana*, *Tobacco*, and *Zea mays)* to act as a references when researching (orthologous) genes identified from distant plant relatives. As apomixis has only been found in close crop relatives such as *Tripsacum* and *Boechera* (Hörandl and Hojsgaard, 2012), the implication is that in order for apomixis to be used in crop plants it must be genetically engineered.

Gene expression experiments of female gametophyte development have played an important role in identifying genes that may possess a role in any of the three stages required to form a functional apomictic seed (apomeiosis, parthenogenesis, and endosperm development; Sharbel et al. 2010; Laspina et al. 2008; Podio et al. 2014). Historically, apomictic plants do not have the omics resources available comparable to that of *Arabidopsis thaliana*, but advances in single cell manipulation and DNA sequencing technologies are advancing apomixis research at a rapid pace. The field of functional genomics is similarly undergoing a revolution with regards to gene editing. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) has played a major role by increasing the simplicity and efficiency of artificially executing gene knockouts, inducing mutations with new levels of specificity, and gaining insight into gene expression and cellular localization (Jinek et al. 2012).

Corral et al. (2013) and Mau et al. (2013) discovered two genes which were highly upregulated in the ovules and pollen of apomictic compared to sexual accessions in *Boechera*, and proposed that both were involved in the switch from sex to apomixis. The expression of APOLLO (apomixis-linked locus) is correlated with the generation of a meiotically unreduced megaspore mother cell. APOLLO occurs in both apomictic and sexual *Boechera*, and has apomixis-specific nucleotide polymorphisms and copy number variation which are hypothesized to be associated with functional differences between reproductive modes (Corral et al. 2013). UPGRADE2 (unreduced pollen grain development) is found in apomictic *Boechera* only, and is hypothesized to be associated with meiotically unreduced pollen formation (Mau et al. 2013).

APOLLO and UPGRADE2 were both identified as most significantly differentially expressed genes in two different microarray experiments (Corral et al., 2013). Nonetheless, the underlying data from the original microarray data comparing the reproductive tissues of sexual

and apomictic *Boechera* has not been compared to new genetic data resources. For example, there are a small number of "secondary candidate genes" showing marginal significance from the microarray experiments, and could represent more subtly differentially regulated genes between sex and apomixis. Secondary candidate genes are important because they can act complementarily to primary candidate genes, for example by influencing the variable genetic penetrance (i.e. levels) of apomixis seen in different apomictic *Boechera* accessions (Aliyu et al. 2010)

This chapter of my thesis is thus devoted to two lines of analysis. Firstly, apomixis associated genes found in the literature will be compared to known gene expression data previously obtained in sexual and apomictic *Boechera* (Corral et al., 2013). Secondly, the data in the original APOLLO array project (Corral et al. 2013) will be reanalyzed to identify secondary candidate genes of interest. Both sets of genes have potential implications for engineering apomixis in crops.

3 Materials and Methods

3.1 Apomixis Candidate Gene Identification

Novel genes discovered in apomictic plants are potentially more interesting compared to genes identified from distantly related sexual reproducing species. However, given the potential for apomixis to be artificially induced with the current knowledge of genes involved in sexual reproduction (Khanday et al. 2019), it is potentially useful to compile genes that cause apomictic-like phenotypes in sexually reproducing individuals.

Through extensive literature review, genes of interest from both sexual and apomictic species were taken into consideration based on their function(s) during (i) apomeiosis and meiosis, (ii) parthenogenesis, and (iii) endosperm development. In all cases, orthologues to any genes of interest were identified in the model *Arabidopsis*. Using TAIR (Berardini et al. 2015) and UniProt (UniProt Consortium, 2017) the protein identifiers for each gene in *Arabidopsis thaliana* was retrieved for submission to InParanoid8 (Sonnhammer and Östlund, 2014). InParanoid is an accredited database containing 273 proteomes and uses an algorithm to identify functional orthologues between full proteomes of two organisms through bootstrapping and applying confidence values to orthologue clusters. All inparalogs with scores >0.05 were excluded from consideration.

3.2 Microarray Analysis

The microarray data used in Corral et al. (2013) to identify APOLLO was reanalyzed to identify marginally significant array probes whose expression differed between the live ovules microdissected independently from 10 apomictic and 8 sexual *Boechera.* The data was analyzed for differential expression using the R statistical software (Version 3.2.3) in combination with the Bioconductor limma software (Smyth et al. 2004). The microarray data was subjected to

normalization using the quantile method in addition to background correction with 95% confidence for multiple testing. From this statistical analysis, the top twenty probes were chosen to be the foundation for all analyses conducted in this study (Table 4.3 and 4.4).

3.3 **DNA Sequence Annotation**

Complementary DNA (cDNA) libraries from three sexual and three apomictic *Boechera* accessions at various flowering stages were pooled together and sequenced using 454 (FLX) technology (Corral et al. 2013). The sequences were assembled using CLC Genomics Workbench (CLC bio, version 5.5) and after trimming and sequence quality filtering 36 289 contigs and 154 468 non-assembled singleton sequences resulted. From this assembly ImaGenes (GmbH. Berlin, Germany) designed custom Agilent microarray chips based on 60bp long (i.e. 60mers) oligonucleotides. For every contig in the genome assembly there was 14 matching oligos on the microarray chip and for every singleton there was 8 oligos. This design was performed to have all possible gene expression product from different growth stages and tissues to minimize bias in the resulting data.

The first step in my differential expression analysis was to determine to which *Boechera* cDNAs the top 20 probes (60mers) corresponded. Using CLC Genomics Workbench by CLC Bio (8.5.1), the top 20 probe sequences were blasted to a number of *Boechera* specific cDNAs and genomic databases in order to identify potential target sequences (Table 3.1). Nucleotide BLAST (standard parameters) was used to find similar sequences in whole genome sequencing assemblies (WGS) from an apomictic *Boechera divaricarpa* and sexual *B. stricta*, in addition to *B. stricta* coding sequence (CDS) and transcript libraries. In addition, BLASTp and tBLASTn to a *Boechera stricta* protein library were performed.

From each BLAST result of the probe sequences from the microarray (60mers) to the *Boechera* cDNA libraries (Table 3.1), significant matches were defined as hits having the highest alignment length. This process led to the generation of five FASTA files containing all matching sequences extracted from the different *Boechera* sequence libraries. These five files were then used for NCBI (National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/) and TAIR (The Arabidopsis Information Resource (TAIR), http://www.arabidopsis.org) BLAST searches (Table 3.1). Using CLC Genomics Workbench for the NCBI BLAST, each FASTA file containing unannotated *Boechera* sequences was compared to the NCBI database using default search settings, and the organism filter set to limit results to virdiplantae. To align these BLAST results to a model organism, I then moved to TAIR (The Arabidopsis Information Resource (TAIR), http://arabidopsis.org) database, as *Boechera* and *Arabidopsis* are closely related and share genome synteny (Schranz et al. 2006).

Only the protein, transcript, and CDS sequences from *B. stricta* were submitted to the TAIR10 databases for BLAST analysis. Using default search settings, BLASTp was used to search the TAIR10 protein library, and BLASTn to search the TAIR10 transcripts and CDS libraries to identify orthologous sequences in *Arabidopsis*. The top hits from these analyses were picked based on percent identity, alignment length, and e-value which distinctly separated significant hits from the rest. After cutoff for these thresholds, the remaining *Arabidopsis thaliana* loci were checked for duplicates and compiled into one table. This group of loci represented the orthologues of secondary candidate genes linked to apomixis in *Boechera,* in the model organism *Arabidopsis thaliana* (Table 4.2).

Table 3.1 Databases used for annotation

Query	BLAST Type	Database	Results Table #
Table 4.3	BLASTn	Reference *B. stricta* cDNA transcript library	Table 4.11
Table 4.3	BLASTn	Reference *B. stricta* cDNA CDS library	Table 4.13
Table 4.3	BLASTn	Apomictic *B. divaricarpa* WGS cDNA library	Table 4.6
Table 4.3	tBLASTx	Reference *B. stricta* protein library	Table 4.14
Table 4.3	BLASTn	Sexual *B. stricta* WGS cDNA library	Table 4.5
Table 4.11	BLASTn	TAIR10 Transcript	Table 4.12
Table 4.13	BLASTn	TAIR10 CDS	Table 4.7
Table 4.14	BLASTp	TAIR10 Protein	Table 4.8
Table 4.5	BLASTn	NCBI nucleotide collection	Table 4.15
Table 4.6	BLASTn	NCBI nucleotide collection	Table 4.16
Table 4.14	tBLASTn	NCBI nucleotide collection	Table 4.17

4 Results

4.1 Compilation of Apomixis Candidate Genes

In Table 4.1 and 4.2, it is noticeable that there are very few novel apomixis candidate genes known to the scientific community, which directly reflects how little is known about the evolution and maintenance of apomixis in natural systems, or that apomixis arises via deregulation of genes usually associated with sex (*sensu* Carman, 1987). Notably absent from the tables previously mentioned are regions such as *Happy* (Schallau et al. 2010) and DIP (Vijverberg et al. 2010) which have been directly linked to apomixis in their respected species, but for which gene products contained within the loci have have not been identified. Gene expression studies into apomicts have observed global de-regulation of gene expression (Koltunow and Grossniklaus, 2003), making it statistically difficult to identify candidate genes, while in sexual plants many of these same genes are involved in the regulatory transition of a plant changing from vegetative to reproductive growth. Several transcription factor genes that are candidates for apomixis also pose potential for being used in plant regeneration studies because of their ability to transcriptionally reprogram somatic cells into somatic embryos (Junker and Bäumlein, 2012; Lowe et al. 2016; Zuo et al. 2002).

Table 4.1 Spreadsheet of documented candidate genes for apomixis with their corresponding species of interest and gene ID's in other important crop plants for phylogenic interest. Literature background for all the genes listed here are in Table 4.2. In each gene's corresponding literature, *Arabidopsis thaliana* is highly referenced making it possible for retrieval of the protein ID through TAIR. The protein ID's in this table are the means of identifying orthologous gene ID's in other plants via InParanoid8.

Species of Interest Documented	Gene Name	Protein ID	*Arabidopsis thaliana* Gene Name	*Arabidopsis thaliana* ID	*Maize* Gene Name	*Oryza sativa* Gene ID	*Glycine max* Gene ID	*Brachyopodium distachyon* Gene ID	*Vitis vinifera* Gene ID	*Solanum lycopersicum* Gene ID
Maize	AFD1				AFD1					
Arabidopsis thaliana	AGL15	Q38847	AGL15	AT5G13790						
Arabidopsis thaliana	AGL6	P29386	AGL6	AT2G45650		OS02G45770	GLYMA07G081300	BRADI3G51800	VIT_15S0048G01270	SOLYC01G093960
Maize	AGO104	Q84VQO	AGO9	AT5G21150	AGO104					
Arabidopsis thaliana	AGP18	Q9FPR2	AGP18	AT4G37450			GLYMA05G102100		VIT_07S0129G00560	SOLYC02G092790
Arabidopsis thaliana	AINTEGUMENTA (ANT)	Q38914	CKC1	AT4G37750						
Boechera	APOLLO	Q9CA74	NEN3	AT1G74390						SOLYC12G036220
Poa pratensis	APOSTART	F4JSE7	EDR2	AT4G19040						
Panicum maximum	ASG-1	Q08298	RD22	AT5G25610		OS01G53240	GLYMA14G141000	BRADI2G49000	VIT_04S0008G03930	SOLYC08G068150
Arabidopsis thaliana	ATROPOS (ATO)	Q9FG01	ATO	AT5G06160						
Cenchrus ciliaris Pennisetum	BABY BOOM-like									
Arabidopsis thaliana	BBM1	Q6PQQ4	BABY BOOM	AT5G17430		OS04G42570	GLYMA09G248200			SOLYC11G008560
Arabidopsis thaliana	BLH1 (BEL1)	Q38897	BEL1	AT5G41410		OS06G0108900 / OS06G01934	GLYMA08G016900	BRADI1G76940	VIT_02S0025G00200	SOLYC08G081400
Maize x Tripsacum Hybrid	CHR106		DDM1		CHR106					
Arabidopsis thaliana	CHR11	F4JAV9	CHR11	AT3G06400		OS01G0367900	GLYMA15G097000	BRADI2G12950	VIT_05S0020G01780	
Arabidopsis thaliana	CKH2	Q9S775	PICKLE	AT2G25170		OS06G08480	GLYMA06G063400	BRADI1G47367	VIT_04S0008G05880	SOLYC06G065730
Arabidopsis thaliana	CLOTHO (CLO)	Q8RWY3	CLO	AT1G06220						
Arabidopsis thaliana	DMC1	Q39009	DMC1	AT3G22880						
Maize x Tripsacum Hybrid	DRM1	Q9LXE5	DRM1	AT5G15380	DMT103					
Maize x Tripsacum Hybrid	DRM2	Q9M548	DRM2	AT5G14620	DMT103					
Arabidopsis thaliana	DYAD/ SWITCH (SWI1)	Q9FGN8	SWI1	AT5G51330	AM1	OS03G0650400	GLYMA15G30830	BRADI1G13530	VIT_00S0199G00200	SOLYC03G093370
Hieracium	FIE	Q9LT47	FIS3	AT3G20740		OS08G0137100	GLYMA10G022000	BRADI3G14520	VIT_19S0014G05210	SOLYC07G064090
Arabidopsis thaliana	FIS2	P0DKJ8	FIS2	AT2G35670						
Arabidopsis thaliana	FUS3	Q9LW31	FUS3	AT3G26790			GLYMA19G100900		VIT_14S0068G01290	SOLYC02G094460
Hypericum perforatum	HAPPY LOCUS	Q84RR0	Ariadne7	AT2G31510			GLYMA13G343300			
Oryza sativa	HOP1/PAIR2	F4HRV8	ASY1	AT1G67370		OS09G32930	GLYMA18G006800	BRADI4G34650	VIT_01S0010G03590	SOLYC04G007330
Arabidopsis thaliana	HOP2	Q9FX64	AHP2	AT1G13330			GLYMA08G260400	BRADI1G11260	VIT_01S0011G03320	SOLYC05G009560
Maize	IG1	O04479	LBD6	AT1G65620			GLYMA13G191000	BRADI2G25270	VIT_00S0340G00090	SOLYC11G008830
Arabidopsis thaliana	LACHESIS (LIS)	O22212	LIS	AT2G41500			GLYMA10G000800	BRADI1G62257	VIT_08S0007G05930	SOLYC04G016510
Arabidopsis thaliana	LEC1	Q9SFD8	LEC1	AT1G21970						
Arabidopsis thaliana	LEC2	Q1PFR7	LEC2	AT1G28300		OS04G0676650	GLYMA20G035800		VIT_10S0071G00270	
Arabidopsis thaliana	MEDEA (MEA)	O65312	FIS1	AT1G02580						
Alfalfa	MOB1	Q9FHI1	MOB1A	AT5G45550						
Arabidopsis thaliana	MPC	Q9LT82	MPC	AT3G19350					VIT_18S0001G10190	
Arabidopsis thaliana	MSI1	O22467	ATMSI1	AT5G58230		OS03G43890	GLYMA08G085900	BRADI1G13930	VIT_16S0013G01550	SOLYC01G104510
Arabidopsis thaliana	OSD1	Q9M2R1	GIG1	AT3G57860		OS02G37850	GLYMA19G144400	BRADI3G47430	VIT_08S0007G07790	SOLYC10G080400
Pennisetum ciliare	PCA21									
Pennisetum ciliare	PCA24									
Arabidopsis thaliana	REC8/SYN1	Q9S7T7	SYN1	AT5G05490		OS05G50410	GLYMA02G263700	BRADI2G15127	VIT_13S0067G01660	SOLYC06G074870
Arabidopsis thaliana	RETINOBLASTOMA RELATED (RBR)	Q9LKZ3	RBR1	AT3G12280		OS08G42600	GLYMA15G230300	BRADI3G41630	VIT_04S0008G02780	SOLYC09G091280
Arabidopsis thaliana	RKD1	Q9M9U9	RKD1	AT1G18790		OS02G0744950		BRADI3G59410	VIT_17S0000G06250	
Arabidopsis thaliana	RKD2	Q9CA66	RKD2	AT1G74480			GLYMA05G01030			SOLYC03G116010
Arabidopsis thaliana	RKD4	Q9LVU8	GRD	AT5G53040			GLYMA06G129900	BRADI2G61760	VIT_16S0098G01730	
Poa pratensis	SERK1	Q94AG2	SERK1	AT1G71830				BRADI3G46747	VIT_18S0164G00070	
Poa pratensis	SERK2	Q9XIC7	SERK2	AT1G34210		OS08G0174700	GLYMA20G173000	BRADI5G12227	VIT_07S0031G01410	SOLYC04G072570
Arabidopsis thaliana	SPO11	Q9M4A2	SPO11-1	AT3G13170						
Arabidopsis thaliana	SPOROCYTELESS (SPL)/ NOZZLE	O81836	NZZ	AT4G27330						
Arabidopsis thaliana	SSN1	A8MSF8	RAD51d	AT1G07745		OS09G01680	GLYMA11G143900	BRADI1G27880		SOLYC11G073220
Boechera	UPGRADE2									
Arabidopsis thaliana	WEE1	Q8L4H0	WEE1	AT1G02970		OS02G0135300	GLYMA03G211900	BRADI3G03112	VIT_07S0104G01740	SOLYC09G074830
Arabidopsis thaliana	WUSCHEL	Q9SB92	WUS	AT2G17950			GLYMA01G166800		VIT_04S0023G03310	SOLYC02G083950

Table 4.2 Apomixis candidate genes identified through published literature including to which stage of apomictic development they are relevant. The majority of genes identified come from sexual backgrounds, but it is indicated if a gene is associated with an apomictic case study.

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
AFD1	Curtis and Doyle, 1991 Nan et al. 2011 Golubovskaya et al. 2006	•Rec8/ RAD21 homolog found in *Maize* •Involved in sister chromatid cohesion •Absence of First Division 1		yes		
AGL15	Harding et al. 2003 Perry et al. 1999 Serivichyaswat et al. 2015 Wang et al. 2004 Zheng et al. 2016	•Positive regulator of mir-156 •DNA binding transcriptional regulator •Promotes somatic embryo development •MADS-box family transcription factor			yes	
AGL6	Dreni and Zhang, 2016 Guimaraes et al. 2013 Ma et al. 1991	•Differentially expressed in ovules of apomictic and sexual *Brachiaria* •MADS-box family transcription factor	yes			
AGO104	Armenta-Medinam et al. 2011 EcKardt, 2011 Singh et al. 2011	•AGO9 in Arabidopsis thaliana •Mutants formed unreduced fertile gametes in maize •Involved in RNA silencing through small RNA pathways		yes	yes	yes
AGP18	Acosta-García and Vielle-Calzada, 2004 Demesa-Arévalo and Vielle-Calzada, 2013 Farquharson, 2013	•Regulates selection and survival of the functional megaspore resulting from meiotic reduction		yes	yes	
AINTEGUMENTA (ANT)	Elliot et al. 1996 Klucher et al. 1996 Krizek, 2003 Krizek, 2011 Mizumaki and Fischer, 2000 Randall et al. 2015	•AP2/ERF -type transcription factor •Regulates flower organ growth			yes	
APOLLO	Corral et al. 2013	•Differentially expressed in ovules of sexual and apomictic *Boechera*	yes	yes		

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
APOSTART	Albertini et al. 2005	•Differed in expression between sexual and apomictic *Poa pratensis* •*Arabidopsis* homolog is AT4G19040 •There are two copies: APOSTART1 & APOSTART2 •Assumed to be involved in sporogenesis •Downregulated in apomixis	yes	yes		
ASG-1	Chen et al. 1999 Chen et al. 2005	•Apomixis specific gene to *Panicum maximum* •Expressed in apomictic flower buds; not in sexual •Related to RD22 in *Arabidopsis*	yes		yes	
ATROPOS (ATO)	Moll et al. 2008	•Loss of function allele •Encodes a pre-mRNA splicing factor			yes	
***BABY BOOM-like**	Conner et al. 2008 Conner et al. 2015	•Conserved multiple copies in ASGR of both *Pennisetum* and *Cenchrus* apomictic species •Hypothesized to be responsible for parthenogenesis	yes		yes	
BBM1	Boutilier et al. 2002 Florez et al. 2015 Nic-Can et al. 2013 Khanday et al. 2019	•Induces somatic embryogenesis when ectopically expressed •Member of the AP2/ERF transcription factor family •Related to ANTIGUMENTA			yes	yes
BLH1 (BEL1)	Moll et al. 2008 Pagnussat et al. 2007	•Can trigger synergids to form gametes •Causes the *eostre* phenotype			yes	
CHR11	Huanca-Mamani et al. 2005	•Chromatin remodeling factor involved in transcription/repression •Sporophytic and gametophytic functions		yes		

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
CKH2	Furuta et al. 2011 henderson et al. 2004 Ogas et al. 1997 Ogas et al. 1999 Rider et al. 2003 Zhang et al. 2012	•Also known as PICKLE in other species •Is a loss of function mutation causing plant tissues to take on embryo identity •Chromatin remodeling factor involved in histone trimethylation			yes	
CLOTHO (CLO)	Moll et al. 2008	•Pre-mRNA splicing factor Snu114 •Loss of function mutation •Restricts the expression of LIS during egg cell specification		yes	yes	
DMC1	Bishop, 1994 Couteau et al. 1999 Doutriaux et al. 1998 Kathiresan et al. 2002 Kurzbauer et al. 2012 Pradillo et al. 2012 Seeliger et al. 2012	•Yeast homolog of RecA which is required for homologous recombination in *E. coli* •Cell cycle regulatory element crucial for meiosis •Expressed in both meiosis and mitosis		yes		
DMT102	Garcia-Aguilar, 2010	•Loss of function allele with downregulated expression in both apomictic *Boechera* and *Maize-Tripsacum* hybrid •Produced significantly larger pollen grains that was a mix of reduced and unreduced		yes	yes	
DMT103	Garcia-Aguilar, 2010	•Loss of function allele with downregulated expression in both apomictic *Boechera* and *Maize-Tripsacum* hybrid •Produced significantly larger pollen grains that was a mix of reduced and unreduced		yes	yes	
DYAD/ SWITCH (SWI1)	Agashe et al. 2002 Che et al. 2011 Golubovskaya et al. 1993 Mercier et al. 2001 Mercier et al. 2003 Motamayor et al. 2000 Nan et al. 2011 Pawlowski et al. 2009 Ravi et al. 2008 Sezer et al. 2016	•Also known as AM1 in other species •Expression is crucial for early events in meiosis •Mutants have different meiotic phenotypes between male and female meiocytes in *A. thaliana*		yes		

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
FIE	Gerashchenkov et al. 2015 Kohler et al. 2003 Liu et al. 2012 Luo et al. 2000 Ohad et al. 1999 Rodrigues et al. 2008 Spillane et al. 2000 Yadegari et al. 2000	•Gene mutation causes endosperm development without fertilization •Is needed for both sexual and apomictic endosperm development in *Hieracium* •Protein interactions are hypothesized to be different in *Hieracium* than *A. thaliana* •Candidate for autonomous endosperm development	yes			yes
FIS2	Gerashchenkov et al. 2015 Luo et al. 2000	•Imprinted gene •Suppresses endosperm development				yes
FUS3	Baumbusch et al. 2004 Kagaya et al. 2005 Tsuchiya et al. 2004 Yamamoto et al. 2010	•Expression induced by LEC1 •Transcription factor regulating seed maturation				yes
HOP1/PAIR2	Armstrong et al. 2002 Boden et al. 2007 Caryl et al. 2000 Nonomura et al. 2004 Sanchez-Moran et al. 2007	•Meiotic yeast gene •Mediates chromosome pairing		yes		
HOP2	Petukhova et el. 2005 Uanschou et al. 2013	•Binds single stranded and double stranded DNA •Recombination protein		yes		
IG1	Evans, 2007 Guo et al. 2004 Huan and Sheridian, 1996 Huan and Sheridian, 1998	•Mutant allele causes several defects in megasporogenesis •Notable differences are variance in endosperm ploidy and polyembryony			yes	yes
LACHESIS (LIS)	Groß-Hardt et al. 2007 Völz et al. 2012	•Homolog of yeast splicing factor PRP4 •Determines cell fate in the female gametophyte •Mutants form excess gametes			yes	

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
LEC1	Baumbusch et al. 2004 Baybrook and Harada, 2008 Gruszczyńska and Rakoczy-Trojanowska, 2011 Junker and Bäumlein, 2012 Klimaszewska et al. 2010 Maillot et al. 2009 Nic-Can et al. 2013 Shibukawa et al. 2009 Xie et al. 2008 Yamamoto et al. 2009 Yazawa and Kamada, 2007	•Transcription factor regulating embryogenesis •Ectopic expression triggers expression of LEC2 and FUS3 •LEC1 is a duplicated HAP3 gene that has evolved and further specialized its function			yes	
LEC2	Belide et al. 2013 Braybrook et al. 2006 Kim et al. 2014 Angeles-Nunez and Tiessen, 2012 Stone et al. 2001 Stone et al. 2008 Wojcikowska et al. 2013 Wojcikowska et al. 2015	•Upregulates FUS3 and ABI3 •Can also induce embryogenesis like LEC1			yes	
MEDEA (MEA)	Grossniklaus et al. 1998 Kinoshita et al. 1999	•Parent of origin dependent imprinted gene involved in endosperm development				yes
MOB1	Citterio et al. 2005 Cui et al. 2016 Vitulo et al. 2007	•Required for cytokinesis to occur		yes		
MPC	Tiwari et al. 2008	•Imprinted gene involved in transcriptional regulation of endosperm development				yes
MSI1	Guitton and Berger, 2005 Ingouff et al. 2005 Jullien et al. 2008 Kohler et al. 2003 Rodrigues et al. 2010	•WD40 repeat protein •P55 homologue from *Drosophila*			yes	

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
OSD1	Bao and Hua, 2004 Cromer et al. 2012 d'Erfurth et al. 2009 d'Erfurth et al. 2010 Mieulet et al. 2010	•Mutant allele leads to functional unreduced diploid male and female gametes •Mutant used in MiMe phenotype		yes		
***PCA21**	Singh et al. 2007	•Novel gene in *Pennisetum ciliare* •Differentially expressed b between apomictic and sexual ovules	yes	yes	yes	yes
***PCA24**	Singh et al. 2007	•Novel gene in *Pennisetum ciliare*	yes	yes	yes	yes
RDR6	Peragine et al. 2004 Qu et al. 2008 Rajeswaran et al. 2012	•Post transcriptional gene silencing •Promotes production of trans-acting siRNA's		yes	yes	yes
REC8/ SYN1	Bhatt et al. 1999 Brar et al. 2006 Cai et al. 2003 Golubovskaya et al. 2006 Grishaeva et al. 2007 Mieulet et al. 2016 Molnar et al. 1995 Watanabe and nurse, 1999 Zhang et al. 2006	•Mutant used in the MiMe phenotype •Orthologue of REC8 •Essential gene for meiosis		yes		
RETINOBLASTOMA RELATED (RBR)	Ach et al. 1997 Ebel et al. 2004 Jullien et al. 2008	•Mitosis of the developing megagametophyte is deregulated causing continuous growth within the embryo sac with excess nuclei forming •Autonomous endosperm formation occurs •Co-activates FIS2 •Represses cell proliferation •Crucial for controlling male and female gametogenesis			yes	yes
RKD1	Chardin et al. 2014 Koszegi et al. 2011	•RKD (A) subfamily transcription factor member			Yes	

Gene Name	Literature	Notes	Apomictic Plant Species Association	Apomeiosis	Parthenogenesis	Endosperm Development
SERK	Albertini et al. 2005 Cueva et al. 2012 Gruszczyńska and Rakoczy-Trojanowska, 2011 Hecht et al. 2001 Kantama et al. 2006 Podio et al. 2014 Santos and Aragão, 2009 Savona et al. 2012 Singla et al. 2009	•Somatic embryogenesis receptor kinase (SERK) There are two copies: SERK1 & SERK2 •Expression of SERK2 in *Paspalum notatum* is linked to onset of apomixis	yes		yes	
SPO11	Baudat et al. 2000 Cole et al. 2010 Hartung and Puchta , 2000 Hartung et al. 2007 Keeney et al. 1997 Klapholz et al. 1985 Merino et al. 2000 Puizina et al. 2004 Romanienko and Camerini-Otero, 2000 Stacey et al. 2006	•Generates double strand breaks for the initiation of meiosis •Mutant used in the MiMe phenotype •Meiosis specific protein		yes		
SPOROCYTELESS (SPL)/ NOZZLE	Bencivenga et al. 2012 Chen et al. 2014 Schiefthaler et al. 1999 Wei et al. 2015 Yang et al. 1999 Yuan, 2015	•MADS-box transcription factor •Crucial for male and female sporogenesis				
SSN1	Durrant et al. 2007 Ma et al. 2016 Richter et al. 2016 Serra et al. 2013 Wang et al. 2014	•Suppressor of SNI1 •Also known as RAD51D		yes		
***UPGRADE2**	Mau et al. 2013	•Associated with unreduced pollen in *Boechera* species	yes	yes		
WEE1	Beck et al. 2012 Cook et al. 2013 De Schutter et al. 2007 Gonzalez et al. 2004 Shimotohno et al.2006 Sorrell et al. 2002	•Overexpression inhibits cell division •Downregulation stimulates DNA replication •Regulates CDK's		yes	yes	
WUSCHEL	Chen et al. 2014 Lenhard et al. 2001 Nic-Can et al. 2013 Zuo et al. 2002	•Ectopic expression results in somatic embryo formation •Activates AGAMOUS			yes	

*Only found in apomictic species

4.2 **Microarray Analysis**

The top twenty probes identified as significantly differentially expressed in Table 4.3 showed that only nineteen out of twenty had a significance cutoff (unadjusted) value of $P < 0.05$. The adjusted P value, the P value resulting from background correction for multiple statistical tests, excluded all but three probes (Sharb1189846, Sharb0640787, and Sharb0123118) from the $P < 0.05$ threshold (Table 4.4). Despite these significance values, which can be highly influenced by sample size, all 20 probes were considered as secondary candidate genes in subsequent analyses.

Table 4.3 Microarray probe sequences of secondary candidate genes that were statistically isolated to be differentially expressed between sexual and apomictic megaspore mother cells.

Probe Name	60nt Probe Sequence (5' to 3')
Sharb1189846	GAAAAATCATTCCTCTGTCTCTTCCTGGTCAGTCGGTGAGCACTTTCGGTCATTTCCATT
Sharb0640787	AAATGATCTGTCTTGTACCGATCAATGGAGACACCACACTCATAATGTAACAGTCATCCT
Sharb0123118	TCAAGCACAAGCTCAGATTGGAGGCCTACCGTTATGAGGAAAAAAGGCTTTGCCAACTAT
Sharb0715958	TCATCCTGCGTGTATCGTTAAGACTTATGTTCTGATGTTGGTTAATAATTGTTGCTTATG
Sharb0931225	CTATCTGGGCGAATATTTGCAGGTGGTGTGCAAATTTAGTAATTTCAAGTTCATATTTGT
Sharb0892351	TCCCTGCGTGTTTTGACTCTGCTATGGCTTTTGATATTTAATGAAGAACGAGAATTCAAG
Sharb0547429	AGCGTCCCAACGACAAATATGCACAGATTCATTACAATATTGTGAAATGACAATGTCTAG
Sharb1697874	CAATTGACTCTCGATTCAAGCCTGTGTTCTCTGATAAGCGATTTACCATGGGTTCTGCTC
Sharb1113991	AGTCTTGTTTTTCAGACTAGTTTTCTCGAATAAAACCGAAACCGGGTAGAGATTGCATAA
Sharb0472474	CGTGTTTATTTGTCAGAACAATTTGTTGGAGTTTGTTAATTGTGGAAGAGTTTGGAGTTG
Sharb1068047	TTACGATGGAACTATGCGATTTGGTTCTACCCAGGAGGTGTGTTTTATTGTTCGGTTCCA
Sharb1650340	CCGGAAGTATGTGGATCATGTAGGTCGTCCAAAGATGAATATTGTTGTAGACATACCTCC
Sharb0565881	GTGTGTGTGTGTTCTTGAATAAGTCTTGGTTGTGAGTTCCAATTCAATGTAATGTTAACA
Sharb0487238	GTGTTGAGTGTGTTGGATTGATTCAGTTGAAATCAGTTCTCAACTCTGGATTCAATTAGG
Sharb0917101	GGAATCACTTAATCCTTCCTGCTTCCGGCTGACACAATATTTATCATTAGATCTATCATT
Sharb0595864	AAGGACAATGGCTTTGGATTTTTTCGATGTTGTATTGAGTTAGCATTGACAAAGTGTTTC
Sharb0789321	GCTTCTTCTTCTTCTGGTTGTTGTTGTTGTTGGAAACCTGATGTTGGAATCCAATAATAA
Sharb0895390	TGCCAAGTTCTGCCCCACAAGAGAAACGTCATCAACACTAACACTAGTGCAACGAAATCT
Sharb1459220	AATTTGCTGATACTGAGAGGCTAAATTGGGGTCAATGGGGAGTTTGCATTGCGATTGCGT
Sharb0444742	TCCCTGCGTGTTGCAAGACCTCTTCATATTACAATAACACTAATTATCAATGACCTAAAC

Table 4.4 Differential expression statistics for top 20 candidate probes. LogFC is the fold change statistic with positive values indicating upregulation and negative being downregulated in apomictic ovules. AveExpr is the mean log2 expression of a probe across all 18 arrays. t is the value derived from a moderated t-test statistic. The adjusted P value is the P value adjusted to control the false discovery rate across multiple samples via background correction.

Probe Name	logFC	AveExpr	t	P.Value	adj.P.Val
Sharb1189846[a]	2.59E+00	6.18E+00	8.32E+00	5.72E-08	4.33E-03
Sharb0640787[a]	2.37E+00	6.56E+00	8.03E+00	1.01E-07	4.33E-03
Sharb0123118[a]	1.49E+00	4.46E+00	7.92E+00	1.24E-07	4.33E-03
Sharb0715958	-2.51E+00	5.17E+00	-6.29E+00	3.65E-06	8.13E-02
Sharb0931225[a]	2.43E+00	4.00E+00	6.26E+00	3.87E-06	8.13E-02
Sharb0892351	-2.20E+00	6.59E+00	-6.12E+00	5.33E-06	9.34E-02
Sharb0547429	2.43E+00	5.06E+00	5.73E+00	1.25E-05	1.87E-01
Sharb1697874	1.71E+00	4.62E+00	5.52E+00	2.04E-05	2.68E-01
Sharb1113991	-1.37E+00	4.68E+00	-5.36E+00	2.92E-05	3.41E-01
Sharb0472474	-2.00E+00	6.21E+00	-5.23E+00	3.95E-05	4.15E-01
Sharb1068047	1.48E+00	3.86E+00	5.18E+00	4.37E-05	4.17E-01
Sharb1650340[a]	9.88E-01	4.10E+00	5.14E+00	4.82E-05	4.22E-01
Sharb0565881	-1.98E+00	5.85E+00	-5.09E+00	5.36E-05	4.33E-01
Sharb0487238	-1.94E+00	5.15E+00	-5.01E+00	6.49E-05	4.87E-01
Sharb0917101	-1.10E+00	3.99E+00	-4.95E+00	7.45E-05	4.91E-01
Sharb0595864	2.15E+00	3.99E+00	4.95E+00	7.48E-05	4.91E-01
Sharb0789321	2.38E+00	6.08E+00	4.89E+00	8.47E-05	5.22E-01
Sharb0895390[a]	1.50E+00	4.38E+00	4.87E+00	8.95E-05	5.22E-01
Sharb1459220	1.98E+00	5.64E+00	4.83E+00	9.77E-05	5.39E-01
Sharb0444742	-1.07E+00	4.05E+00	-4.81E+00	1.03E-04	5.39E-01

[a]Probes linked to primary candidate gene APOLLO

4.3 **Identification of APOLLO**

The adjustment of the P value via background correction proved to be very important in validation of the repeated analysis of the microarray data. The BLAST of probes to the WGS assemblies (Tables 4.5 and 4.6) resulted in several probes aligning to common contigs. In both the TAIR loci bulk retrieval ((Table 4.10) loci identified in Tables 4.7, 4.8, and 4.12) and BLASTn results of *B. stricta* transcripts (Table 4.11) submitted to the TAIR10 transcript database (Table 4.9), the orthologue of APOLLO (Corral et al. 2013) was identified as expected, and thus served as an internal positive control.

Table 4.5 BLASTn of candidate probes to a sexual *B. stricta* WGS assembly

Query	Number of hits	Lowest E-value	Accession (E-value)
Sharb1189846[a]	15	7.08E-24	contig_9533
Sharb0640787[a]	11	7.08E-24	contig_9533
Sharb1697874	14	7.08E-24	contig_390
Sharb1650340[a]	27	7.08E-24	contig_9533
Sharb0895390[a]	15	7.08E-24	contig_2911 contig_2910 contig_18391 contig_4886 contig_13489 contig_8009
Sharb1459220	10	8.63E-23	contig_2554
Sharb0123118[a]	6	3.67E-21	contig_9533
Sharb0931225[b]	34	8.08E-17	contig_13220
Sharb0547429	28	8.08E-17	contig_3215
Sharb0472474	42	4.18E-14	contig_2676
Sharb0565881[b]	22	1.46E-13	contig_6287
Sharb1068047[b]	5	5.10E-13	contig_8025
Sharb0917101	50	6.21E-12	contig_174
Sharb0789321	405	7.57E-11	contig_2442
Sharb0444742[b]	26	7.57E-11	contig_2283
Sharb0595864	30	9.22E-10	contig_3399
Sharb0715958[b]	30	3.22E-09	contig_3563 contig_9399
Sharb1113991	33	3.22E-09	contig_1643
Sharb0892351	38	1.37E-07	contig_2428
Sharb0487238[b]	19	4.77E-07	contig_14126

[a]Probes linked to APOLLO

[b]Probes with CNV between apomictic and sexual *Boechera*

Table 4.6 BLASTn of candidate probes to apomictic *B. divaricarpa* WGS assembly

Query	Number of hits	Lowest E-value	Accession (E-value)
Sharb1189846[a]	19	1.11E-23	contig_57299 contig_57298 contig_14417
Sharb0640787[a]	16	1.11E-23	contig_57299 contig_57298 contig_14417
Sharb0123118[a]	12	1.11E-23	contig_57299 contig_57298 contig_14417
Sharb1697874	16	1.11E-23	contig_45500
Sharb1068047[b]	8	1.11E-23	contig_32910 contig_27630 contig_97
Sharb1650340[a]	48	1.11E-23	contig_57299 contig_57298 contig_14417
Sharb0895390[a]	16	1.35E-22	contig_3140 contig_3245 contig_36745 contig_3585 contig_104589
Sharb1459220	16	1.35E-22	contig_36006
Sharb0931225[b]	62	4.72E-22	contig_13243 contig_4303
Sharb0547429	45	4.72E-22	contig_14892
Sharb0472474	90	6.56E-14	contig_31281
Sharb0565881[b]	33	2.29E-13	contig_36362 contig_36363
Sharb0917101	91	2.29E-13	contig_258
Sharb0789321	418	1.19E-10	contig_54780
Sharb0444742[b]	52	1.19E-10	contig_13797 contig_958
Sharb0715958[b]	51	5.05E-09	contig_32071
Sharb1113991	51	5.05E-09	contig_28058
Sharb0487238[b]	36	1.76E-08	contig_55638 contig_25269
Sharb0892351	56	2.15E-07	contig_51902

[a]Probes linked to APOLLO

[b]Probes with CNV between apomictic and sexual *Boechera*

Table 4.7 BLASTn results of candidate *B. stricta* CDS sequences submitted to TAIR10 CDS database

Query ID	Subject ID	% Identity	Alignment Length	Mismatches	Gap Openings	e-value	Bit Score
Bostr.1460s0103.1	AT3G01160.1	88.12	1608	137	10	0	1681
Bostr.19424s0368.2	AT3G22910.1	92.91	1369	94	1	0	1941
Bostr.19424s0368.2	AT3G22910.1	92.76	815	59	0	0	1148
Bostr.19424s0368.1	AT3G22910.1	91.49	1646	128	2	0	2157
Bostr.19424s0368.1	AT3G22910.1	92.91	1369	94	1	0	1941
Bostr.3288s0084.1	AT1G74390.1	88.77	1514	140	3	0	1683
Bostr.3288s0084.1	AT1G74390.2	86.72	956	97	3	0	918

Table 4.8 BLASTp results of candidate *B. stricta* protein sequences submitted to TAIR10 protein database

Query ID	Subject ID	% Identity	Alignment Length	Mismatches	e-value	Bit Score
Bostr.19424s0368.2.p	AT3G22910.1	94.23	745	40	0	1323
Bostr.19424s0368.1.p	AT3G22910.1	93.54	1022	59	0	1777
Bostr.19424s0368.1.p	AT3G63380.1	66.8	1000	312	0	1207
Bostr.3288s0084.1.p	AT1G74390.1	85.38	506	64	0	800
Bostr.3288s0084.1.p	AT1G74390.2	83.2	506	64	0	768
Bostr.1460s0103.1.p	AT3G01160.1	82.88	514	73	0	638

Table 4.9 *Arabidopsis thaliana* loci and their corresponding transcript, protein, and CDS candidate sequences from *B. stricta*

***B. stricta* CDS**	***B.stricta* Protein**	***B. stricta* Transcripts**
AT3G01160.1	AT3G22910.1	AT5G27770.1
AT3G22910.1	AT3G63380.1	AT3G05560.3
AT1G74390.1[a]	AT1G74390.1[a]	AT3G05560.2
AT1G74390.2[a]	AT1G74390.2[a]	AT3G05560.1
	AT3G01160.1	AT5G19510.1
		AT5G19500.1
		AT5G12110.1
		AT2G43460.1
		AT3G59540.1
		AT2G36160.1
		AT3G52580.1
		AT3G11510.1
		AT3G22910.1
		AT5G50350.1
		AT1G74390.1[a]
		AT1G74390.2[a]
		AT3G01160.1
		AT3G28500.1
		AT2G02130.1
		AT2G02100.1
		AT2G02120.1

[a] Orthologue to primary candidate gene APOLLO

Table 4.10 Identified *Arabidopsis orthologues* of *Boechera* cDNA containing microarray probe alignment(s).

Up/Down Regulated in Apomictic	Sharb ID	Locus Identifier	Gene Description
Down	Sharb0444742	AT5G19510.1	Translation elongation factor EF1B/ribosomal protein S6 family protein; FUNCTIONS IN: translation elongation factor activity; INVOLVED IN: translational elongation, defense response to bacterium.
Down	Sharb0472474	AT3G11510.1	Ribosomal protein S11 family protein; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation; LOCATED IN: cytosolic small ribosomal subunit, cytosolic ribosome, nucleolus.
Down	Sharb0472474	AT2G36160.1	Ribosomal L22e protein family; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation.
Up	Sharb1650340[a] Sharb0123118[a] Sharb1189846[a] Sharb0640787[a]	AT1G74390.1	Polynucleotidyl transferase, ribonuclease H-like superfamily protein; FUNCTIONS IN: exonuclease activity, nucleic acid binding.
Down	Sharb0487238	AT2G43460.1	Ribosomal L38e protein family; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation, ribosome biogenesis.
Down	Sharb1113991	AT3G05560.1	Ribosomal L22e protein family; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation
Up	Sharb1697874	AT3G01160.1	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown.CONTAINS InterPro DOMAIN/s: NUC153 (InterPro:IPR012580)

Up/Down Regulated in Apomictic	Sharb ID	Locus Identifier	Gene Description
Up	Sharb0547429	AT5G50350.1	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to oxidative stress; LOCATED IN: chloroplast.
Up	Sharb1459220	AT3G63380.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein; FUNCTIONS IN: calcium-transporting ATPase activity, calmodulin binding; INVOLVED IN: calcium ion transport, cation transport, metabolic process, ATP biosynthetic process.
Down	Sharb0444742	AT5G19500.1	Encodes a putative amino acid transporter that localizes to the chloroplast inner envelope membrane.
Down	Sharb0487238	AT3G59540.1	Ribosomal L38e protein family; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation, ribosome biogenesis
Up	Sharb145920	AT3G22910.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein; FUNCTIONS IN: calcium-transporting ATPase activity, calmodulin binding; INVOLVED IN: cation transport, calcium ion transport, metabolic process, ATP biosynthetic process.
Down	Sharb0715958	AT3G28500.1	60S acidic ribosomal protein family; FUNCTIONS IN: structural constituent of ribosome.
Down	Sharb1113991	AT5G27770.1	Ribosomal L22e protein family; FUNCTIONS IN: structural constituent of ribosome.
Down	Sharb0444742	AT5G12110.1	Glutathione S-transferase, C-terminal-like;Translation elongation factor EF1B/ribosomal protein S6; FUNCTIONS IN: translation elongation factor activity.

Up/Down Regulated in Apomictic	Sharb ID	Locus Identifier	Gene Description
Down	Sharb0472474	AT3G52580.1	Ribosomal protein S11 family protein. BEST Arabidopsis thaliana protein match is: Ribosomal protein S11 family protein (TAIR:AT3G11510.1)
Down	Sharb0565881	AT2G02120.1	Predicted to encode a PR (pathogenesis-related) protein. Belongs to the plant defensin (PDF) family
Down	Sharb0487238	AT2G02130.1	Predicted to encode a PR (pathogenesis-related) protein. Belongs to the plant defensin (PDF) family
Down	Sharb0565881	AT2G02100.1	Predicted to encode a PR (pathogenesis-related) protein. Belongs to the plant defensin (PDF) family

[a]Probes linked to APOLLO

Table 4.11 BLASTn of candidate probes to *B. stricta* transcript database

Query	Number of hits	Lowest E-value	Accession
Sharb1189846[a]	19	1.79E-24	Bostr.3288s0084.1
Sharb1697874	5	1.79E-24	Bostr.1460s0103.1
Sharb1650340[a]	13	1.79E-24	Bostr.3288s0084.1
Sharb1459220	13	2.18E-23	Bostr.19424s0368.2 Bostr.19424s0368.1
Sharb0547429	11	7.62E-23	Bostr.15774s0182.1
Sharb0123118[a]	13	3.95E-20	Bostr.3288s0084.1
Sharb0472474	27	1.06E-14	Bostr.23794s0481.1
Sharb1068047	3	1.29E-13	Bostr.13671s0403.1
Sharb0565881	13	4.51E-13	Bostr.0556s0037.1 Bostr.055620039.1
Sharb0444742	11	1.92E-11	Bostr.26527s0249.1
Sharb0640787[a]	10	8.15E-10	Bostr.3288s0084.1
Sharb0715958	18	8.15E-10	Bostr.0556s0507.1
Sharb1113991	13	8.15E-10	Bostr.29827s0100.1
Sharb0892351	26	3.46E-08	Bostr.15697s0322.1
Sharb0487238	15	1.21E-07	Bostr.25993s0435.1
Sharb0789321	445	2.19E-04	Bostr.26675s0094.1

[a]Probes linked to APOLLO

Table 4.12 BLASTn results of candidate *B. stricta* transcripts submitted to TAIR10 transcript database

Query ID	Subject ID	% Identity	Alignment Length	Mismatches	Gap Openings	e-value	Bit Score
Bostr.29827s0100.1	AT5G27770.1	90.47	556	35	5	0	668
Bostr.29827s0100.1	AT3G05560.3	89.11	349	38	0	9.00E-108	391
Bostr.29827s0100.1	AT3G05560.2	89.11	349	38	0	9.00E-108	391
Bostr.29827s0100.1	AT3G05560.1	89.11	349	38	0	9.00E-108	391
Bostr.26527s0249.1	AT5G19510.1	91.82	880	68	2	0	1162
Bostr.26527s0249.1	AT5G19500.1	88.29	444	51	1	2.00E-128	460
Bostr.26527s0249.1	AT5G12110.1	85.63	348	50	0	2.00E-78	293
Bostr.26527s0249.1	AT5G12110.1	81.16	207	39	0	2.00E-20	101
Bostr.25993s0435.1	AT2G43460.1	95.22	314	13	1	4.00E-140	498
Bostr.25993s0435.1	AT2G43460.1	91.73	133	8	3	3.00E-36	153
Bostr.25993s0435.1	AT3G59540.1	89.78	225	23	0	1.00E-69	264
Bostr.23794s0481.1	AT2G36160.1	90.28	607	52	4	0	710
Bostr.23794s0481.1	AT3G52580.1	91.97	461	37	0	5.00E-177	620
Bostr.23794s0481.1	AT3G11510.1	87.64	453	56	0	7.00E-127	454
Bostr.19424s0368.2	AT3G22910.1	92.91	1369	94	1	0	1941
Bostr.19424s0368.2	AT3G22910.1	92.56	1357	101	0	0	1889
Bostr.19424s0368.1	AT3G22910.1	91.49	1646	128	2	0	2157
Bostr.15774s0182.1	AT5G50350.1	90.54	1829	146	6	0	2248
Bostr.13671s0403.1	AT5G46240.1	100	20	0	0	0.083	40.1
Bostr.3288s0084.1	AT1G74390.1	88.59	1709	162	4	0	1867
Bostr.3288s0084.1	AT1G74390.2	92.26	530	41	0	0	726
Bostr.3288s0084.1	AT1G74390.2	86.75	1147	119	4	0	1094
Bostr.3288s0084.1	AT1G74390.2	92.26	530	41	0	0	726
Bostr.1460s0103.1	AT3G01160.1	88.15	1620	138	10	0	1697
Bostr.1460s0103.1	AT3G01160.1	90.73	496	34	2	3.00E-156	553
Bostr.0556s0507.1	AT3G28500.1	89.58	355	28	1	4.00E-116	418
Bostr.0556s0039.1	AT2G02130.1	87.12	365	31	6	1.00E-87	323
Bostr.0556s0039.1	AT2G02100.1	84.01	369	37	7	5.00E-62	238
Bostr.0556s0039.1	AT2G02120.1	89.01	191	21	0	3.00E-54	212
Bostr.0556s0037.1	AT2G02100.1	86.74	445	43	4	1.00E-112	406
Bostr.0556s0037.1	AT2G02130.1	86.5	400	44	5	3.00E-91	335
Bostr.0556s0037.1	AT2G02120.1	89.47	228	24	0	4.00E-69	262

4.4 **Presence and Absence of Probes in Sexual and Apomictic Genomes**

There are several probes that did not match to both the sexual and apomictic WGS libraries. Probe Sharb0595864, which was only found in the sexual *B. stricta* WGS library (Table 4.5), was submitted to NCBI BLASTn where the gene (contig 3399) matched a sequence from the third chromosome of *Arabidopsis thaliana* (Table 4.15). The gene alignment matched that of AT3G18430, a calcium binding EF hand binding protein.

The probe Sharb0715958 was homologous to two contigs (contigs 3563 and 9399) in the sexual *B. stricta* WGS database (Table 4.5), while in the apomictic WGS database (Table 4.6) it showed homology to only one contig (contig 32071). Upon searching for matches in the sexual *B. stricta* databases ((CDS, transcript, protein) Tables 4.13, 4.14, and 4.15), the results indicated Sharb0715958 had a single transcript match (Table 4.9), thus supporting the observation for a single copy in the sexual WGS database. When the transcript was submitted to TAIR BLASTn the only locus retrieved was AT3G28500 (Table 4.12), a 60s acidic ribosomal family protein (Table 4.10). Contig 3563 (Table 4.5) and 32071 (Table 4.6) have the same BLASTn description linking them to a sequence in *Capsella rubella* (Table 4.15 and Table 4.16) while the outlier, contig 9399, has an orthologue on *Arabidopsis thaliana's* second chromosome (Table 4.15). Alignment through NCBI determined that contig 9399 matched AT2G05520, a glycine rich protein 3.

Table 4.13 BLASTn of candidate probes to *B. stricta* CDS database

Query	Number of hits	Lowest E-value	Accession
Sharb1697874	5	1.43E-24	Bostr.1460s0103.1
Sharb1650340[a]	11	1.43E-24	Bostr.3288s0084.1
Sharb1459220	12	1.74E-23	Bostr.19424s0368.2 Bostr.19424s0368.1
Sharb0123118[a]	13	3.14E-20	Bostr.3288s0084.1
Sharb1068047	3	1.03E-13	Bostr.13671s0403.1
Sharb0892351	19	1.43E-05	Bostr.7365s0038.1
Sharb0789321	466	1.74E-04	Bostr.26675s0094.1

[a]Probes linked to APOLLO

Table 4.14 tBLASTx of candidate probes to *B. stricta* protein database

Query	Number of hits	Lowest E-value	Accession
Sharb1459220	8	1.14E-08	Bostr.19424s0368.1.p Bostr.19424s0368.2.p
Sharb1650340[a]	8	1.13E-07	Bostr.3288s0084.1.p
Sharb0123118[a]	6	2.18E-06	Bostr.3288s0084.1.p
Sharb1697874	1	1.66E-05	Bostr.1460s0103.1.p
Sharb1068047	2	8.76E-03	Bostr.13671s0403.1.p

[a]Probes linked to APOLLO

Table 4.15 BLASTn results of candidate sequences from sexual *B. stricta* WGS submitted to NCBI nucleotide database with the taxonomic filter limiting results to Virdiplantae

Query	Number of hits	Lowest E-value	Accession (E-value)	Greatest identity %	Description
contig_1416	963	0	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_1349	1200	0	KF705603	100	Boechera sp. IPK Bsp 9 isolate BAC2b_4 APOLLO gene, complete cds
contig_1320	184	0	HF954100	100	Boechera divaricarpa BAC clone C8B11
contig_9533	201	0	KF705604	100	Boechera sp. IPK Bsp 9 isolate BAC5_7 APOLLO gene, complete cds
contig_9399	408	5.28E-107	XM_010490463	97.5	Arabidopsis thaliana chromosome 2, complete sequence
contig_8025	10417	0	AC241020	100	Arabidopsis thaliana chromosome 3, complete sequence
contig_8009	206	5.37E-172	KF705603	96.55	Boechera sp. IPK Bsp 9 isolate BAC2b_4 APOLLO gene, complete cds
contig_6287	619	0	HM222924	100	Thellungiella parvula phosphatase, cyclin J18, clathrin assembly protein-related protein, BPC1, SGR5/ATIDD15, leucine-rich receptor kinase, endomemebrane protein 70, salt overly sensitive 1, glutamate decarboxylae 4, oligopeptide transport protein, peptide transporter, and NADH-ubiquinone oxidoreductase B18 genes, complete cds; and unknown genes
contig_4886	3114	0	KF705603	96.15	Boechera sp. IPK Bsp 9 isolate BAC2b_4 APOLLO gene, complete cds
contig_3563	1328	0	AC254580	100	Capsella rubella clone JGIBSIA-5G22, complete sequence
contig_3399	579	0.00E+00	CP002686	100	Arabidopsis thaliana chromosome 3, complete sequence
contig_3215	585	0	CP002688	100	Arabidopsis lyrata haplotype Aly-S50 S-locus region genomic sequence
contig_2911	269	9.02E-112	KF705603	96.3	Arabidopsis thaliana chromosome 5 sequence
contig_2910	761	0	CP002685	97.5	Arabidopsis thaliana chromosome 2, complete sequence
contig_2676	588	0	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_2554	379	0	CP002686	98.88	Arabidopsis thaliana chromosome 3, complete sequence
contig_2283	631	0	CP002688	99.03	Arabidopsis thaliana chromosome 5 sequence
contig_1643	816	0	CP002688	100	Arabidopsis thaliana chromosome 5 sequence
contig_390	1070	0	CP002686	100	Arabidopsis thaliana chromosome 3, complete sequence
contig_174	6715	0	JX205495	100	Aethionema grandiflorum chloroplast DNA, complete sequence

Table 4.16 BLASTn results from candidate sequences from apomictic *B. divaricarpa* WGS submitted to NCBI nucleotide database with the taxonomic filter limiting results to Virdiplantae

Query	Number of hits	Lowest E-value	Accession (E-value)	Greatest identity %	Description
contig_57299	148	0	KF705604	100	Boechera sp. IPK Bsp 9 isolate BAC5_7 APOLLO gene, complete cds
contig_57298	149	0	KF705604	100	Boechera sp. IPK Bsp 9 isolate BAC6 APOLLO gene, complete cds
contig_55638	291	3.48E-133	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_51902	218	0	CP002684	96.37	Arabidopsis thaliana chromosome 1 sequence
contig_45500	108	1.39E-155	XM_006297777	100	PREDICTED: Camelina sativa pre-rRNA-processing protein esf1-like (LOC104707960), mRNA
contig_36745	15	0.72	XR_731699	95.83	PREDICTED: Brachypodium distachyon BTB/POZ and MATH domain-containing protein 2-like (LOC100827490), misc_RNA
contig_36363	554	0	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_36362	501	0	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_36006	333	0	CP002686	95.37	Arabidopsis thaliana chromosome 3, complete sequence
contig_32910	292	0	XM_010497370	93.1	Arabidopsis thaliana chromosome 3, complete sequence
contig_32071	626	0	AC254580	100	Capsella rubella clone JGIBSIA-5G22, complete sequence
contig_28058	222	5.01E-96	CP002688	98.08	Arabidopsis thaliana chromosome 5 sequence
contig_27630	1301	0	JX185681	100	PREDICTED: Brassica napus uncharacterized LOC106378981 (LOC106378981), mRNA
contig_25269	308	7.74E-132	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_14892	756	0	CP002688	98.52	Arabidopsis thaliana chromosome 5 sequence
contig_14417	220	0	KF705605	100	Boechera sp. IPK Bsp 9 isolate BAC6 APOLLO gene, complete cds
contig_13797	472	0	CP002688	100	Arabidopsis thaliana chromosome 5 sequence
contig_13243	308	0	HF954101	100	Boechera divaricarpa BAC clone E7K5
contig_4303	269	0	HF954100	100	Boechera divaricarpa BAC clone C8B11
contig_3585	85	0.04	CP002685	100	Arabidopsis thaliana chromosome 2, complete sequence
contig_3245	351	0	KF705603	96.3	Boechera sp. IPK Bsp 9 isolate BAC2b_4 APOLLO gene, complete cds
contig_3140	347	1.39E-109	KF705603	96.3	Boechera sp. IPK Bsp 9 isolate BAC2b_4 APOLLO gene, complete cds
contig_958	476	0	CP002688	100	Arabidopsis thaliana chromosome 5 sequence
contig_258	6638	0	JX205495	100	Cardamine resedifolia plastid, complete genome
contig_96	276	1.10E-106	AC241133	96.3	Brassica rapa subsp. pekinensis clone KBrH046K16, complete sequence

4.5 **Copy Number Variation of Secondary Candidate Genes**

There are four probes with no homology to APOLLO that were characterized by an increase (i.e. >1) in matches to the apomictic WGS library (Table 4.17). The four probes; Sharb0931225, Sharb0565881, Sharb0444742, and Sharb0487238 had two significant matches in the apomictic WGS library, and only one in the sexual WGS library. In all cases but Sharb0565881, the sexual and apomictic WGS homologues to each probe also had matching annotations in NCBI BLASTn analysis.

The contig sequence from the sexual WGS, which had homology to probe Sharb0565881, was resubmitted to NCBI BLASTn and aligned with locus AT2G02090, which contains three domains; a Helicase superfamily c-terminal domain, a DEAD-like helicase superfamily domain, and a SNF2 family N-terminal domain (Marchler-Bauer et al. 2014). Probe Sharb0565881 has two homologous contigs from the apomictic WGS (contigs 36362 and 36363 (Table 4.6)), both of which are linked to a different locus in *Arabidopsis thaliana* (AT2G02100, encoding a member of the plant defensin family of proteins) (Table 4.10).

Table 4.17 CNV between apomictic and sexual *Boechera* secondary candidate probes not associated with APOLLO

Gene	Apo WGS	Sexual WGS
Sharb0715958	1	2
Sharb0487238	2	1
Sharb0444742	2	1
Sharb1068047	3	1
Sharb0565881	2	1
Sharb0931225	2	1

4.6 Is Gene Regulation Influenced by Copy Number Variation?

Out of the three probes that were characterized by a hypothesized duplication in the apomict WGS library and for which the duplicated copies corresponded to the same annotation, only the expression of probe Sharb0931225 showed that upregulation of a gene product in a cell is positively correlated with the number of copies of the gene in the genome.

Gene duplication does not always imply the result will be two fold expression. There are many outcomes resulting from gene duplications such as subfucntionalization, neofunctionalization, and gene loss. Gene duplications rarely occur in tandem pairs on homologous chromosomes and more than often end up in an gene being established elsewhere in the genome with no location based homologous template for crossing over to occur. With no homologous pair to facilitate recombination a duplicated gene accumulates mutations which can affect gene expression, protein function, and downstream genetic factors.

4.7 Preliminary Identification of a Gene only Found in Boechera

The BLAST tables involving microarray probe Sharb1068047 stood out. A strong probe alignment was obtained when matching the probe to all DNA libraries sourced from *Boechera*, but any attempt to find orthologues in other species resulted in nothing. The transcript and coding sequence bearing the identity Bostr.13671s0403.1 is the only hit and was submitted to TAIR, Phytozome, NCBI, Pfam, and Ensembl with multiple attempts using various settings, and no orthologues emerged. Hence, it is hypothesized that this gene is completely novel to *Boechera* and found in no other existing genome (all BLAST searches started with virdiplantae filter which after no results in Phytozome and NCBI all organism filters were removed) that has been sequenced thus far. Very interestingly, there is not a single recognizable known protein domain, yet this has a transcript and coding sequence that can be confirmed by the *Boechera stricta* genome published

on Phytozome. If future studies confirm the involvement of this gene in reproductive development, I propose this gene be named after a Celtic deity of fertility- Damara.

5 Discussion

Through extensive literature review a list of known apomixis candidate genes has been compiled which creates a reference for associating genes with megasporogenesis, fertilization, and endosperm/embryo development. Two primary candidate genes underlying apomixis in *Boechera* have already been identified (APOLLO, UPGRADE2), but through reanalyzing the microarray data from the Sharbel lab used in the discovery of APOLLO (Corral et al. 2013) with the latest genomic data from multiple *Boechera* species, the top 20 statistically significant probes showing differential expression between apomictic and sexual *Boechera* ovules could be further investigated. Secondary candidate genes have yet to be identified that may aid the genetic engineering of apomixis into crop plants. Through preliminary bioinformatics, the top twenty probes identified by statistical analysis were characterized.

5.1 Current Methodology for Investigating Apomixis

A common workflow for identifying gene function and further interactions is the initial discovery of a candidate gene in an apomictic species, followed by identification and analysis of its orthologue in a sequenced genome, which has typically been *Arabidopsis thaliana*. Transformation and regeneration protocol development in these often-underdeveloped apomictic model plants requires experienced technicians and years of testing for success, hence using a proven model is faster and easier. Candidate genes for apomixis involve a range of functions because in addition to deviations in meiosis there also has to be stable endosperm development, which is additionally perturbed during apomictic seed formation. For a true apomictic seed, the female gamete has to undergo apomeiosis followed by parthenogenetic development of the unreduced egg cell to yield and embryo with the same ploidy level as the mother plant. Endosperm development of an apomict can occur with (pseudogamy) or without fertilization (autonomous

endosperm formation) (Gerashchenkov et al. 2015). Due to multiple pathways for endosperm development, ploidy level variation in this normally sexually-derived tissue can arise within natural populations of facultative apomicts, and provides variation which can be used to test the parental conflict theory and endosperm balance number (Vinkenoog and Scott, 2001).

5.2 **Apomixis is Multidimensional**

Considering the complex nature of apomictic seed development, it is unlikely to be under monogenic control, and furthermore the variation associated with apomictic seed development (e.g. Aliyu et al., 2010) introduces background noise in identifying underlying factors leading to apomixis. Apomixis-specific genes, genes with no identified ancestry in sexual plants, are of special interest because of their potential association with apomixis, but in reality the formation of new genes is evolution in action. The novel genes associated with apomixis (Table 4.2) are potentially important, although functional data supporting their role is still required before their eventual transformation into crop plants. In addition to known novel genes there are known apomixis specific loci that are not described in either Table 4.1 or 4.2 because the gene products within the apomixis specific region(s) of the genome have yet to be identified (Schallau et al. 2010; Vijverberg et al. 2010). A common characteristic of apomixis are apomixis specific genomic regions that are non-recombining and hemizygous in nature, which can harbor genetic elements that affect reproduction (Barcaccia and Albertini, 2013). Several novel genes identified in Table 4.2 such as UPGRADE2, BBM-like, PAC21 and PAC24 in addition to allelic variants such as BBM-like (Conner et al. 2015) can be associated with such genomic regions because of mutation accumulation in the absence of recombination leading to the creation of novel variants (Campos et al. 2012). Besides the identification of a novel gene in an apomictic species, it is expected that experiments such as genome walking and inverted PCR can be used to identify flanking regions

around the gene of interest to identify other genomic elements of interest (Reddy et al. 2008; Triglia, 2000).

Apart from novel genes, many of the candidates identified in Table 4.1 are from sexual plant species, whereby they play important roles in female gamete development and/ or zygote maturation. Through functional genomic studies, namely in *Arabidopsis thaliana*, phenotypes of several mutants produce apomixis-like phenotypes except that they fail to produce fully functional offspring without lethal consequences (Dresselhaus et al. 2001). Attempts have been made at using several gene mutants to artificially induce apomixis in *Arabidopsis* through the MiMe (Mitosis instead of Meiosis) phenotype (d'Erfurth et al. 2009). The MiMe phenotype uses the triple mutant of *Atspo11-1, Atrec8,* and *AtOSD1* to prevent chromosome pairing during the second division in meiosis and have a 2n female gamete mitotically proliferate into a functional zygote. The triple mutant described was recently implemented into rice with success under controlled conditions (Mielut et al. 2016). Given the success however, there is a drawback in that the triple mutants still accepted pollen, which would contribute to ploidy variation in offspring.

Transcription factors have the inherit ability to cause major changes in cells, but further add to the concept of apomixis being immensely complex. Because transcription factors often influence multiple genes, a cause and effect must be identified for the regulation of the transcription factor, its effect on primary targets, and the following downstream interactions. This encompasses the dilemma of functional genomics involving highly regulated processes that occur once per offspring in a plant's lifecycle.

5.3 Emergence of BABY BOOM as a Significant Apomictic Factor for Asexuality in Field Crops

BABY BOOM (BBM) has been making waves in the world of reproductive biology after successful transgenic events in rice resulting in apomictic events with endosperm development (Khanday et al. 2019). Previously, BBM had been making an impact in transformation studies since its discovery by Boutilier et al. 2002 through studying the embryonic pathway in *Brassica napus*. Phenotypes of BBM overexpression in both *Arabidopsis* and *Brassica* showed hormone free regeneration of explants in addition to the spontaneous formation of somatic embryos on cotyledons. Since publication in 2002, there have been several studies published in a range of plants where the orthologous genes for BBM in *Arabidopsis* have been cloned and expressed in explants, resulting in significant improvements to plant transformation recovery. Several plants such as the Chinese white Poplar (Deng et al. 2009), chocolate tree (Florez et al. 2015), Sweet peppers (Heidmann et al. 2011), Tobacco (Srinivasan et al. 2007), and Maize (Lowe et al. 2016) all showed consistent phenotypes that showed it is possible to increase the recovery rate of transformed explants. In addition to Maize, Lowe et al. 2016 showed that BBM in combination with WUSCHEL (another important transcription factor involved in plant development (Lenhard et al. 2001; Zuo et al. 2002; Nic-Can et al. 2013; Cheng et al. 2014)) were able to significantly increase the recovery rate of transformed explants. The results from Lowe et al. 2016 have changed DuPont Pioneer's approach to transforming maize, sorghum, sugarcane, and indica rice.

In the apomictic plant *Pennisetum squamulatum*, a BBM-like gene has been identified in an asexual genomic region (ASGR) of the genome, and has a role in apomixis (Conner et al. 2008, 2015). The expression of ASGR-BABY BOOM-like before fertilization was shown to promote embryonic parthenogenesis where the unreduced aposporous initial cell goes through three rounds

of mitosis, resulting in eight nuclei of the same ploidy (Figure 1.4). At this point BABY BOOM is capable of triggering parthenogenesis in an aposporous plant, but in the big picture of transferring apomixis into sexual crop plants, BABY BOOM does not address the production of an unreduced megaspore mother cell or the triggering of endosperm development. This all changed with the use of an apomeiosis phenotype termed MiMe (mitosis instead of meiosis) and another discovery involving BABY BOOM in the male gametophyte.

The MiMe phenotype is a triple mutant system using OSD1 (omission of second division 1; d'Erfurth et al. 2009), Rec8 (meiotic recombination protein; Chelysheva et al. 2005), and SPO11 (meiotic recombination protein; Hartang et al. 2007). This system was first experimented with in *Arabidopsis thaliana*, but phylogenetic conservation of all three genes advocates that MiMe should be possible to achieve in several crop species (Mieulet et al. 2016). The MiMe phenotype is attained through the replacement of meiosis I with a mitosis-like division meditated by Rec8 and SPO11 mutants, which eliminate recombination to ensure that the two cells at prophase II are genetic copies of the mother plant. Any further progression into meiosis II is aborted due to OSD1 ensuring that the two megaspore mother cells are all that remain for degenerative selection. Therefore, the triple mutant MiMe system achieves production of an unreduced megaspore mother cell.

Measuring gene expression of an unfertilized egg cell is a difficult task but was only the beginning for Khanday et al. 2019, who found that BABYBOOM1 is an embryonic proliferation trigger expressed by the fertilizing male sperm cell. BBM until recently has always been associated with the female gamete, as previously mentioned with its ability to initiate embryo formation from somatic tissue. When BBM is expressed periodically in a fertilized or unfertilized egg cell, it activates a developmental checkpoint in the zygote, allowing for cell division and proliferation to

an embryo. By combining egg cell specific expression of BBM1 with the MiMe phenotype from Mieulet et al. 2016, Khanday et al. 2019 were able to produce gametophytic apomictic embryos and a phenotype transmissible through several generations. However, not to be overlooked in Khanday et al. 2019 is the issue of endosperm development. BBM does not play a role in promoting endosperm development, which is a crucial component in seed viability. For viable seed to be produced using this synthetic genetically engineered apomictic system, fertilzation of the central cell leads to variability in the seed ploidy ratio (embryo:endosperm) of 2:5 (haploid pollen) or 2:6 if the MiMe system also acts on microgametogenesis to yield diploid pollen. It is not the perfect 2:4 apomict that is desired, but it works!

5.4 AT2G02090, a Locus Important to DNA Repair and Recombination Identified in Sexual *B. stricta* was not Identified in Apomictic *B. divaricarpa*

The probe Sharb0565881 was identified in both the sexual and apomictic WGS results, but the matches between libraries did not result in identical loci. In the sexual WGS library, Sharb0565881 was homologous to a locus with biological importance in DNA repair and recombination, while it was homologous to plant defensin proteins in apomictic *Boechera* (Table 4.10).

The orthologue from *Arabidopsis thaliana* matching Sharb0565881 in the sexual WGS library is AT2G02090. This locus contains a helicase superfamily C-terminal domain, DEAD-like helicase superfamily domain, and a SNF2 family N-terminal domain. The domains contained within this locus play a significant role in DNA repair, DNA unwinding, transcriptional regulation, and chromatin unwinding (Schmid and Linder, 1992; Schmidt et al. 2011; Shaked et al. 2006). The absence of this gene could very well be another significant factor in the cascade of events leading to a faulty meiosis in *Boechera* megaspore mother cells.

DNA repair is an essential process in eukaryotic cells. Damage to the DNA of both somatic and reproductive cells occurs as a continual process, and in order for a cell to function properly the DNA must be repaired in order for a gene to code for the proper gene product. There are several types of DNA errors (Harper and Elledge, 2007), but in apomicts the absence of a DNA repair gene brings light to a notable hypothesis. If a DNA repair gene is absent it is assumed that DNA repair is inhibited, thus possibly contributing to mutation accumulation over time, which is consistent with Muller's ratchet (Muller, 1964).

Muller's ratchet is famously known for explaining why sexual reproduction is advantageous. Genetic mutations occur randomly and can have a positive or negative effect on the fitness of an organism, and 99% of the time these mutations have slightly negative (i.e. deleterious) effects (Sharp and Agrawal, 2018). Sexually reproducing organisms undergo recombination and syngamy (i.e. joining of male and female gametes) which can occasionally segregate different deleterious mutations into a single offspring, leading to death of that offspring and purging of the mutations from the population. With asexually reproducing organisms, mutations can never be segregated, and as such would continue to accumulate and be fully replicated in the successive progeny. Each time a mutation occurs, adding to the mutations already present, Muller's ratchet clicks to a higher mutational class, a process which will continue through time. Deleterious mutation accumulation and the fitness of an organism have two proposed models, (i) the first being a direct inverse linear relationship between mutation accumulation and decreasing fitness, (ii) the second being mutation accumulation occurs with no effect on fitness until it reaches a threshold (point of departure)(Lynch et al. 1995; Keightley and Lynch, 2003).

5.5 Is Downregulation of Ribosomal Protein Genes Linked to Upregulated Expression of a Novel Nucleolar Protein Family?

Downregulation of several ribosomal proteins (L22e, S11, L38e, S6, and 60S) in apomictic *Boechera* (Table 4.10) could have the ability to induce genome wide gene regulation changes. The 60s large subunit and the 40s subunit connect to form the 80s ribosome (Ceci et al. 2003), and the ribosomal L22e and L38e family proteins are subunits of the 60s large subunit while S6 and S11 are subunits of the 40s small subunit. Interestingly, the downregulation of ribosomal proteins measured here is consistent with the genome wide pattern of gene downregulation measured in apomictic ovules (Sharbel et al. 2009, 2010).

In cell biology, the nucleolus is where ribosomal subunits are manufactured (Kobayashi, 2008). The identification of a novel nucleolar protein family member containing a NUC153 domain that is upregulated in apomictic *Boechera* is an interesting find considering the identification of several ribosomal sub units (Table 4.10). It is unclear whether the upregulation directly influences the formation of ribosomal subunits L22e, S11, L38e, S6, and 60S or is involved in other processes currently unknown.

5.6 Downregulation of Plant Defensin Proteins in Apomictic Boechera

Another trend identified in the bulk retrieval of locus descriptors from TAIR (Table 4.10) is the downregulation of pathogenesis related proteins of the plant defensin family (PDF) in apomictic *Boechera*. Plant defensin proteins enable a host plant to respond to pathogen infection (Penninckx et al. 1996). In sexually reproducing organisms the evolutionary arms race between pathogen and host is constantly evolving, whereas in apomictic plants that reproduce clonally a pathogen can adapt to particular genotypes, leading to their extinction. This is explained by the Red Queen hypothesis (Pearson, 2001).

The Red Queen hypothesis is best described as antagonistic coevolution, which is thought to favor sexual reproduction. By creating genetically variable offspring, the imposed selection on particular "weak" genotypes by pathogens would lead to continuous evolution of the host through sex, securing its existence. Pathogens must similarly change over generations to adapt to genetically different host genotypes. What happens when there is no sex? Asexual organisms are at a disadvantage when faced with a pathogen, since the pathogen is expected to adapt to the unchanging asexual genotypes which cannot escape, leading to a decrease in their fitness or extinction.

In sexually reproducing plants, it would be assumed that PDF proteins would be upregulated, being involved with maintenance of plant defenses in a variable environment. Downregulation of said genes in asexually reproducing plants is consistent with both Muller's ratchet and the Red Queen. Loss of function of plant defensin genes could be attributed to general mutation accumulation (i.e. knockout of gene function), while downregulation of PDF proteins in apomictic *Boechera* is an indicator that asexual lineages are increasingly susceptible to pathogenic infection, as would be expected in the Red Queen hypothesis.

5.7 Copy Number Variation for Secondary Candidate Genes between Apomictic and Sexual Whole Genome Assemblies

The copy number variation (CNV) discovered in APOLLO is consistent with the hypothesis that apomixis is caused by incomplete gene copies located in hemizygous regions (Corral et al. 2013) as has been found in other species (e.g. *Pennisetum*; Conner et al. 2015). If CNV is so important for the upregulation of APOLLO in apomictic ovules, as was suggested by Corral et al. (2013), could secondary candidates be experiencing the same thing despite insignificant statistical evidence?

In Table 4.17, six probes showed differences in copy numbers between the apomictic and sexual *Boechera* WGS libraries. Further gene sequencing will have to be done to confirm this observation. Given the Arabidopsis orthologue identifiers and positions (Table 4.10) and a close microsynteny between Arabidopsis and Boechera genomes (Mandáková and Lysak, 2008; Mandáková et al. 2015), had all loci demonstrating CNV in the apomicts been physically linked (e.g. on a single chromosome arm) then duplication of that chromosome fragment, leading to two copies of all genes on that fragment, would have been a simple explanation for the observation here (Table 4.17). Since no physical linkage was found, multiple duplications are hypothesized to have led to the patterns of CNV seen in this dataset.

The accumulation of mutations, such as gene duplication in apomicts, is hypothesized to result from the absence or circumvention of meiosis (i.e. Muller's ratchet; Muller, 1964). Meiosis creates genetic variation whereby different mutations are segregated unequally between progeny. If one offspring accumulates too many deleterious mutations affecting its fitness, it is eliminated and the mutations possessed by that individual can no longer be reintroduced to the gene pool for further breeding. In asexual organisms, mutations are expected to accumulate until they lead to eventual extinction.

6 Conclusion

The results of the bioinformatic analyses identified several genes of interest related to the reproductive divergence observed in the genus *Boechera*. APOLLO's identity as a primary candidate gene for apomixis was confirmed, but future work will have to be done to further annotate and characterize the function of the identified targets. There are several genes that remain unidentified, but for those identified with the use of TAIR and NCBI, they have added to the pool of information already compiled by the Sharbel lab in the search to find the molecular interactions causing apomixis.

Transition

Chapter one of this thesis is an in depth review of the genomic implications and evolutionary changes that have occurred for asexual reproduction to take place and stabilize in plants. Apomixis is a complex trait involving multiple independent developmental steps, and the implications of introducing it to crop breeding merit research into this potential revolutionary tool. Nonetheless, the mechanism continues to evade the understanding of the world's scientists, so it is essential to have a thorough understanding on the basics of sexual reproduction in plants and the role it has in today's plant breeding as a frame of reference to understand asexuality.

Steady improvements in biotechnology and omics technologies have led to the identification of many factors associated with apomixis. Sex is such a highly regulated process that every gene involved seems to have the capability to cause downstream effects leaning in favor or apomixis. Genes such as DMT102, DYAD/SWITCH, CLOTHO, CKH2, and FIE to name a few of many genes in Table 4.2 provide evidence for the previous statement.

With context on the evolution of apomixis in *Boechera* and the genes previously discovered to show apomixis-like phenotypes, the candidate APOLLO gene is the subject of the next chapter, with a focus on the promoter regions of sexual and apomictic alleles. As the sexual and apomictic APOLLO alleles exibit different tissue-specific expression patterns (Corral et al., 2013), the promoters of both alleles are candidates underlying this differential expression.

7 *In Silico* Promoter Analysis

7.1 Introduction

Cells are complex regulated environments whose genes experience alterations to their expression, structure, and end use during transcription and translation when considering cell specificity. Promoters are upstream sequences of a gene where RNA polymerase complexes assemble to begin transcription, and are essentially the ON:OFF (up:down) switch for gene expression. If RNA polymerase cannot assemble to the promoter, transcription does not occur (Ishihama. 2000). Promoters can evolve through critical changes that have influential downstream effects involving quantitative and qualitative traits (e.g. disease resistance to abiotic stresses), making them an area of interest to evolutionary biologists and plant breeders alike.

APOLLO was discovered through a differential expression anlysis consisting of eighteen accessions of *Boechera* with equal representation from sexual and apomictic reproductive accessions (Corral et al. 2013). APOLLO, a DEDD 3'-5' exonuclease, is significantly upregulated in the ovules of apomictic *Boechera,* and is characterised by both sexual and apomictic alleles which are defined by the occurrence of an apomixis-specific 20nt insertion-deletion polymorphism (Corral et al., 2013). The sequence analysis of the apomixis-specific allele of APOLLO demonstrates that the 20bp insertion that characterises its 5'-UTR is phylogenetically ancestral as it is almost identical to that the sexual alleles in *Arabidopsis* and *Brassica* (Corral et al. 2013). Furthermore, its association with apomixis is highly conserved on a large biogeographical level (Mau et al. 2015). The orthologue for APOLLO in *Arabidopsis thaliana* (AT1G74390) has not been studied in depth, but interestingly the 20nt polymorphism that is apomixis specific in *Boechera* is present in this sexual relative, leaving room for speculation regarding its function. The 20nt polymorphism in *Boechera* is associated with a substitution of several transcription factor

binding sites which may play a role in causing upregulation of APOLLO in apomictic *Boechera* (Corral et al. 2013).

8 Literature Review

A gene is a structured entity consisting of exons, introns, 5' UTR (untranslated region), and a 3' UTR. All components listed are transcribed into mRNA, but during translation via ribosomes in a eukaryotic cell only the exons are read in three letter units known as codons. Since only the exons are translated into proteins, the non-coding sequences of a gene have held little interest to the scientific community because of the linear thinking that protein function is the gold standard for functional genomics (Palazzo and Gregory, 2014). However, there are many steps before protein assembly that can influence mRNA expression, which is why non-coding DNA has taken center stage recently. Logically thinking, if non coding DNA was not required for gene regulation it would not be conserved in intergenic regions between protein coding sequences (Chorev and Carmel, 2012). The term "garbage DNA" (Zimmer, 2015) is no longer relevant because there is a lot to be discovered in the evolution and interactions mediated by non coding regions.

8.1 **Non-Coding DNA Evolution**

Mutations in DNA can occur through replication, exposure to mutagens (induced), or spontaneously. Protein sequences are measurable units that allow the assessment of DNA mutations because they likely are evolutionary conserved, are easy to screen for duplications, and conclusions can be made regarding downstream implications via codon changes, substitutions in amino acids, and changes in protein structures/ interacting partners. Non coding DNA mutation rates are difficult to measure without fully sequenced genomes, as the majority of a genome is ncDNA (non coding DNA), making it a problematic topic because conservation implies selection for an important function.

The formation of new introns is not well understood but hypothesized to be a product of exon shuffling (Kolkman and Stemmer, 2001), random insertions (O'Neill et al. 1997), duplication(s) (Hellsten et al. 2011), or transposon insertion (Giroux et al. 1994), resulting in the formation of complex proteins with new structures. Once an intron is formed as a spacer between exon units it must overcome initial disadvantages to remain fixed in a gene's structure while at the same time being prone to variability in sequence and length (Lynch, 2002). A gene possessing a newly established intron is proposed to encounter transcriptional disadvantages because said intron may not possess appropriate flanking markers for accurate excision during mRNA maturation, leaving the intron to be processed along with the exon coding sequences (Duret, 2001). On rare occasions introns can also reside in the 5' and 3' untranslated regions. For example, it is known that introns tend to reside in the 5' UTR closer to the start codon which leaves room for speculation on whether or not it plays a role in secondary structure and element recruitment before being spliced (Chung et al. 2006). Proteins with amino acids coded from an intronic sequence are assumed to be deleterious (Dvinge and Bradley, 2015). Variability in intron size and sequence is due to a process called intron sliding, where a mutation can add or delete flanking sites required for intron removal in transcription, causing the formation of new larger exons or shortening of the intron region. The retention of an intron to form a larger exon is additionaly a rare form of alternative splicing, allowing for one gene to code for multiple different proteins (Ner-Gaon et al. 2004).

The 5' UTR is a region situated in front of the start codon for the coding sequence of a gene that posseses elements affecting translational control such as hairpins, interacting protein/transcription factor sites, and other cis elements (Mignone et al. 2002). The 5' UTR does not contribute to the protein sequence it preceeds, and unlike introns, 5' UTR's remain connected

to the matured mRNA and can form structures that affect ribosomal translation because the 5'UTR is the leading sequence that enters the ribosome (Wilkie et al. 2003). This makes it an attractive region for the study of post transcriptional regulation. There are two fates for mRNA once it has been transcribed, it either makes it to the ribosome for translation or fails due to post transcriptional regulation. The 5' UTR is the genetic chaperone of coding sequences that determine their fate. On average, the 5' UTR is approximately 100-200bp in length which is significantly shorter than the 3' UTR which occupies the tail end of a matured RNA (Pesole et al. 2001).

The 3' UTR is the tail end of a matured mRNA possessing functions in subcellular localization, stability, polyadenylation, and together with the 5' UTR complements translational control (Mazumber et al. 2003). Cells can change their gene expression profiles in response to bioitic and abiotic stresses, and during these responses different genes are transcribed with appropriate functional response while previously-transcribed mRNAs are degraded (i.e. post-transcriptional regulation). Polyadenylation is the term used to describe the addition of a poly(A) tail to the the 3' UTR of mRNA (Calvo and Manley, 2003). The function of a poly tail is to protect the mRNA during export form the nucleus into the cytoplasm, and the deadenylation (shortening) of the tail commences mRNA degradation (Funakoshi et al. 2007). Once deadenylation occurs, fully removing the poly(A) tail of a mature mRNA, the transcript is exposed to exonucleases in the exosome thus leading to termination (Elkon et al. 2013). Lastly, in addition to deadenylation, endonucleases are capable of mRNA elimination in the same or similar manner as known gene editing systems like ZFN's, TALENS, meganucleases, and CRISPR/Cas9.

8.2 **Promoter Evolution**

Promoters are regions of non-coding DNA abundant with cis regulatory elements that affect gene expression, and are located in front of the 5' UTR start site. Regulating gene expression

at the transcriptional level is the most energy efficient method for organisms compared to post transcriptional regulation eliminating the transcribed product, as it takes energy to recruit the cellular machinery required for processing interactions that initiate transcription. That initial investment of energy is like starting a car and transcription is hitting the highway. When the transcribed product is completed it will either go through further steps in protein synthesis which involve more energy investment, or it can be negatively regulated via splicing/degradation. Metaphorically speaking, if a farmer receives enough precipitation in the growing season to produce a 60 bushel/ acre canola crop, and she has purchased fertilizer for an 80 bushel/acre canola, the farmer is in an investment defecit. In other words, why invest in transcribing a protein if it is not going to be used?

Transcriptional regulation is in part controlled by the promoter, which is a hotbed for a range of cis regulatory elements such as silencers, enhancers, TATA boxes, and recognition sites for thousands of protein complexes. The general understanding of promoters is that they are directional in which way they promote transcription, but in fact that is only characteristic of evolutionarily conserved promoters. Novel promoters demostrate that the ground state of newly-formed promoters stimulate transcription bidirectionally because over time evolutionary selection has constrained conserved motifs to make them more directional (Jin et al. 2017). The size of motifs in promoter regions ranges from 6-20nt, and with promoters ranging in size from 1-10kb it is difficult to identify important binding sites due to the volume of motifs and motif turnover through mutations. It has even been found that random DNA sequences can take on promoter identity as demonstrated in Yona et al. (2018) by using the expression of reporter genes as a measure of promoter activity.

8.3 **The Meselson Effect and Gene Duplication Influence on Promoter evolution**

The "Meselson effect" is the term used to describe sequence divergence between two alleles at the same locus through a lack of recombination in asexual organisms, and was first documented by Birky (1996) in asexual bdelloid rotifers. Without recombination, genes accumulate mutations (Welch et al. 2004), and with no homologous repair template to maintain functionality genes begin to drift away from each other in both sequence and structure. Another method for gene evolution is through duplication which can occur through several different methods, and if the duplicated gene is retained within the genome there are three common fates which are (i) pseudogenization (ii) subfunctionalization and (iii) neofunctionalization (Prince and Pickett, 2002). Gene dosage is commonly observed in genes that still possess the same role/ function and have to be appropriately regulated to maintain stability. Subfunctionalization occurs when a gene is duplicated and over time the function(s) segregate and become exclusive to one of the gene duplicates. Neofunctionaliztion results in the duplicate gene acquiring novel functions via mutation accumulation or through changes in non-coding DNA regions such as the promoter. Due to APOLLO having copy number variation in asexual *Boechera* (Corral et al. 2013), it is important to understand whether CNV is connected to apomixis.

Research into the Meselson Effect has only focused on the DNA that makes up the core organization of a gene (exons and introns)(Taylor and Raes, 2004; Weiblen et al. 2015; Dowell et al. 2018) because they are structured measurable units for comparing homology. In an asexual organism it is assumed that on an evolutionary time scale the longer a gene has gone without homologous recombination the more mutations its alleles should have accumulated (Welch and Meselson, 2000). However, given the current status of our understanding of promoter evolution, is the Meselson effect expected to be observed between promoters of homologous genes? This is

a big unknown and carries ambiguity in that when a gene is duplicated, what is the status of the promoter; was it also duplicated or was it left behind? These questions can also be directed towards gene conversion events where one homologous gene completely replaces the other that had accumulated mutations resulting in both loci having 100% homology (Chen et al. 2007).

9 Materials and Methods

9.1 Accession Selection

Boechera accessions selected in Table 9.1 were chosen based on overlapping pre-existing data collected from previous experiments conducted in the Sharbel Lab. These species cover the spectrum of species diversity in the *Boechereae*, including species in the *Borodinia* and *Boechera* groups (Alexander *et al.* 2013).

Table 9.1. *Boechera* germplasm used in study. IPK_ID = Accession ID used at IPK Gatersleben with the first two letters being initials of the collector. MOR = Mode of Reproduction.

#	IPK_ID	ID in Corral et al 2013	Taxa	Ploidy	Embryo:Endosperm ratio	APOLLO apoallele	MOR	Sequenced 1kb Clones
1	ES773		*B. crandallii*	2	2:3	0	Sexual	6
2	ES758		*B. glareosa*	2	2:3	1	Sexual	7
3	JL107		*B. stricta*	2	2:5	1	Apomictic	2
4	JL99		*B. retrofracta*	2	2:5	1	Apomictic	6
5	JL100		*B. retrofracta*	2	2:5	1	Apomictic	3
6	JL106		*B. spatifolia X stricta*	2	2:5	1	Apomictic	5
7	JL73		*B. stricta*	2	2:6	0	Apomictic	16*
8	ES776	111A2	*B. holboellii*	2	2:6	1	Apomictic	5
9	ES805		*B. holboellii*	2	2:6	1	Apomictic	34*
10	TS290		*B. retrofracta*	2	2:6	1	Apomictic	19*
11	TS300		*B. retrofracta*	2	2:6	1	Apomictic	8
12	TS78		*B. retrofracta*	2	2:6	1	Apomictic	7
13	ES704	104A3	*B. divaricarpa*	3	3:9	1	Apomictic	4
14	ES596	215A3	*B. divaricarpa*	3	3:9	1	Apomictic	5
15	ES517	168A2	*B. divaricarpa*	2	2:6	1	Apomictic	8
16	TS29	329S2	*B. polyantha*	2	2:3	0	Sexual	7
17	ES552	369S2	*B. stricta*	2	2:3	0	Sexual	9
18	ES558	380S2	*B. stricta*	2	2:3	0	Sexual	6
19	ES910	385S2	*B. polyantha*	2	2:3	0	Sexual	7
20	ES913	390S2	*B. polyantha*	2	2:3	0	Sexual	5

*Apo allele and sex allele promoter clones included, others only have sex alleles

9.2 **DNA Extractions and Quality Testing**

Young leaves from each accession in Table 9.1 were sampled and frozen overnight at -80 to increase DNA extraction quantity and quality via improved tissue homogenization and cell lysis. The DNA extraction kit used was the NucleoSpin® Plant II kit from Macherey-Nagel. Once DNA extractions were complete, DNA quality was assessed by the use of Internal Transcribed Spacer (ITS) PCR amplification. ITS regions are widely used in taxonomy and barcoding/phylogeny studies because they have highly conserved flanking sequences that allow universal primers to be used in a taxon-inclusive manner (Cheng et al 2016). The primers used in Table 9.2 amplified a product approximately 650bp in length (Figure 9.1).

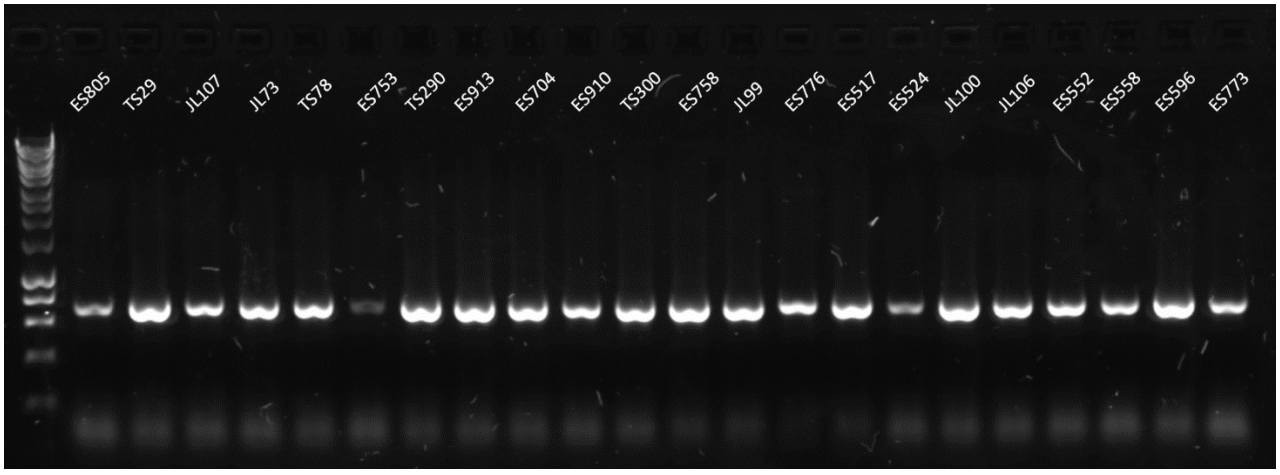

Figure 9.1 Using the ITS PCR protocol in Table 9.3, spacer DNA was amplified from 1µl of DNA solution from all accessions in Table 21 using 2x Taq FroggaMix polymerase and primers in Table 9.2. 10µl of PCR product was run on 0.8% agarose gel for 40 minutes at 100kv with a Bioline HyperLadder™ 1kb.

Table 9.2 Primers from two different publications had been previously experimentally verified for use in amplification of the *Boechera* ITS region (Mummenhoff et al. 1997 provided primer 18F and White et al. 1990 for primer 25R).

Primer	Sequence 5'-3'
ITS1 18F Hei	GGAAGGAGAAGTCGTAACAAGG
ITS2 25R Hei	GCATCGATGAAGAACGCAGC

Table 9.3 PCR protocol for amplification of ITS DNA using primers in Table 9.2

Step	Temperature (C°)	Time
Polymerase activation	94	5 minutes
Denaturation	94	30 seconds
Annealing	55	45 seconds
Extension	72	2 minutes
Final extension	72	5 minutes

9.3 **Polymerase Chain Reaction (PCR)**

Primers were designed using sequencing data from bacterial artificial chromosomes (BACS) created during the initial gene discovery experiment of APOLLO (Corral et al. 2013). BAC sequences are also available on NCBI (https://www.ncbi.nlm.nih.gov/). Primers were chosen to amplify an approximate 1Kb region upstream of the APOLLO start codon.

Table 9.4 BAC sequences containing APOLLO were initially isolated in Corral et al. 2013 and characterized to contain either the sex or the apo allele. Sequences can be retrieved using the NCBI accession number.

NCBI Accession	BAC ID (Corral et al. 2013)	Allele
KF705605.1	BAC 6	Apo
KF705604.1	BAC 5_7	Sex
KF705603.1	BAC 2b_4	Apo

Table 9.5 Primers designed from the BAC sequences in Table 9.4 to amplify the promoter of the apo and sex allele of APOLLO. TSP1R is from Corral et al. 2013 and is a universal reverse primer located within the first exon of APOLLO that is used to screen for the 5' UTR indel that is the distinguishing feature between apo and sex alleles.

Primer	Sequence 5'-3'
pAPOLLO Universal F-1kb[F]	CACCAAACAGAAACGACCATACATCCAAA
TSP1R[R]	GATAGCCCCAAACTCCAAAATCGC
APO-CDS-MAR2[R]	AGAGCTCCACTAGCTTCATAG
APO-Pro[F]	GAATCCATAGAGATTGGCGACAT

[F] Forward Primer

[R]Reverse Primer

Table 9.6 Expected size of PCR product for each primer pair. Estimates are based on BAC sequences containing APOLLO in Table 9.4

Forward Primer	Reverse Primer	Expected Size (bp)
pAPOLLO Universal F-1kb	TSP1R	1221[S]
APO-Pro	APO-CDS-MAR2	1296[A]

[S] Sex allele
[A] Apo allele

Table 9.7 PCR touchdown protocol for TSP1R and pAPOLLO Universal F-1kb primers using KOD HotStart Polymerase (TOYOBO®) which has proofreading capability qualifying it for use with amplicons that need to be sequenced.

Step	Temperature (C°)	Time
Polymerase activation	95°	2 minutes
Denaturation	95°	20 seconds
Annealing	75°x2, 73°x2, 71°x2	10 seconds
Extension	70°	2 minutes
Denaturation	95°	20 seconds
Annealing	68°	10 seconds
Extension	70°	2 minutes
Final extension	70°	5 minutes

Table 9.8 PCR protocol used for amplifying the apo allele promoter using Phusion Hot Start II High Fidelity PCR Master Mix.

Step	Temperature (C°)	Time
Polymerase activation	98	2 minutes
Denaturation	98	10 seconds
Annealing	60	15 seconds
Extension	72	30 seconds
Final extension	72	1 minute

9.4 **Cloning and Sequencing**

All PCR products were isolated using the GeneJET gel extraction kit (ThermoFisher) and blunt end ligated using a CloneJET PCR cloning kit (ThermoFisher). Ligation reactions were desalted and transformed via electroporation into DH-5α electro competent cells at 1.7kv. Selection for transformed competent cells took place on Ampicillin LB agar media at 100µg/mL grown overnight at 37°celsius. Positive colonies were sub cultured into liquid LB media containing 100µg/mL ampicillin and incubated for 12 hours at 37°celsius before colony PCR using CloneJET specific primers. pJET1.2/blunt is the linearized plasmid provided with the CloneJET PCR cloning kit that has forward and reverse sequencing primer sites flanking the blunt end cloning sites. The primers (pJet1.2 forward and pJet1.2 reverse (Table 9.10)) were used in both colony PCR and Sanger Sequencing reactions. Positive colonies verified via colony PCR were submitted via liquid culture 96 well plate format for Sanger Sequencing using TempliPhi™.

Table 9.9 Colony PCR protocol used for screening DH5α transformed liquid cultures. 2x Taq Master Mix from Frogga Bio™ was used with pJET 1.2 primers in Table 9.10 to amplify all cloned promoter inserts (apo and sex).

Step	Temperature (C°)	Time
Polymerase activation	94	5 minutes
Denaturation	94	30 seconds
Annealing	55	30 seconds
Extension	72	70 seconds
Final extension	72	4 minutes

Table 9.10 pJET primers used for sequencing and colony PCR. Primers come with the pJET 1.2 blunt end-cloning vector provided by ThermoFisher Scientific™.

Primer Name	Sequence (5'-3')
pJET 1.2 Reverse	AAGAACATCGATTTTCCATGGCAG
pJET 1.2 Forward	CGACTCACTATAGGGAGAGCGGC

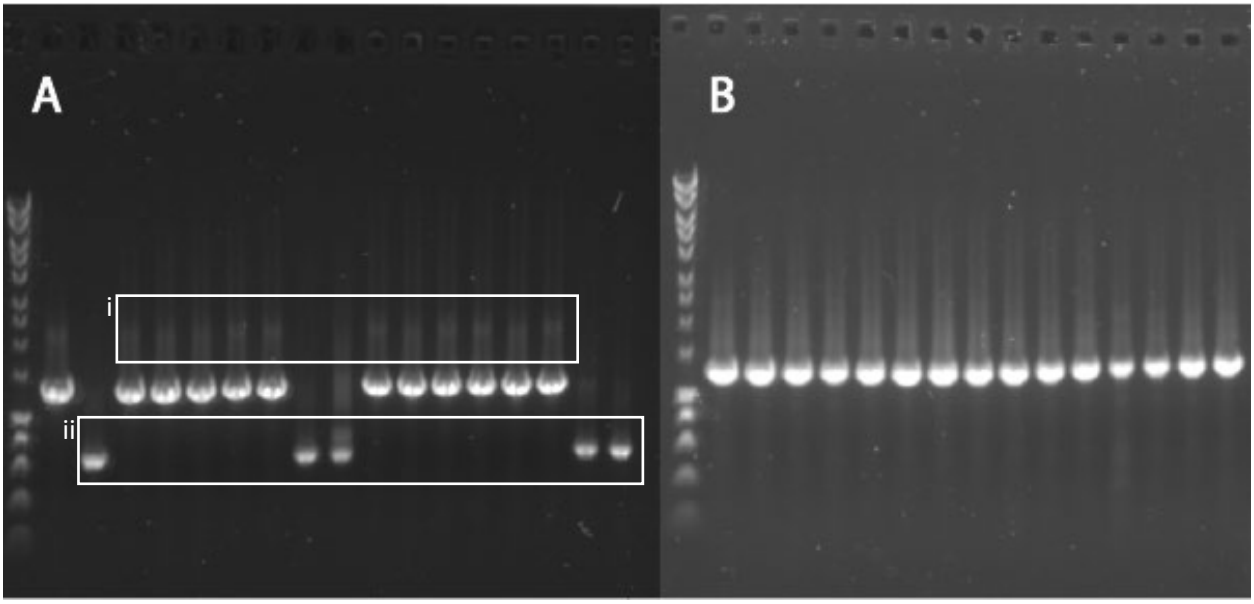

Figure 9.2 Colony PCR results for both the sex allele (A) and apo allele (B) on 0.8% agarose gel with a Bioline HyperLadder™ 1 kb. Faint bands above the 1kb positive results (i) is the plasmid backbone (approx. 2kb) and the smaller products (ii) are off target amplification that were sequenced and found to be the result of the TSP1R acting as both the forward and reverse primer for this small piece of DNA. The sex allele colonies in (A) are from accession JL100 and apo allele colonies (B) are from accession TS290**.**

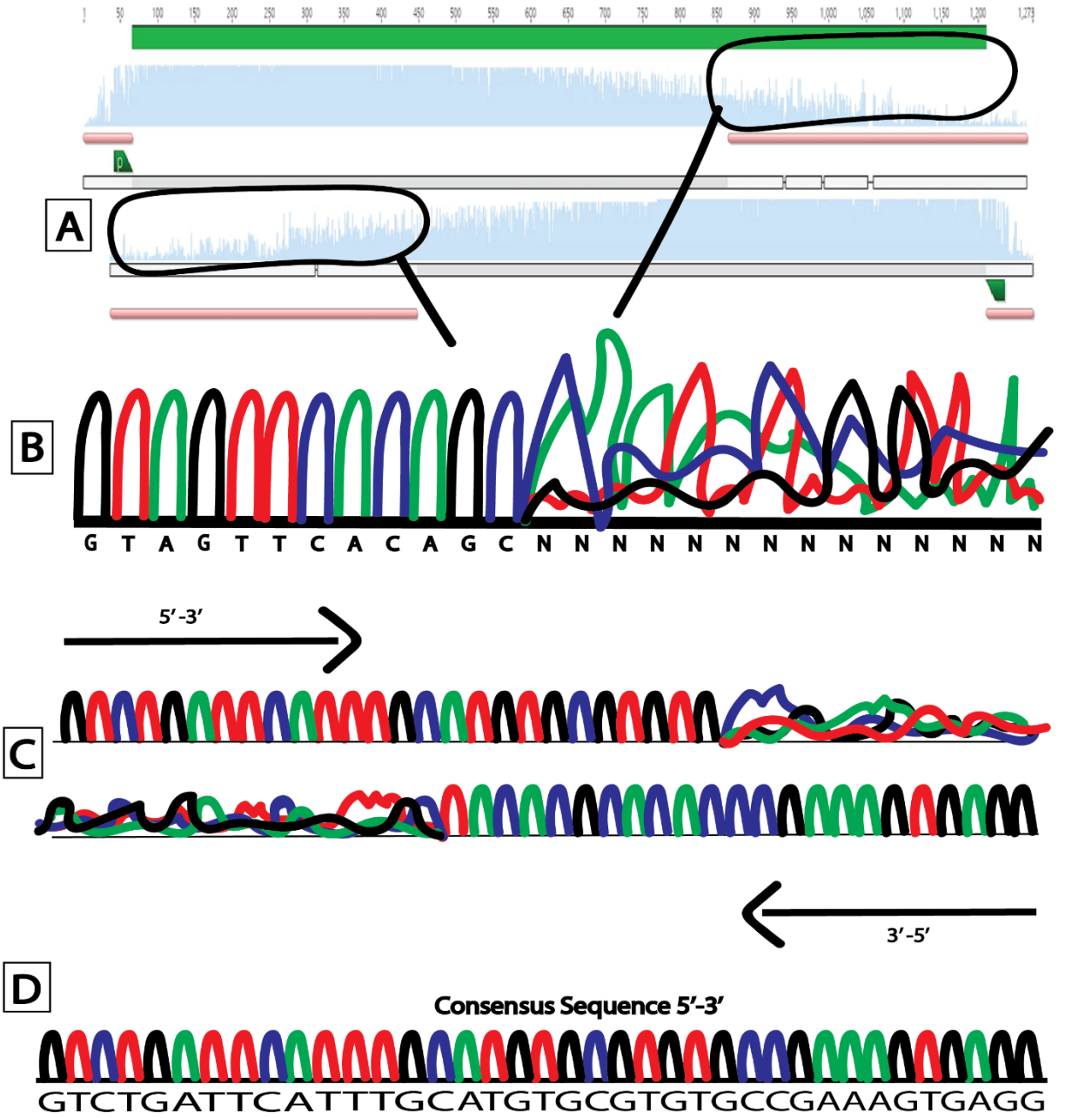

Figure 9.3 The Sanger sequencing method used in this study dye terminator sequencing which uses fluorescent dyes to label each of the four nucleic acids. Sanger sequencing is a PCR reaction requiring a 5'-3' directional primer to anchor as the starting point on the DNA template. Sanger sequencing consistently provides sequencing reads upwards of >600bp in one direction until quality drops off and the chromatograph becomes unusable (A and B). Because the Apollo promoter is approximately 1.1kb in size, sequencing had to be done in both directions to create an overlap region in the middle that would act as a bridge (C). The alignment of both sequencing reads exemplified in A and C completely covers the 1.1kb region with high quality chromatograph peaks (D).

9.5 **Promoter Analysis**

All sequencing data was assembled using Geneious R10 (http://www.geneious.com/). Assessing nucleotide diversity and haplotypes per accession was done with DNAsp (V6) (Rozas et al. 2017) while functional analysis was done using the Genomatix Software Suite (http://www.genomatix.de). Transcription factor identification was completed with MatInspector (Cartharius et al. 2005) using a core similarity of 1.00 and an optimized matrix similarity. The resultant search results were then again filtered by increasing the matrix similarity to 95% to identify a collection of transcription factors with a high potential for binding to the regulatory region of APOLLO. Another method to identify interacting factors using MatInspector is to look for overrepresented transcription factor families within the promoter region. This can statistically compare the occurrence of a motif in a genome to that of the promoter. By comparing how often motifs occurred in *Arabidopsis thaliana* promoters (TAIR10 Genome), statistics were calculated using Z-scores from the deviation in number of motifs contained in the APOLLO promoter.

10 Results

During the study it became evident, after sex alleles of all accessions (Table 9.1) were cloned and sequenced, that the forward primer *pAPOLLO Universal F-1kb* was not universal. With >100 clones and high quality sequencing, the distinctive 5' UTR indel that characterized the APOLLO apo allele could not be found. The BAC's that contained APOLLO from Corral et al. (2013; Table 9.4) were compared, and it became evident that beyond the 5' UTR, APOLLO apo alleles have a completely different promoter region compared to sex alleles (Fig. 10.1). Primers were thus redesigned to amplify the apo allele promoter and tested on all accessions using the phusion protocol in Table 9.8. Out of all accessions in Table 9.1, only three accessions (CJ73, ES805, and TS290) were successfully amplified, cloned, and sequenced.

From early comparisons of the sex allele promoter to its *A. thaliana* orthologue (AT1G74390) it was evident that significant changes in sequence homology have occurred because the promoter of AT1G74390 is small and bi-directional in nature (Figure 10.2). It is unknown if the promoters for the apo and sex alleles of APOLLO are bidirectional until sequencing ≥10kb upstream can identify neighbouring genes. With 1kb of upstream sequencing completed, some unique observations arise.

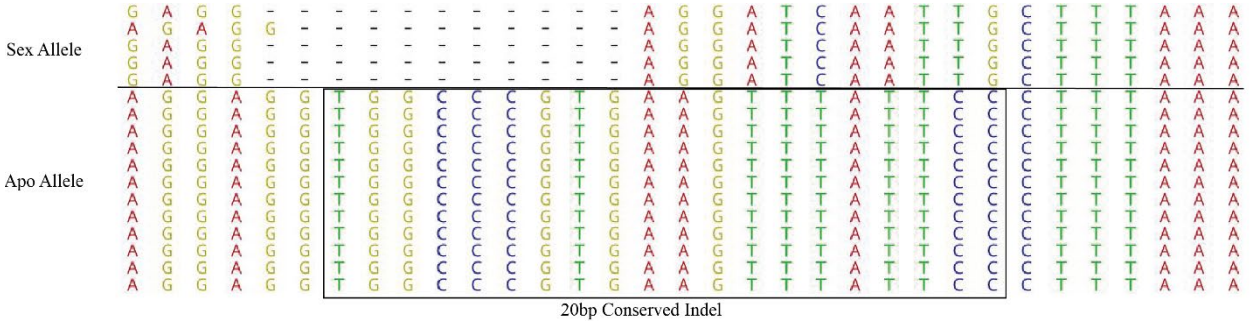

Figure 10.1 APOLLO 5' UTR alignment 44bp upstream of the start codon for exon 1 of all sequenced clones for accession TS290 (*B. retrofracta*). Within the black rectangle is the conserved 20bp indel that characterizes apo alleles from sex alleles as highlighted in Corral et al. 2013

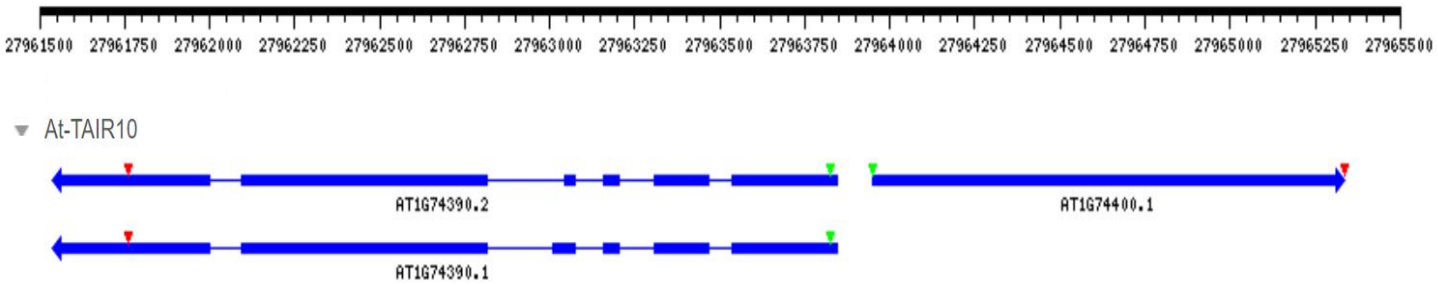

Figure 10.2 Using the TAIR genome browser to capture the size of the promoter for AT1G74390 (APOLLO's orthologue) it became evident that the 200bp region (between green arrows) must promote transcription bi-directionally because transcription of AT1G7440 occurs in the opposite direction.

10.1 **Assessment of genetic diversity between the apo and sex allele promoters**

Since the apo and sex allele promoters showed little similarity till near the start codon of APOLLO (Figure 10.1) they were analyzed separately for allelic diversity. Through initial comparisons of the 120 sequenced sex allele alignment (Table 10.3, Figure 10.3) there was no pattern of alleleic diversity based on haplotypes. Despite patterns being evident when visualized in Figure 10.3 there was no statistical significant differences in allelic diversity between groups, regardless whether the sequences were analyzed as a single group based on sex or apomixis, or by accession. Both MOR (modes of reproduction) showed extremes such as sexual accession ES910 having the most segregating sites, and apomictic accession JL107, JL99, ES776, and JL 100 having less than three segregating sites that separated their haplotypes (Table 10.1). The number of segregating sites also had no correlation to the number of haplotypes because accessions can have $\geq$ 20 segregating sites and still have the same number of haplotypes as an accession with 3 $\leq$ segregating sites. Thus there is evidence for duplication in Table 10.1 based on number of haplotypes per accession, and there are alleles that appear to be more divergent than others based on their segregating sites.

Despite the small sample size, haplotype diversity was low in the apo allele, as only eight SNPs were found in the 1.2kb region that additionally contained the 20bp indel in the 5' UTR (Table 10.2, Figure 10.4). In addition to the 8 SNPs, polynucleotide repeats (Figure 10.4) were difficult to sequence and not included in statistical analyses.

Table 10.1 Analysis of all sequences per accession measuring genetic diversity of the APOLLO sex allele promoter. HD = a measure of haplotype diversity; k = the average number on nucleotide differences; pi = nucleotide diversity per site; pi(JC) = pi with Jukes-Cantor correction.

Accession	Seg. Sites	Haplotypes	Hd	k	pi	pi(JC)	MOR
JL107	2	2	1.000	2.000	0.00175	0.00175	Apomict
JL99	3	3	0.600	1.000	0.00088	0.00088	Apomict
JL100	1	2	0.667	0.667	0.00058	0.00058	Apomict
JL106	20	2	0.400	8.000	0.00701	0.00709	Apomict
JL73	26	4	0.900	10.80	0.00947	0.00960	Apomict
ES776	3	3	0.800	1.600	0.00140	0.00140	Apomict
ES805	27	3	0.639	13.22	0.01159	0.01177	Apomict
TS290	11	3	0.700	5.800	0.00508	0.00511	Apomict
TS300	11	5	0.857	5.393	0.00473	0.00475	Apomict
TS78	9	6	0.952	4.762	0.00417	0.00419	Apomict
ES704	24	3	0.833	12.167	0.01066	0.01079	Apomict
ES596	25	2	0.400	10.00	0.00876	0.00889	Apomict
ES517	24	7	0.964	9.179	0.00804	0.00810	Apomict
ES773	0	1	0	0	0	0	Sexual
ES758	21	2	0.476	10.00	0.00876	0.00887	Sexual
TS29	0	1	0	0	0	0	Sexual
ES552	7	3	0.417	1.556	0.00136	0.00137	Sexual
ES558	1	2	0.333	0.333	0.00029	0.00029	Sexual
ES910	30	5	0.905	16.10	0.01411	0.01430	Sexual
ES913	2	3	0.700	0.800	0.00070	0.00070	Sexual

*see Table 9.1 for accession ploidy

Table 10.2 . SNP Table of apo allele promoters. Location is based on its position upstream of the start codon of APOLLO (right to left). The total number of sequences aligned from the three amplified accessions (CJ73, ES805, and TS290) is 50, which is also the denominator in calculating the variant frequency.

Location	Variant Frequency
114	12%
114	88%
136	70%
136	30%
370	90%
370	10%
391	84%
391	16%
452	74%
452	26%
673	50%
673	50%
691	12%
691	88%
876	20%
876	80%

Table 10.3 SNP table based on the alignment of 120 sex allele promoter sequences from all accessions in Table 9.1. Location refers to the genetic position upstream of the APOLLO start codon (right to left).

Location	Length	Coverage	Polymorphism Type	Variant Frequency
4	1	120	SNP	64.70%
4	1	120	SNP	35.30%
48	1	120	SNP	64.70%
48	1	120	SNP	35.30%
60	1	120	SNP	35.30%
60	1	120	SNP	64.70%
75	1	120	Indel	13.40%
75	1	120	Indel	86.60%
84	1	120	SNP	26.90%
84	1	120	SNP	70.60%
115	1	120	SNP	30.30%
115	1	120	SNP	69.70%
244	1	120	SNP	46.20%
244	1	120	SNP	53.80%
284	1	120	SNP	25.20%
284	1	120	SNP	74.80%
346	1	120	SNP	26.10%
346	1	120	SNP	73.90%
378	1	120	SNP	25.20%
378	1	120	SNP	74.80%
390	1	120	SNP	37.00%
390	1	120	SNP	62.20%
429	1	120	SNP	63.00%
429	1	120	SNP	37.00%
473	1	120	SNP	89.90%
473	1	120	SNP	10.10%
531	1	120	SNP	36.10%
531	1	120	SNP	63.90%
536	1	120	SNP	26.90%
536	1	120	SNP	73.10%
552	1	120	SNP	63.90%
552	1	120	SNP	36.10%
600	1	120	SNP	37.00%
600	1	120	SNP	63.00%
616	1	120	SNP	64.70%
616	1	120	SNP	35.30%
627	1	120	SNP	89.10%
627	1	120	SNP	10.90%
657	1	120	SNP	35.30%
657	1	120	SNP	64.70%
728	1	120	SNP	58.00%
728	1	120	SNP	42.00%
777	1	120	SNP	42.00%
777	1	120	SNP	58.00%
780	1	120	SNP	42.90%
780	1	120	SNP	57.10%
916	1	120	SNP	45.40%
916	1	120	SNP	54.60%
988	1	120	SNP	89.10%
988	1	120	SNP	10.90%
1003	1	120	SNP	56.30%
1003	1	120	SNP	43.70%

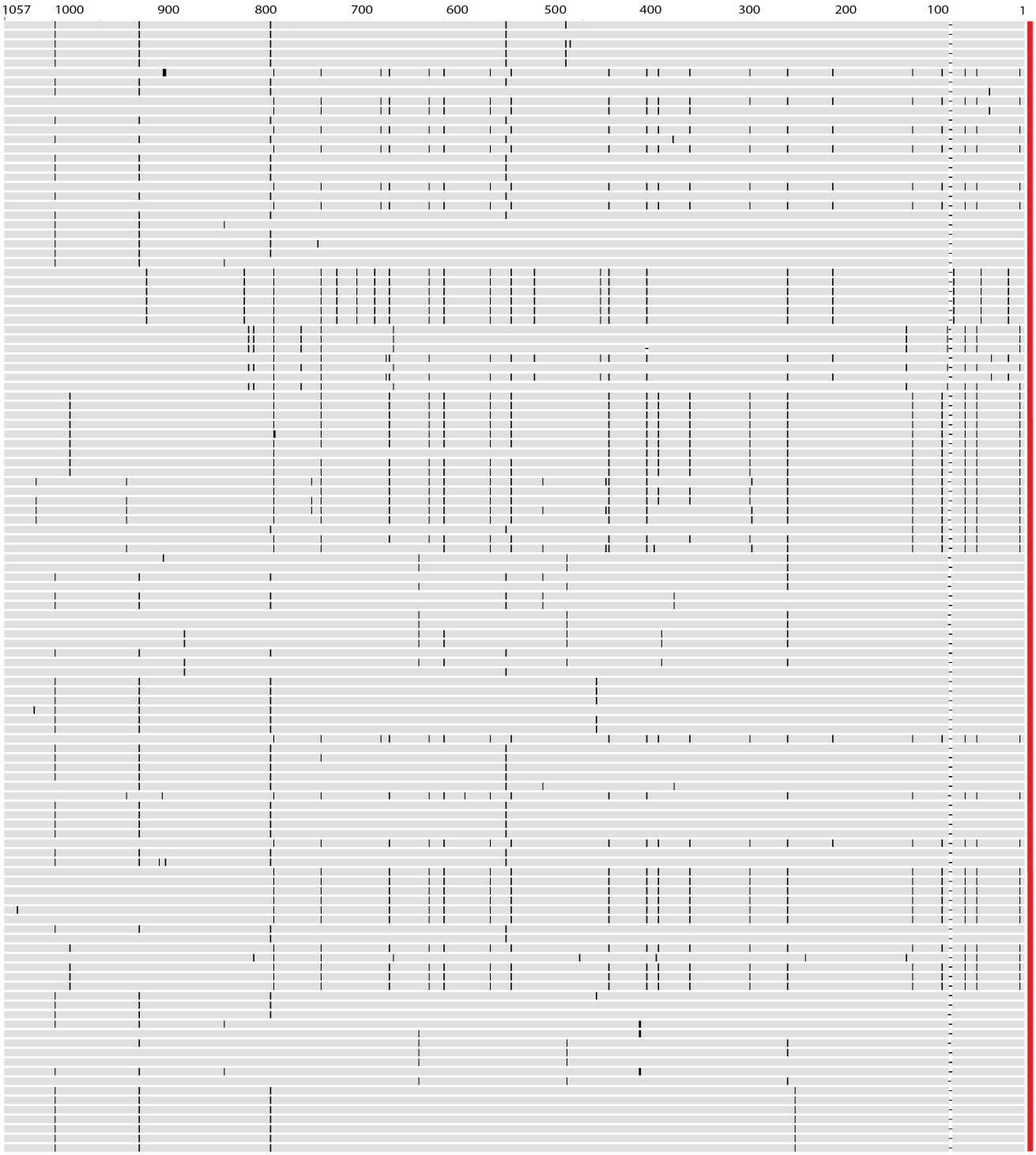

Figure 10.3 Sex promoter SNP alignment showing the placement of all SNP's in Table 10.3 across all 120 cloned sequences. Red bar on the right represents the start codon for APOLLO. At 80bp, the gap representing an indel is another poly repeat region causing sequencing errors. All black vertical lines represent SNP's.

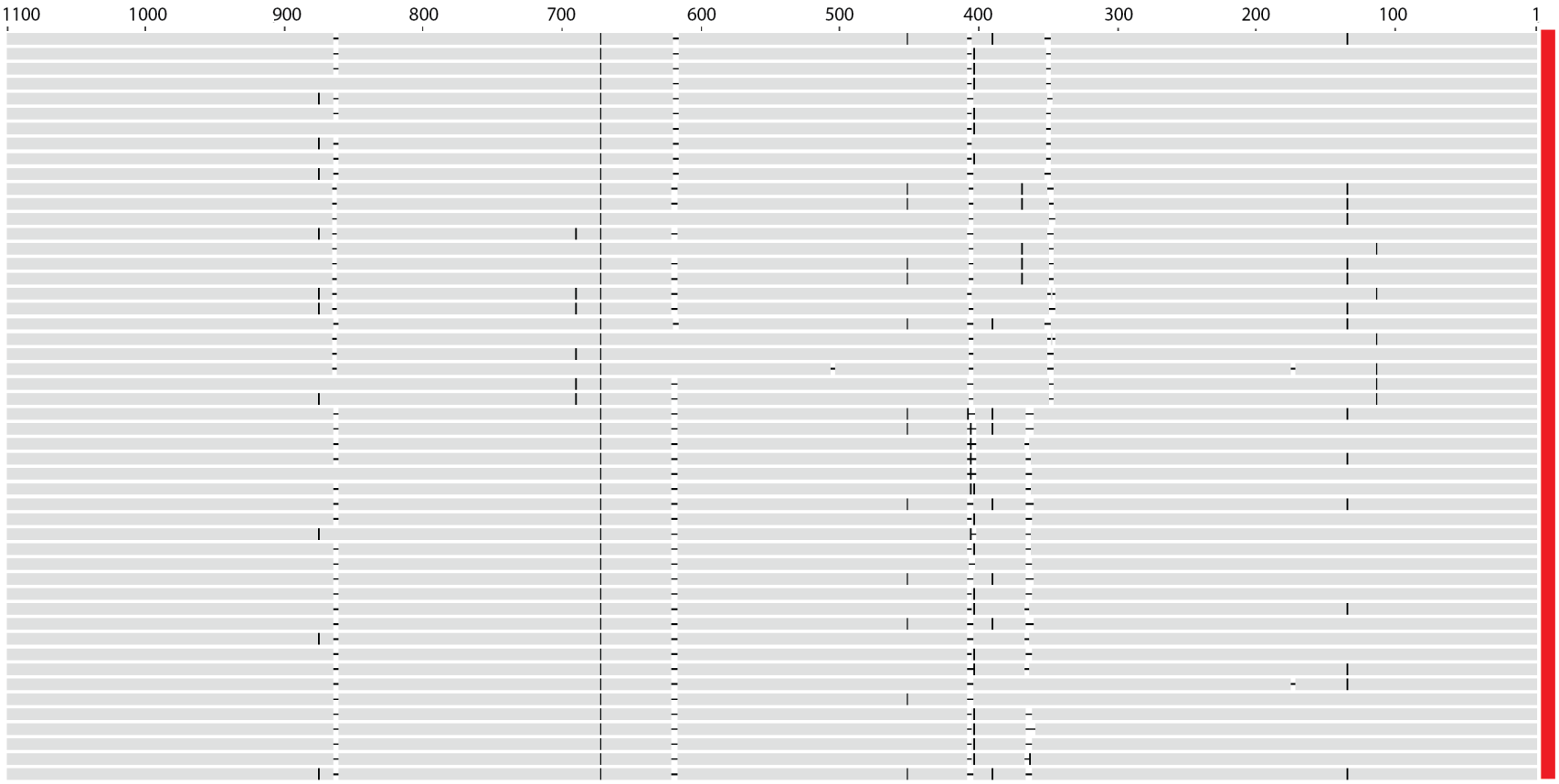

Figure 10.4 Apo promoter SNP alignment visualizing the SNP's in Table 10.2 across all the sequenced clones. The red bar on the right represents the start codon of APOLLO. White gaps are locations of indels, but as referred to in Fig. 10.3 they are the locations of long poly repeats causing sequencing difficulties. Black vertical lines are the SNP's.

10.2 *In Silico* Promoter Analysis

To adequately model the functional domains contained within the promoter sequences for identification in MatInspector, consensus sequences were created because the program can only analyze single sequences. Using the sex allele consensus sequence derived from a collective of the 47 sex allele haplotypes (Table 10.1), MatInspector identified ten motifs of interest (Table 10.4) with the top hit being WUSCHEL, an apomicitc candidate gene (Table 4.2; Zuo et al. 2002) . A similar analysis using the apo allele consensus sequences resulted in the identification of genes such as RAV1, TESMIN, and DIVARICATA 1 (Table 10.5). In addition to the apo (Table 10.6) and sex alle analysis, the promoter of AT1G74390 (APOLLO orthologue in *Arabidopsis thaliana*) was also analyzed as an external reference (Table 10.7). It concluded that the only DNA binding domain consistent between the APOLLO sex allele promoter and AT1G74390 is RKD2 (RWP-RK domain containing 2). Given the variability that comes along with *in silico* analyses of non coding DNA regions, matrix similarity filters were strictly set to $P<0.05$ so that a small list of high confidence domains would be identified instead of hundreds by using relaxed stringency. With the goal of identifying potential factors implicated in apomeiotic development, every identified transcription factor was checked for involvement in megasporogenesis and the formation of the ovule primordium.

With the apo allele promoter having high similarity between apomictic accessions (8 SNP's over 1.2kb), it is unlikely that any one SNP could introduce a new DNA binding domain. An analysis for overrepresented TF motifs using relaxed motif identification criteria over the 1.2 kb apo allele promoter of APOLLO using genome statistics (likelihood of a transcription factor occuring in the promoter of any gene) from *Arabidopsis thaliana* showed that TSO1-like proteins collectively have 18 matches (Table 10.6) and have an important role in floral organ development.

Table 10.4 Analysis of transcription factor binding site (TFBS) in the APOLLO sex allele using MatInspector. A consensus sequence was created using a 50% threshold to eliminate ambiguous nucleotides from the alignment of all 47 identified sex allele haplotypes. Matrix similarities were filtered for results only containing P<0.05 due to the small size of TFBS'. Sequence orientation is left to right with the APOLLO start codon located at 1058bp.

Transcription Factor	From	To	Anchor	Strand	Matrix Sim.	Sequence
WUSCHEL-related homeobox 13	30	40	35	+	0.952	cctCAATcatg
RWP-RK domain containing 2	458	472	465	+	0.953	tTTTCcgcttccatt
Trihelix transcription factor AT1G76880 (DF1)	604	622	613	+	0.95	cgccgcagaGTAAaaacct
Maize C1 myb-domain protein	611	631	621	-	0.953	atctctgGTAGgtttttactc
PHR1-like 2 (G2-like family)(AT3G24120)	813	829	821	+	0.957	tctccccagATTCaaga
DIVARICATA 1 (myb related R-R-type factor AT5G58900)	820	838	829	-	0.953	gaagcatTATCttgaatct
DIVARICATA 6 (myb/SANT-like DNA-binding domain-containing protein AT5G05790)	862	880	871	-	0.954	tctacaTTATcatacgaaa
Class I GATA factors	865	881	873	+	0.957	cgtatGATAatgtagaa
Nodulin consensus sequence 1	925	935	930	+	0.951	cAAAAgatgac

Table 10.5 Analysis of TFBS in the APOLLO apo allele using MatInspector. A consensus sequence from all apo alleles being aligned was created with a 50% threshold therefore eliminating ambiguous nucleotides, which are problematic for TF binding domain identification. Sequence orientation is left to right with the APOLLO start codon located at 1103bp. Matrix similarity was filtered for P<0.05

TF Description	From	To	Anchor	Strand	Matrix Sim.	Sequence
Homeobox-leucine zipper protein ATHB-6	152	170	161	+	0.959	tatatCAATgatacaatta
Cis-element involved in SA (salicylic acid) induction of secretion-related genes via NPR1	342	356	349	+	0.957	caaGAAGaaaaagat
AP2/ERF and B3 domain-containing transcription factor RAV1	450	462	456	-	0.958	atcAACAaaagct
Arabidopsis 6b-interacting protein 1-like 1	719	737	728	-	0.953	aaggagaaaGTGAatggtg
Class I GATA factors	773	789	781	-	0.958	ggaaaGATAattactca
DIVARICATA 1 (myb related R-R-type factor AT5G58900)	774	792	783	+	0.954	gagtaatTATCtttcctta
Homeobox protein 32	836	854	845	+	0.957	aaatattaTTATtctaaat
TESMIN/TSO1-like CXC 6 (At2G20110)	861	891	876	-	0.952	tttgtaacaaaTTTTaaaacaaaatttcaaa
Trihelix transcription factor GTL1	1016	1034	1025	+	0.956	ttttttccgTAAAaagag

Table 10.6 Transcription factor overrepresentation analysis of the apo allele promoter of APOLLO

TF Description	# of Matches	Expected	Std.dev. ±	Over rep.	Z-Score
Dof zinc finger protein DOF5.1 (AT5G02460)	3	0.12	0.35	24.66	6.82
CXC domain containing TSO1-like protein 1 (TCX3)	10	1.81	1.34	5.54	5.73
Dof zinc finger protein DOF1.10 (AT1G69570, HPPBF-2b)	3	0.17	0.42	17.17	5.56
Zinc finger protein OBP3	3	0.26	0.51	11.73	4.44
Cyclic DOF factor 3	5	0.75	0.86	6.69	4.34
Dof zinc finger protein DOF2.2 (AT2G28810)	4	0.51	0.72	7.79	4.17
TESMIN/TSO1-like CXC 6 (At2G20110)	8	1.85	1.36	4.33	4.16
Dof zinc finger protein DOF5.8 (AT5G66940)	2	0.12	0.35	16.34	3.94
B-box domain protein 31	3	0.32	0.57	9.36	3.85
Dof zinc finger protein DOF2.4	3	0.34	0.58	8.93	3.73
Dof zinc finger protein DOF1.8 (AT1G64620)	3	0.35	0.59	8.54	3.63
High Cambial Activity 2 (AT5G62940, DOF5.6)	2	0.15	0.38	13.51	3.51
Dof zinc finger protein DOF2.4	4	0.74	0.86	5.37	3.19
DOF Affecting Germination 2	3	0.46	0.68	6.57	3.02
KH and zinc finger CCCH domain-containing protein	5	1.24	1.11	4.03	2.93
Member of the EPF family of zinc finger transcription factors	2	0.21	0.46	9.54	2.82
Dof zinc finger protein DOF4.7 (AT4G38000)	3	0.51	0.72	5.85	2.78
Opaque-2 regulatory protein	3	0.52	0.72	5.74	2.73
Homeobox protein 40	6	1.94	1.39	3.09	2.56
Homeobox 51, Late Meristem Identity 1	7	2.54	1.59	2.76	2.49
Dof zinc finger protein DOF4.5	2	0.26	0.51	7.66	2.42
Homeobox-leucine zipper protein ATHB-20	8	3.19	1.78	2.51	2.42
Suppressor of overexpression of CO 1 (AGL20)	2	0.27	0.52	7.47	2.38
Myb family transcription factor REVEILLE 1	6	2.12	1.45	2.83	2.33
Dof zinc finger protein DOF3.4 (OBF binding protein 1)	3	0.65	0.81	4.6	2.29
C2H2-like zinc finger protein ENHYDROUS (AtIDD1) (AT5G66730)	2	0.29	0.54	6.83	2.23
WRKY DNA binding protein 18	2	0.31	0.56	6.37	2.12
Homeobox-leucine zipper protein ATHB-13	6	2.32	0.52	2.59	2.1
ABA (abscisic acid) inducible transcriptional activator	2	0.32	0.57	6.21	2.08
Arabidopsis thaliana Zinc-dependent Activator Protein-1 (ZAP1)	2	0.33	0.57	6.15	2.06

Table 10.7 Using MatInspector, the promoter of AT1G74390 was analyzed for TFBS with a matrix similarity score filter of p<0.05

Transcription Factor	From	To	Anchor	Matrix Sim.	Motif
Yabby transcription factor CRABS CLAW	90	102	96	0.953	agcaTGATtatgg
H. vulgare dehydration-response factor 1	448	468	458	0.956	catctgcACCGccccgagat
RWP-RK domain containing 2	518	532	525	0.953	tTTTCcgcttccatt
PHR1-like 2 (G2-like family)(AT3G24120)	873	889	881	0.957	gttccccagATTCaaga
Arabidopsis 6b-interacting protein 1-like 1	963	981	972	0.957	tcactccaaGTGAatgcag
Trihelix transcription factor GTL1	1030	1048	1039	0.956	atttttttccgTAAAaagag
bZIP protein G-Box binding factor 1	1042	1060	1051	0.953	acgcgcCACGtcctctttt
Rice transcription activator-1 (RITA), basic leucin zipper protein, highly expressed during seed development	1044	1060	1052	0.958	aagaggACGTggcgcgt

*Captital letters in motif sequence are the most conserved bases in the motif

10.3 **The Isolated Apo Allele Promoter is Novel**

To gain insight on the origin of the apo allele promoter, the roughly 1.1kb sequence was searched against a new gold standard genome assembly of a sexual *Boechera* (Sharbel lab - unpublished) with varying word sizes, scoring matrixes, and e-value cutoffs. No BLAST approach led to alignments longer than 40bp, with results coming from locations throughout the genome, a result consistent with the idea that the promoter is novel. Taking the investigation a step further, again, multiple BLAST searches using the NCBI database with varying filters (megablast, discontiguous megablast, and blastn) also led to no significant hits (no taxonomic BLAST filters).

11 Discussion

The significant differential expression between sex and apo APOLLO alleles (Corral et al, 2013) is correlated with divergent promoter(s) and brings to question the origin of the apo allele promoter. Novel promoters can arise through either fragment duplication that have accumulated mutations (Meselson effect) or novel assembly.

11.1 Is there more than one Novel Promoter of APOLLO?

Discovering a novel promoter involved in a trait as important as apomixis is a significant stepping-stone in understanding the factors that regulate transcription. Significant work needs to be done on the APOLLO promoter to determine if there are different types of novel apo allele promoters, as the primer set tested on twenty accessions only successfully amplified from three accessions (Table 9.5). The reverse primers in Table 9.5 are within the first exon of APOLLO and hence likely conserved anchors with proven reliability in multiple studies. What remains to be done is to use a method such as genome walking to amplify areas upstream of the reverse primer in the apomictic accessions that did not amplify with APO-Pro (forward primer) to uncover what has happened.

Unfortunately, promoters can only be modeled bioinformatically (*in-silico*) to a certain extent because motifs can range from 6-20nt in regions that span multiple kilobase pairs. This can lead to the identification of hundreds of motifs when only a select few will truly be an interacting factor. Functional bioinformatics on promoter regions is a thus a prediction game until *in vivo* studies (plant transformation) and *in vitro* approaches (ChiP–Seq) are completed. All *in-silico* data has to be compared with proper expression data capturing whether or not a TF of interest is expressed during a specific period of development. Expression data helps limit false positives; however, obtaining data for every stage of megasporogenesis, which is required for further

investigation into the results mentioned in this chapter, is a large and difficult task. Since there is no knowledge regarding what binds to new apo promoter, a prediction tool with a binding matrix statistic could be used, but it should be kept in mind that matches below a 95% confidence level may still hold significance in promoter element prediction.

There are several transcription factors involved in this process, possessing functions from cytokinesis to organ development and gametophyte architecture. Conservation in the function of master regulators in these processes is vital for plant reproduction and most often mutations in such genes are lethal.

11.2 **Promoter Motifs of Interest**

Conservation of nucleic acids between orthologous genes implies that natural selection has maintained function. Interestingly, RKD2 is the only DNA binding domain (DBD) to be significantly maintained between the sex allele of APOLLO in *Boechera* and the promoter of AT1G74390 in Arabidopsis. There are five members of the RKD family of transcription factors in *Arabidopsis thaliana* with no such analysis undertaken in *Boechera* as of yet. All five members have shown expression restricted to reproductive development, with RKD2 showing peak expression during the late stages of embryo sac formation (Tedeschi et al. 2016). Further work needs to be done on the interaction between RKD2 and DEDDH Exonucleases (e.g. APOLLO) to examine whether further downstream factors also demonstrate conserved function.

With previous knowledge of genes identified to be potential factors in apomictic development, it is surprising to see a WUSCHEL-related homeobox transcription factor identified and conserved in the sex allele of APOLLO. WUSCHEL is a gene previously identified to promote the transition of vegetative cells to take on embryonic fate, and is thus being capable of promoting somatic embryogenesis with an increase in expression (Zuo et al. 2002). However, genes in the

WOX family (WUSCHEL-related homeobox domains) possess functions beyond just embryogenesis, as demonstrated in Cheng et al. (2014) who showed that WOX genes play roles in hormone signaling and abiotic stress responses in rice.

DIVARICATA 1 and 6 are transcription factors containing a MYB DNA binding domain that is highly conserved in plants with over 200 family members (Zhang et al. 2018). Being the largest family of plant transcription factors, DIVARICATA 1 and 6 do not have reported function outside of being involved in dorsoventral asymmetry of snapdragon flowers (Galego and Almeida, 2002). DF1 is a repressor of trichome growth and is a duplicated version of GTL1. Loss of function mutants resulted in increased trichome size and modified expression of cell cycle genes. Given the polyploid nature of trichomes and epidermal plant tissue that have higher ploidy than the growing tips of meristematic tissue, it is interesting to hypothesize that a DF1 mutant could have an effect on ploidy during megasporogenesis.

TESMIN is a gene found to be transcribed during developmental stages of sexual gametes in both animal and plant models (Andersen et al. 2007; Hauser et al. 2000; Olesen et al. 2004). A characteristic common between TESMIN homologs is the presence of two CXC domains that remain crucial for Tesmin-like protein (TXC) function. There is an entire family of proteins containing CXC domains, and two of them have been identified as overrepresented in the apo allele promoter of APOLLO. TCX3 (*Arabidopsis thaliana* locus AT3G22760), otherwise known as SOL1, is a close homolog of TSO1 and is a type 1 TCX protein inferred from phylogenetic analysis of all TCX proteins in Andersen et al. (2007). In regards to expression in sexual gametes, it was found to be highly expressed in pollen but no further conclusion on function/ staging could be made. Little is known about the other Tesmin-like protein TCX6 other than it is a homolog of

TSO1, which is crucial for both male and female development in *Arabidopsis thaliana* (Andersen et al. 2007).

12 Conclusion

This analysis is the base of the pyramid establishing the groundwork for future functional genomic experiments involving APOLLO and its promoters. *In silico* analyses are difficult as databases are heavily biased towards *Arabidopsis thaliana,* and thus the data presented here require further validation in *Boechera*. The apo allele of APOLLO is exclusively correlated with a highly conserved promoter of novel origin which is an observation not previosuly documented in any apomictic candidate gene. Phylogenetic analyses based on SNPs and indels are a beginning step in identifying allele copy number per accession, but to validate these data there needs to be complementary fluorescent *in situ* hybridization (FISH) and genome sequence analyses.

Sliding window analysis is routinely done on protein coding sequences, and can illuminate the location significance of a SNP. However, relatively less can be said for the significance of a SNP in a promoter region as it is a non-coding region. Available online programs are capable of identifying hundreds of motifs, yet the true method of identifying promoter function *in vivo* depends on transformation work which has been the achilles heel of working with *Boechera*. Future work must be focused on transformation and genome sequencing to find more about the genomic context/landscape of APOLLO. These data can be added to future datasets where the APOLLO gene, its promoter (1-5kb in length), and the 3' region are all together in one line of DNA for the same comparative analysis executed in this chapter.

13 Thesis Conclusion

So what does the entirety of this thesis do for the future direction of research initiatives directed to solving the basis of apomixis in *Boechera*? The data presented in this manuscript presents a full review of genes associated with apomixis-like phenotypes, as referenced for the public literature. Considering the increasing ease with which gene function can be tested (e.g. CRISPR technology), this database provides the basis of a next generation of apomixis research to test combinations of genes in forwards genetics approaches.

Functional genetics in a niche organism like *Boechera* is difficult. Gathering funding and the appropriate resources to accomplish what has been done in *Boechera* thus far is impressive. For perspective, the genetic diversity contained within the genus means that for genome analysis to be applied there needs to be genomes available for both sexual and apomictic accessions of different species, and this is not including accessions with supernumerary chromosomes. *Boechera* is special considering variation for sex and apomixis when studying molecular genetics, and this complicates our understanding about the evolutionary development of the forms of reproduction. My analyses of the promoter region of a candidate gene in a non-model organism has provided a building block in the analysis of the complete gene expression interactome of the APOLLO gene in sexual and apomictic *Boechera*.

References

Ach, R. A., Taranto, P., & Gruissem, W. (1997). A conserved family of WD-40 proteins binds to the retinoblastoma protein in both plants and animals. The Plant Cell, 9(9), 1595-1606.

Acosta-García, G., & Vielle-Calzada, J. P. (2004). A classical arabinogalactan protein is essential for the initiation of female gametogenesis in Arabidopsis. The Plant Cell, 16(10), 2614-2628.

Agashe, B., Prasad, C. K., & Siddiqi, I. (2002). Identification and analysis of DYAD: a gene required for meiotic chromosome organisation and female meiotic progression in Arabidopsis. Development, 129(16), 3935-3943.

Agrawal, A. F. (2006). Similarity selection and the evolution of sex: revisiting the Red Queen. PLoS biology, 4(8), e265.

Albertini, E., Marconi, G., Reale, L., Barcaccia, G., Porceddu, A., Ferranti, F., & Falcinelli, M. (2005). SERK and APOSTART. Candidate genes for apomixis in Poa pratensis. Plant Physiology, 138(4), 2185-2199.

Aliyu, O. M., Schranz, M. E., & Sharbel, T. F. (2010). Quantitative variation for apomictic reproduction in the genus Boechera (Brassicaceae). American Journal of Botany, 97(10), 1719-1731.

Alleman, M., & Doctor, J. (2000). Genomic imprinting in plants: observations and evolutionary implications. Plant molecular biology, 43(2-3), 147-161.

Andersen, S. U., Algreen-Petersen, R. G., Hoedl, M., Jurkiewicz, A., Cvitanich, C., Braunschweig, U., ... & Jensen, E. Ø. (2007). The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc in vitro and TSO1 is required for both male and female fertility in Arabidopsis thaliana. Journal of experimental botany, 58(13), 3657-3670.

Angeles-Núnez, J. G., & Tiessen, A. (2012). Regulation of AtSUS2 and AtSUS3 by glucose and the transcription factor LEC2 in different tissues and at different stages of Arabidopsis seed development. Plant molecular biology, 78(4-5), 377-392.

Armenta-Medina, A., Demesa-Arévalo, E., & Vielle-Calzada, J. P. (2011). Epigenetic control of cell specification during female gametogenesis. Sexual plant reproduction, 24(2), 137-147.

Armstrong, S. J., Caryl, A. P., Jones, G. H., & Franklin, F. C. H. (2002). Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. Journal of cell science, 115(18), 3645-3655.

Bao, Z., & Hua, J. (2014). Interaction of CPR5 with cell cycle regulators UVI4 and OSD1 in Arabidopsis. PLoS One, 9(6), e100347.

Barcaccia, G., & Albertini, E. (2013). Apomixis in plant reproduction: a novel perspective on an old dilemma. Plant Reproduction, 26(3), 159-179.

Barrett, S. C. (2002). Evolution of sex: the evolution of plant sexual diversity. Nature reviews genetics, 3(4), 274.

Baudat, F., Manova, K., Yuen, J. P., Jasin, M., & Keeney, S. (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Molecular cell, 6(5), 989-998.

Baumbusch, L. O., Hughes, D. W., Galau, G. A., & Jakobsen, K. S. (2004). LEC1, FUS3, ABI3 and Em expression reveals no correlation with dormancy in Arabidopsis. Journal of experimental botany, 55(394), 77-87.

Beck, H., Nähse-Kumpf, V., Larsen, M. S. Y., O'Hanlon, K. A., Patzke, S., Holmberg, C., ... & Sørensen, C. S. (2012). Cyclin-dependent kinase suppression by WEE1 kinase protects the genome through control of replication initiation and nucleotide consumption. Molecular and cellular biology, 32(20), 4226-4236.

Belide, S., Zhou, X. R., Kennedy, Y., Lester, G., Shrestha, P., Petrie, J. R., & Singh, S. P. (2013). Rapid expression and validation of seed-specific constructs in transgenic LEC2 induced somatic embryos of Brassica napus. Plant Cell, Tissue and Organ Culture (PCTOC), 113(3), 543-553.

Bencivenga, S., Simonini, S., Benková, E., & Colombo, L. (2012). The transcription factors BEL1 and SPL are required for cytokinin and auxin signaling during ovule development in Arabidopsis. The Plant Cell, 24(7), 2886-2897.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. genesis, 53(8), 474-485.

Bhatt, A. M., Lister, C., Page, T., Fransz, P., Findlay, K., Jones, G. H., ... & Dean, C. (1999). The DIF1 gene of Arabidopsis is required for meiotic chromosome segregation and belongs to the REC8/RAD21 cohesin gene family. The Plant Journal, 19(4), 463-472.

Birky, C. W. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics, 144(1), 427-437.

Bishop, D. K. (1994). RecA homologs Dmc1 and Rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. Cell, 79(6), 1081-1092.

Boden, S. A., Shadiac, N., Tucker, E. J., Langridge, P., & Able, J. A. (2007). Expression and functional analysis of TaASY1 during meiosis of bread wheat (Triticum aestivum). BMC Molecular Biology, 8(1), 65.

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., ... & Custers, J. B. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. The Plant Cell, 14(8), 1737-1749.

Brar, G. A., Kiburz, B. M., Zhang, Y., Kim, J. E., White, F., & Amon, A. (2006). Rec8 phosphorylation and recombination promote the step-wise loss of cohesins in meiosis. Nature, 441(7092), 532.

Braybrook, S. A., & Harada, J. J. (2008). LECs go crazy in embryo development. Trends in plant science, 13(12), 624-630.

Buunk, A. P., & Massar, K. (2014). A night on the town: When the importance of mate acquisition overrides intrasexual competition. Anthropological Review, 77(3), 273-285.

Bürger, R. (1999). Evolution of genetic variability and the advantage of sex and recombination in changing environments. Genetics, 153(2), 1055-1069.

Cai, X., Dong, F., Edelmann, R. E., & Makaroff, C. A. (2003). The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing. Journal of Cell Science, 116(14), 2999-3007.

Calvo, O., & Manley, J. L. (2003). Strange bedfellows: polyadenylation factors at the promoter. Genes & Development, 17(11), 1321-1327.

Campos, J. L., Charlesworth, B., & Haddrill, P. R. (2012). Molecular evolution in nonrecombining regions of the Drosophila melanogaster genome. Genome biology and evolution, 4(3), 278-288.

Caryl, A. P., Armstrong, S. J., Jones, G. H., & Franklin, F. C. H. (2000). A homologue of the yeast HOP1 gene is inactivated in the Arabidopsis meiotic mutant asy1. Chromosoma, 109(1-2), 62-71.

Chardin, C., Girin, T., Roudier, F., Meyer, C., & Krapp, A. (2014). The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development. Journal of Experimental Botany, 65(19), 5577-5587.

Charlesworth, D. (2002). Plant sex determination and sex chromosomes. Heredity, 88(2), 94.

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. Nature reviews genetics, 10(11), 783.

Charnov, E. L. (1979). Simultaneous hermaphroditism and sexual selection. Proceedings of the National Academy of Sciences, 76(5), 2480-2484.

Che, L., Tang, D., Wang, K., Wang, M., Zhu, K., Yu, H., ... & Cheng, Z. (2011). OsAM1 is required for leptotene-zygotene transition in rice. Cell research, 21(4), 654.

Chelysheva, L., Diallo, S., Vezon, D., Gendrot, G., Vrielynck, N., Belcram, K., ... & Mercier, R. (2005). AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. Journal of cell science, 118(20), 4621-4632.

Chen, G. H., Sun, J. Y., Liu, M., Liu, J., & Yang, W. C. (2014). SPOROCYTELESS is a novel embryophyte-specific transcription repressor that interacts with TPL and TCP proteins in Arabidopsis. Journal of Genetics and Genomics, 41(12), 617-625.

Chen, J. M., Cooper, D. N., Chuzhanova, N., Férec, C., & Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. Nature Reviews Genetics, 8(10), 762.

Chen, L., Miyazaki, C., Kojimai, A., Saito, A., & Adachi, T. (1999). Isolation and characterization of a gene expressed during early embryo sac development in apomictic guinea grass (Panicum maximum). Journal of plant physiology, 154(1), 55-62.

Chen, L., Guan, L., Seo, M., Hoffmann, F., & Adachi, T. (2005). Developmental expression of ASG-1 during gametogenesis in apomictic guinea grass (Panicum maximum). Journal of plant physiology, 162(10), 1141-1148.

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Molecular Ecology Resources, 16(1), 138-149.

Chorev, M., & Carmel, L. (2012). The function of introns. Frontiers in genetics, 3.

Chung, B. Y., Simons, C., Firth, A. E., Brown, C. M., & Hellens, R. P. (2006). Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. BMC genomics, 7(1), 120.

Citterio, S., Albertini, E., Varotto, S., Feltrin, E., Soattin, M., Marconi, G., ... & Barcaccia, G. (2005). Alfalfa Mob1-like genes are expressed in reproductive organs during meiosis and gametogenesis. Plant molecular biology, 58(6), 789-807.

Cole, F., Keeney, S., & Jasin, M. (2010). Evolutionary conservation of meiotic DSB proteins: more than just Spo11. Genes & development, 24(12), 1201-1207.

Conner, J. A., Goel, S., Gunawan, G., Cordonnier-Pratt, M. M., Johnson, V. E., Liang, C., ... & Yang, L. (2008). Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of Pennisetum and Cenchrus. Plant physiology, 147(3), 1396-1411.

Conner, J. A., Mookkan, M., Huo, H., Chae, K., & Ozias-Akins, P. (2015). A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. Proceedings of the National Academy of Sciences, 112(36), 11205-11210.

Cook, G. S., Grønlund, A. L., Siciliano, I., Spadafora, N., Amini, M., Herbert, R. J., ... & Rogers, H. J. (2013). Plant WEE1 kinase is cell cycle regulated and removed at mitosis via the 26S proteasome machinery. Journal of experimental botany, 64(7), 2093-2106.

Corral, J. M., Vogel, H., Aliyu, O. M., Hensel, G., Thiel, T., Kumlehn, J., & Sharbel, T. F. (2013). A conserved apomixis-specific polymorphism is correlated with exclusive exonuclease expression in premeiotic ovules of apomictic Boechera species. Plant physiology, 163(4), 1660-1672.

Cosendai, A. C., Wagner, J., Ladinig, U., Rosche, C., & Hörandl, E. (2013). Geographical parthenogenesis and population genetic structure in the alpine species Ranunculus kuepferi (Ranunculaceae). Heredity, 110(6), 560.

Costa, L. M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., & Gutierrez-Marcos, J. F. (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. Current Biology, 22(2), 160-165.

Curtis, C. A., & Doyle, G. G. (1991). Double meiotic mutants of maize: implications for the genetic regulation of meiosis. Journal of Heredity, 82(2), 156-163.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D., & Doutriaux, M. P. (1999). Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of Arabidopsis. The Plant Cell, 11(9), 1623-1634.

Cromer, L., Heyman, J., Touati, S., Harashima, H., Araou, E., Girard, C., ... & Mercier, R. (2012). OSD1 promotes meiotic progression via APC/C inhibition and forms a regulatory network with TDM and CYCA1; 2/TAM. PLoS genetics, 8(7), e1002865.

Cueva, A., Concia, L., & Cella, R. (2012). Molecular characterization of a Cyrtochilum loxense somatic embryogenesis receptor-like kinase (SERK) gene expressed during somatic embryogenesis. Plant cell reports, 31(6), 1129-1139.

Cui, X., Guo, Z., Song, L., Wang, Y., & Cheng, Y. (2016). NCP1/AtMOB1A plays key roles in auxin-mediated Arabidopsis development. PLoS genetics, 12(3), e1005923.

Demesa-Arévalo, E., & Vielle-Calzada, J. P. (2013). The classical arabinogalactan protein AGP18 mediates megaspore selection in Arabidopsis. The Plant Cell, 25(4), 1274-1287.

Deng, W., Luo, K., Li, Z., & Yang, Y. (2009). A novel method for induction of plant regeneration via somatic embryogenesis. Plant science, 177(1), 43-48.

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., & Mercier, R. (2009). Turning meiosis into mitosis. PLoS biology, 7(6), e1000124.

d'Erfurth, I., Cromer, L., Jolivet, S., Girard, C., Horlow, C., Sun, Y., ... & Mercier, R. (2010). The cyclin-A CYCA1; 2/TAM is required for the meiosis I to meiosis II transition and cooperates with OSD1 for the prophase to first meiotic division transition. PLoS genetics, 6(6), e1000989.

De Schutter, K., Joubès, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., ... & De Veylder, L. (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. The Plant Cell, 19(1), 211-225.

De Visser, J. A. G., & Rozen, D. E. (2005). Limits to adaptation in asexual populations. Journal of evolutionary biology, 18(4), 779-788.

Disteche, C. M. (2012). Dosage compensation of the sex chromosomes. Annual review of genetics, 46, 537-560.

Doutriaux, M. P., Couteau, F., Bergounioux, C., & White, C. (1998). Isolation and characterisation of the RAD51 and DMC1 homologs from Arabidopsis thaliana. Molecular and General Genetics MGG, 257(3), 283-291.

Dowell, N. L., Giorgianni, M. W., Griffin, S., Kassner, V. A., Selegue, J. E., Sanchez, E. E., & Carroll, S. B. (2018). Extremely divergent haplotypes in two toxin gene complexes encode alternative venom types within rattlesnake species. Current Biology, 28(7), 1016-1026.

Dreni, L., & Zhang, D. (2016). Flower development: the evolutionary history and functions of the AGL6 subfamily MADS-box genes. Journal of experimental botany, 67(6), 1625-1638.

Dresselhaus, T. H. O. M. A. S., Carman, J. G., & Savidan, Y. (2001). Genetic engineering of apomixis in sexual crops: A critical assessment of the apomixis technology. Flowering of Apomixis: From Mechanisms to Genetic Engineering, Y. Savidan, JG Carman, and T. Dresselhaus, eds (Mexico: CIMMYT, IRD, European Commission DG VI), 229-243.

Duret, L. (2001). Why do genes have introns? Recombination might add a new piece to the puzzle. Trends in Genetics, 17(4), 172-175.

Durrant, W. E., Wang, S., & Dong, X. (2007). Arabidopsis SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response. Proceedings of the National Academy of Sciences, 104(10), 4223-4227.

Dvinge, H., & Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. Genome medicine, 7(1), 45.

Ebel, C., Mariconti, L., & Gruissem, W. (2004). Plant retinoblastoma homologues control nuclear proliferation in the female gametophyte. Nature, 429(6993), 776.

Eckardt, N. A. (2011). A role for ARGONAUTE in apomixis.

Elkon, R., Ugalde, A. P., & Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nature Reviews Genetics, 14(7), 496.

Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D., ... & Smyth, D. R. (1996). AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. The Plant Cell, 8(2), 155-168.

Evans, M. M. (2007). The indeterminate gametophyte1 gene of maize encodes a LOB domain protein required for embryo sac and leaf development. The Plant Cell, 19(1), 46-62.

Farquharson, K. L. (2013). Mirror, Mirror on the Wall: A Role for AGP18 in Functional Megaspore Selection.

Florez, S. L., Erwin, R. L., Maximova, S. N., Guiltinan, M. J., & Curtis, W. R. (2015). Enhanced somatic embryogenesis in Theobroma cacao using the homologous BABY BOOM transcription factor. BMC plant biology, 15(1), 121.

Funakoshi, Y., Doi, Y., Hosoda, N., Uchida, N., Osawa, M., Shimada, I., ... & Hoshino, S. I. (2007). Mechanism of mRNA deadenylation: evidence for a molecular interplay between translation termination factor eRF3 and mRNA deadenylases. Genes & development, 21(23), 3135-3148.

Furuta, K., Kubo, M., Sano, K., Demura, T., Fukuda, H., Liu, Y. G., ... & Kakimoto, T. (2011). The CKH2/PKL chromatin remodeling factor negatively regulates cytokinin responses in Arabidopsis calli. Plant and cell physiology, 52(4), 618-628.

Galego, L., & Almeida, J. (2002). Role of DIVARICATA in the control of dorsoventral asymmetry in Antirrhinum flowers. Genes & development, 16(7), 880-891.

Garcia-Aguilar, M., Michaud, C., Leblanc, O., & Grimanelli, D. (2010). Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes. The Plant Cell, 22(10), 3249-3267.

Gerashchenkov, G. A., Yasybaeva, G. R., Rozhnova, N. A., & Chemeris, A. V. (2015). Isolation of Promoters and Fragments of Genes Controlling Endosperm Development Without Fertilization in Arabidopsis and Engineering of the Antisense Constructions. European Journal of Molecular Biotechnology, (2), 56-62.

Giroux, M. J., Clancy, M., Baier, J., Ingham, L., McCarty, D., & Hannah, L. C. (1994). De novo synthesis of an intron by the maize transposable element Dissociation. Proceedings of the National Academy of Sciences, 91(25), 12150-12154.

Golubovskaya, I., Grebennikova, Z. K., Avalkina, N. A., & Sheridan, W. F. (1993). The role of the ameiotic1 gene in the initiation of meiosis and in subsequent meiotic events in maize. Genetics, 135(4), 1151-1166.

Golubovskaya, I. N., Hamant, O., Timofejeva, L., Wang, C. J. R., Braun, D., Meeley, R., & Cande, W. Z. (2006). Alleles of afd1 dissect REC8 functions during meiotic prophase I. Journal of cell science, 119(16), 3306-3315.

Gonzalez, N., Hernould, M., Delmas, F., Gévaudant, F., Duffe, P., Causse, M., ... & Chevalier, C. (2004). Molecular characterization of a WEE1 gene homologue in tomato (Lycopersicon esculentum Mill.). Plant molecular biology, 56(6), 849-861.

Grimanelli, D., Leblanc, O., Perotti, E., & Grossniklaus, U. (2001). Developmental genetics of gametophytic apomixis. TRENDS in Genetics, 17(10), 597-604.

Grishaeva, T. M., Dadashev, S. Y., & Bogdanov, Y. F. (2007). Meiotic Rec8 cohesins and their mitotic Rad21 orthologs: comparison in silico. Molecular Biology, 41(4), 674-676.

Groß-Hardt, R., Kägi, C., Baumann, N., Moore, J. M., Baskar, R., Gagliano, W. B., ... & Grossniklaus, U. (2007). LACHESIS restricts gametic cell fate in the female gametophyte of Arabidopsis. PLoS biology, 5(3), e47.

Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A., & Gagliano, W. B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science, 280(5362), 446-450.

Gruszczyńska, A., & Rakoczy-Trojanowska, M. (2011). Expression analysis of somatic embryogenesis-related SERK, LEC1, VP1 and NiR ortologues in rye (Secale cereale L.). Journal of applied genetics, 52(1), 1-8.

Guenter, D. (2017). Seed size variation in apomictic and sexual Boechera (Brassicaceae). Plant Science 492. University of Saskatchewan.

Guimarães, L. A., Dusi, D. M. D. A., Masiero, S., Resentini, F., Gomes, A. C. M., Silveira, É. D., ... & Carneiro, V. T. D. C. (2013). BbrizAGL6 is differentially expressed during embryo sac formation of apomictic and sexual Brachiaria brizantha plants. Plant molecular biology reporter, 31(6), 1397-1406.

Guitton, A. E., & Berger, F. (2005). Loss of function of MULTICOPY SUPPRESSOR OF IRA 1 produces nonviable parthenogenetic embryos in Arabidopsis. Current Biology, 15(8), 750-754.

Guo, F., Huang, B. Q., Han, Y., & Zee, S. Y. (2004). Fertilization in maize indeterminate gametophyte1 mutant. Protoplasma, 223(2-4), 111-120.

Haig, D., & Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. The American Naturalist, 134(1), 147-155.

Hand, M. L., & Koltunow, A. M. (2014). The genetic control of apomixis: asexual seed formation. Genetics, 197(2), 441-450.

Harding, E. W., Tang, W., Nichols, K. W., Fernandez, D. E., & Perry, S. E. (2003). Expression and maintenance of embryogenic potential is enhanced through constitutive expression of AGAMOUS-Like 15. Plant Physiology, 133(2), 653-663.

Hartung, F., & Puchta, H. (2000). Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. Nucleic acids research, 28(7), 1548-1554.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S., & Puchta, H. (2007). The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. The Plant Cell, 19(10), 3090-3099.

Hauser, B. A., He, J. Q., Park, S. O., & Gasser, C. S. (2000). TSO1 is a novel protein that modulates cytokinesis and cell expansion in Arabidopsis. Development, 127(10), 2219-2226.

Hecht, V., Vielle-Calzada, J. P., Hartog, M. V., Schmidt, E. D., Boutilier, K., Grossniklaus, U., & de Vries, S. C. (2001). The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiology, 127(3), 803-816.

Heidmann, I., De Lange, B., Lambalk, J., Angenent, G. C., & Boutilier, K. (2011). Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. Plant cell reports, 30(6), 1107-1115.

Hellsten, U., Aspden, J. L., Rio, D. C., & Rokhsar, D. S. (2011). A segmental genomic duplication generates a functional intron. Nature communications, 2, 454.

Henderson, J. T., Li, H. C., Rider, S. D., Mordhorst, A. P., Romero-Severson, J., Cheng, J. C., ... & Ogas, J. (2004). PICKLE acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. Plant Physiology, 134(3), 995-1005.

Hendrichs, J., Katsoyannos, B. I., Papaj, D. R., & Prokopy, R. J. (1991). Sex differences in movement between natural feeding and mating sites and tradeoffs between food consumption, mating success and predator evasion in Mediterranean fruit flies (Diptera: Tephritidae). Oecologia, 86(2), 223-231.

Hörandl, E. (2006). The complex causality of geographical parthenogenesis. New Phytologist, 171(3), 525-538.

Hörandl, E. (2009). Geographical parthenogenesis: opportunities for asexuality. In Lost sex (pp. 161-186). Springer, Dordrecht.

Hörandl, E., & Hojsgaard, D. (2012). The evolution of apomixis in angiosperms: a reappraisal. Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, 146(3), 681-693.

Huanca-Mamani, W., Garcia-Aguilar, M., León-Martínez, G., Grossniklaus, U., & Vielle-Calzada, J. P. (2005). CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 102(47), 17231-17236.

Huang, B. Q., & Sheridan, W. F. (1996). Embryo sac development in the maize indeterminate gametophyte1 mutant: abnormal nuclear behavior and defective microtubule organization. The Plant Cell, 8(8), 1391-1407.

Huang, B. Q., & Sheridan, W. F. (1998). Actin coronas in normal and indeterminate gametophyte1 embryo sacs of maize. Sexual plant reproduction, 11(5), 257-264.

Ingouff, M., Haseloff, J., & Berger, F. (2005). Polycomb group genes control developmental timing of endosperm. The Plant Journal, 42(5), 663-674.

Ishihama, A. (2000). Functional modulation of Escherichia coli RNA polymerase. Annual Reviews in Microbiology, 54(1), 499-518.

Jin, Y., Eser, U., Struhl, K., & Churchman, L. S. (2017). The ground state and evolution of promoter region directionality. Cell, 170(5), 889-898.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096), 816-821.

Jullien, P. E., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., & Berger, F. (2008). Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS biology, 6(8), e194.

Junker, A., & Bäumlein, H. (2012). Multifunctionality of the LEC1 transcription factor during plant development. Plant signaling & behavior, 7(12), 1718-1720.

Kantama, L., Lambert, Y., Hu, H., de Jong, H., de Vries, S. C., & Russinova, E. (2006). Use of the SSLP-based method for detection of rare apomictic events in a sexual AtSERK1 transgenic Arabidopsis population. Sexual plant reproduction, 19(2), 73-82.

Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., ... & Hattori, T. (2005). Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. Plant and Cell Physiology, 46(2), 300-311.

Kathiresan, A., Khush, G. S., & Bennett, J. (2002). Two rice DMC1 genes are differentially expressed during meiosis and during haploid and diploid mitosis. Sexual Plant Reproduction, 14(5), 257-267.

Keeney, S., Giroux, C. N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell, 88(3), 375-384.

Keightley, P. D., & Otto, S. P. (2006). Interference among deleterious mutations favours sex and recombination in finite populations. Nature, 443(7107), 89.

Khanday, I., Skinner, D., Yang, B., Mercier, R., & Sundaresan, V. (2019). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. Nature, 565(7737), 91.

Kim, H. U., Jung, S. J., Lee, K. R., Kim, E. H., Lee, S. M., Roh, K. H., & Kim, J. B. (2014). Ectopic overexpression of castor bean LEAFY COTYLEDON2 (LEC2) in Arabidopsis triggers the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. FEBS open bio, 4, 25-32.

Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (1999). Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. The Plant Cell, 11(10), 1945-1952.

Kinoshita, T., Ikeda, Y., & Ishikawa, R. (2008, December). Genomic imprinting: a balance between antagonistic roles of parental chromosomes. In Seminars in cell & developmental biology (Vol. 19, No. 6, pp. 574-579). Academic Press.

Klapholz, S., Waddell, C. S., & Esposito, R. E. (1985). The role of the SPO11 gene in meiotic recombination in yeast. Genetics, 110(2), 187-216.

Klimaszewska, K., Pelletier, G., Overton, C., Stewart, D., & Rutledge, R. G. (2010). Hormonally regulated overexpression of Arabidopsis WUS and conifer LEC1 (CHAP3A) in transgenic white spruce: implications for somatic embryo development and somatic seedling growth. Plant cell reports, 29(7), 723-734.

Klucher, K. M., Chow, H., Reiser, L., & Fischer, R. L. (1996). The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. The Plant Cell, 8(2), 137-153.

Koch, M. A., Dobeš, C., & Mitchell-Olds, T. (2003). Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American Arabis divaricarpa (Brassicaceae). Molecular Biology and Evolution, 20(3), 338-350.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., & Gruissem, W. (2003). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. The EMBO journal, 22(18), 4804-4814.

Köhler, C., & Weinhofer-Molisch, I. (2010). Mechanisms and evolution of genomic imprinting in plants. Heredity, 105(1), 57.

Kolkman, J. A., & Stemmer, W. P. (2001). Directed evolution of proteins by exon shuffling. Nature biotechnology, 19(5), 423.

Koltunow, A. M. (1993). Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. The plant cell, 5(10), 1425-1437.

Koltunow, A. M., & Grossniklaus, U. (2003). Apomixis: a developmental perspective. Annual review of plant biology, 54(1), 547-574.

Kőszegi, D., Johnston, A. J., Rutten, T., Czihal, A., Altschmied, L., Kumlehn, J., ... & Bäumlein, H. (2011). Members of the RKD transcription factor family induce an egg cell-like gene expression program. The Plant Journal, 67(2), 280-291.

Krizek, B. A. (2003). AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. Nucleic acids research, 31(7), 1859-1868.

Krizek, B. A. (2011). Auxin regulation of Arabidopsis flower development involves members of the AINTEGUMENTA-LIKE/PLETHORA (AIL/PLT) family. Journal of Experimental Botany, 62(10), 3311-3319.

Kurzbauer, M. T., Uanschou, C., Chen, D., & Schlögelhofer, P. (2012). The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in Arabidopsis. The Plant Cell, 24(5), 2058-2070.

Keightley, P. D., & Lynch, M. (2003). Toward a realistic model of mutations affecting fitness. Evolution, 57(3), 683-685.

Laspina, N. V., Vega, T., Seijo, J. G., González, A. M., Martelotto, L. G., Stein, J., ... & Pessino, S. C. (2008). Gene expression analysis at the onset of aposporous apomixis in Paspalum notatum. Plant molecular biology, 67(6), 615-628.

Lenhard, M., Bohnert, A., Jürgens, G., & Laux, T. (2001). Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between WUSCHEL and AGAMOUS. Cell, 105(6), 805-814.

Li, S., Yang, D., & Zhu, Y. (2007). Characterization and use of male sterility in hybrid rice breeding. Journal of Integrative Plant Biology, 49(6), 791-804.

Liu, D. D., Dong, Q. L., Sun, C., Wang, Q. L., You, C. X., Yao, Y. X., & Hao, Y. J. (2012). Functional characterization of an apple apomixis-related MhFIE gene in reproduction development. Plant science, 185, 105-111.

Lovell, J. T., Aliyu, O. M., Mau, M., Schranz, M. E., Koch, M., Kiefer, C., ... & Sharbel, T. F. (2013). On the origin and evolution of apomixis in Boechera. Plant reproduction, 26(4), 309-315.

Lovell, J. T., Williamson, R. J., Wright, S. I., McKay, J. K., & Sharbel, T. F. (2017). Mutation accumulation in an asexual relative of Arabidopsis. PLoS genetics, 13(1), e1006550.

Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M. J., ... & Wang, L. (2016). Morphogenic regulators Baby boom and Wuschel improve monocot transformation. The Plant Cell, 28(9), 1998-2015.

Lundmark, M., & Saura, A. (2006). Asexuality alone does not explain the success of clonal forms in insects with geographical parthenogenesis. Hereditas, 143(2006), 23-32.

Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J., & Chaudhury, A. (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proceedings of the National Academy of Sciences, 97(19), 10637-10642.

Lynch, M., Conery, J., & Burger, R. (1995). Mutation accumulation and the extinction of small populations. The American Naturalist, 146(4), 489-518.

Ma, H., Song, T., Wang, T., & Wang, S. (2016). Influence of human p53 on plant development. PloS one, 11(9), e0162840.

Ma, H., Yanofsky, M. F., & Meyerowitz, E. M. (1991). AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. Genes & Development, 5(3), 484-495.

Maillot, P., Lebel, S., Schellenbaum, P., Jacques, A., & Walter, B. (2009). Differential regulation of SERK, LEC1-Like and Pathogenesis-Related genes during indirect secondary somatic embryogenesis in grapevine. Plant Physiology and Biochemistry, 47(8), 743-752.

Mandáková, T., & Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x= 7 crucifer species (Brassicaceae). The Plant Cell, 20(10), 2559-2570.

Mandáková, T., Schranz, M. E., Sharbel, T. F., de Jong, H., & Lysak, M. A. (2015). Karyotype evolution in apomictic Boechera and the origin of the aberrant chromosomes. The Plant Journal, 82(5), 785-793.

Mau, M., Corral, J. M., Vogel, H., Melzer, M., Fuchs, J., Kuhlmann, M., ... & Sharbel, T. F. (2013). The conserved chimeric transcript UPGRADE2 is associated with unreduced pollen formation and is exclusively found in apomictic Boechera species. Plant physiology, 163(4), 1640-1659.

Mau, M., Lovell, J. T., Corral, J. M., Kiefer, C., Koch, M. A., Aliyu, O. M., & Sharbel, T. F. (2015). Hybrid apomicts trapped in the ecological niches of their sexual ancestors. Proceedings of the National Academy of Sciences, 112(18), E2357-E2365.

Mazumder, B., Seshadri, V., & Fox, P. L. (2003). Translational control by the 3′-UTR: the ends specify the means. Trends in biochemical sciences, 28(2), 91-98.

Mercier, R., Vezon, D., Bullier, E., Motamayor, J. C., Sellier, A., Lefèvre, F., ... & Horlow, C. (2001). SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. Genes & development, 15(14), 1859-1871.

Mercier, R., Armstrong, S. J., Horlow, C., Jackson, N. P., Makaroff, C. A., Vezon, D., ... & Franklin, F. C. H. (2003). The meiotic protein SWI1 is required for axial element formation and recombination initiation in Arabidopsis. Development, 130(14), 3309-3318.

Merino, S. T., Cummings, W. J., Acharya, S. N., & Zolan, M. E. (2000). Replication-dependent early meiotic requirement for Spo11 and Rad50. Proceedings of the National Academy of Sciences, 97(19), 10477-10482.

Mieulet, D., Jolivet, S., Rivard, M., Cromer, L., Vernet, A., Mayonove, P., ... & Mercier, R. (2016). Turning rice meiosis into mitosis. Cell research, 26(11), 1242.

Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). Untranslated regions of mRNAs. Genome biology, 3(3), reviews0004-1.

Mizukami, Y., & Fischer, R. L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proceedings of the National Academy of Sciences, 97(2), 942-947.

Moll, C., Von Lyncker, L., Zimmermann, S., Kägi, C., Baumann, N., Twell, D., ... & Groß-Hardt, R. (2008). CLO/GFA1 and ATO are novel regulators of gametic cell fate in plants. The Plant Journal, 56(6), 913-921.

Molnar, M., Bähler, J., Sipiczki, M., & Kohli, J. (1995). The rec8 gene of Schizosaccharomyces pombe is involved in linear element formation, chromosome pairing and sister-chromatid cohesion during meiosis. Genetics, 141(1), 61-73.

Morran, L. T., Schmidt, O. G., Gelarden, I. A., Parrish, R. C., & Lively, C. M. (2011). Running with the Red Queen: host-parasite coevolution selects for biparental sex. Science, 333(6039), 216-218.

Motamayor, J. C., Vezon, D., Bajon, C., Sauvanet, A., Grandjean, O., Marchand, M., ... & Horlow, C. (2000). Switch (swi1), an Arabidopsis thaliana mutant affected in the female meiotic switch. Sexual Plant Reproduction, 12(4), 209-218.

Muller, H. J. (1964). The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 1(1), 2-9.

MUMMENHOFF, K., FRANZKE, A., & KOCH, M. (1997). Molecular data reveal convergence in fruit characters used in the classification of Thlaspi sl (Brassicaceae. Botanical Journal of the Linnean Society, 125(3), 183-199.

Nan, G. L., Ronceret, A., Wang, R. C., Fernandes, J. F., Cande, W. Z., & Walbot, V. (2011). Global transcriptome analysis of two ameiotic1 alleles in maize anthers: defining steps in meiotic entry and progression through prophase I. BMC plant biology, 11(1), 120.

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., & Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. The Plant Journal, 39(6), 877-885.

Nic-Can, G. I., López-Torres, A., Barredo-Pool, F., Wrobel, K., Loyola-Vargas, V. M., Rojas-Herrera, R., & De-la-Peña, C. (2013). New insights into somatic embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHEL-RELATED HOMEOBOX4 are epigenetically regulated in Coffea canephora. PLoS One, 8(8), e72160.

Nogler, G. A. (1984). Gametophytic apomixis. In Embryology of angiosperms (pp. 475-518). Springer, Berlin, Heidelberg.

Nonomura, K. I., Nakano, M., Murata, K., Miyoshi, K., Eiguchi, M., Miyao, A., ... & Kurata, N. (2004). An insertional mutation in the rice PAIR2 gene, the ortholog of Arabidopsis ASY1, results in a defect in homologous chromosome pairing during meiosis. Molecular Genetics and Genomics, 271(2), 121-129.

Ogas, J., Cheng, J. C., Sung, Z. R., & Somerville, C. (1997). Cellular differentiation regulated by gibberellin in the Arabidopsis thaliana pickle mutant. Science, 277(5322), 91-94.

Ogas, J., Kaufmann, S., Henderson, J., & Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proceedings of the National Academy of Sciences, 96(24), 13839-13844.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J. J., ... & Fischer, R. L. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. The Plant Cell, 11(3), 407-415.

Olesen, C., Møller, M., & Byskov, A. G. (2004). Tesmin transcription is regulated differently during male and female meiosis. Molecular Reproduction and Development: Incorporating Gamete Research, 67(1), 116-126.

O'Neill, R. J. W., Brennan, F. E., Delbridge, M. L., Crozier, R. H., & Graves, J. A. M. (1998). De novo insertion of an intron into the mammalian sex determining gene, SRY. Proceedings of the National Academy of Sciences, 95(4), 1653-1657.

Otto, S. P. (2009). The evolutionary enigma of sex. the american naturalist, 174(S1), S1-S14.

Pagnussat, G. C., Yu, H. J., & Sundaresan, V. (2007). Cell-fate switch of synergid to egg cell in Arabidopsis eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene BLH1. The Plant Cell, 19(11), 3578-3592.

Palazzo, A. F., & Gregory, T. R. (2014). The case for junk DNA. PLoS genetics, 10(5), e1004351.

Pawlowski, W. P., Wang, C. J. R., Golubovskaya, I. N., Szymaniak, J. M., Shi, L., Hamant, O., ... & Cande, W. Z. (2009). Maize AMEIOTIC1 is essential for multiple early meiotic processes and likely required for the initiation of meiosis. Proceedings of the National Academy of Sciences, 106(9), 3603-3608.

Penninckx, I. A., Eggermont, K., Terras, F. R., Thomma, B. P., De Samblanx, G. W., Buchala, A., ... & Broekaert, W. F. (1996). Pathogen-induced systemic activation of a plant defensin gene in Arabidopsis follows a salicylic acid-independent pathway. The Plant Cell, 8(12), 2309-2323.

Pennington, P. D., Costa, L. M., Gutierrez-Marcos, J. F., Greenland, A. J., & Dickinson, H. G. (2008). When genomes collide: aberrant seed development following maize interploidy crosses. Annals of botany, 101(6), 833-843.

Petukhova, G. V., Pezza, R. J., Vanevski, F., Ploquin, M., Masson, J. Y., & Camerini-Otero, R. D. (2005). The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. Nature structural & molecular biology, 12(5), 449.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L., & Poethig, R. S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & development, 18(19), 2368-2379.

Perry, S. E., Lehti, M. D., & Fernandez, D. E. (1999). The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. Plant Physiology, 120(1), 121-130.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., & Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. Gene, 276(1-2), 73-81.

Peters, A. D., & Lively, C. M. (1999). The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. The American Naturalist, 154(4), 393-405.

Podio, M., Felitti, S. A., Siena, L. A., Delgado, L., Mancini, M., Seijo, J. G., ... & Ortiz, J. P. A. (2014). Characterization and expression analysis of SOMATIC EMBRYOGENESIS RECEPTOR KINASE (SERK) genes in sexual and apomictic Paspalum notatum. Plant molecular biology, 84(4-5), 479-495.

Policansky, D. (1982). Sex change in plants and animals. Annual Review of Ecology and Systematics, 13(1), 471-495.

Pradillo, M., López, E., Linacero, R., Romero, C., Cuñado, N., Sánchez-Morán, E., & Santos, J. L. (2012). Together yes, but not coupled: new insights into the roles of RAD51 and DMC1 in plant meiotic recombination. The Plant Journal, 69(6), 921-933.

Prince, V. E., & Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. Nature Reviews Genetics, 3(11), 827.

Puizina, J., Siroky, J., Mokros, P., Schweizer, D., & Riha, K. (2004). Mre11 deficiency in Arabidopsis is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. The Plant Cell, 16(8), 1968-1978.

Qu, F., Ye, X., & Morris, T. J. (2008). Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. Proceedings of the National Academy of Sciences, 105(38), 14732-14737.

Quarin, C. L. (1999). Effect of pollen source and pollen ploidy on endosperm formation and seed set in pseudogamous apomictic Paspalum notatum. Sexual Plant Reproduction, 11(6), 331-335.

Rajeswaran, R., Aregger, M., Zvereva, A. S., Borah, B. K., Gubaeva, E. G., & Pooggin, M. M. (2012). Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. Nucleic acids research, 40(13), 6241-6254.

Randall, R. S., Miyashima, S., Blomster, T., Zhang, J., Elo, A., Karlberg, A., ... & Blajecka, K. (2015). AINTEGUMENTA and the D-type cyclin CYCD3; 1 regulate root secondary growth and respond to cytokinins. Biology open, 4(10), 1229-1236.

Ravi, M., Marimuthu, M. P., & Siddiqi, I. (2008). Gamete formation without meiosis in Arabidopsis. Nature, 451(7182), 1121.

Reddy, P. S., Mahanty, S., Kaul, T., Nair, S., Sopory, S. K., & Reddy, M. K. (2008). A high-throughput genome-walking method and its use for cloning unknown flanking sequences. Analytical biochemistry, 381(2), 248-253.

Richter, K. S., Serra, H., White, C. I., & Jeske, H. (2016). The recombination mediator RAD51D promotes geminiviral infection. Virology, 493, 113-127.

Rider Jr, S., Henderson, J. T., Jerome, R. E., Edenberg, H. J., Romero-Severson, J., & Ogas, J. (2003). Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis. The Plant Journal, 35(1), 33-43.

Rodrigues, J. C., Tucker, M. R., Johnson, S. D., Hrmova, M., & Koltunow, A. M. (2008). Sexual and apomictic seed formation in Hieracium requires the plant polycomb-group gene FERTILIZATION INDEPENDENT ENDOSPERM. The Plant Cell, 20(9), 2372-2386.

Rodrigues, J. C., Okada, T., Johnson, S. D., & Koltunow, A. M. (2010). A MULTICOPY SUPPRESSOR OF IRA1 (MSI1) homologue is not associated with the switch to autonomous seed development in apomictic (asexual) Hieracium plants. Plant science, 179(6), 590-597.

Rojek, J., Kapusta, M., Kozieradzka-Kiszkurno, M., Majcher, D., Górniak, M., Sliwinska, E., ... & Bohdanowicz, J. (2018). Establishing the cell biology of apomictic reproduction in diploid Boechera stricta (Brassicaceae). Annals of botany, 122(4), 513-539.

Romanienko, P. J., & Camerini-Otero, R. D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. Molecular cell, 6(5), 975-987.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution, 34(12), 3299-3302.

Rushworth, C. A., SONG, B. H., LEE, C. R., & MITCHELL-OLDS, T. H. O. M. A. S. (2011). Boechera, a model system for ecological genomics. Molecular ecology, 20(23), 4843-4857.

Sanchez-Moran, E., Santos, J. L., Jones, G. H., & Franklin, F. C. H. (2007). ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. Genes & development, 21(17), 2220-2233.

Santos, M. O., & Aragão, F. J. (2009). Role of SERK genes in plant environmental response. Plant signaling & behavior, 4(12), 1111-1113.

Savona, M., Mattioli, R., Nigro, S., Falasca, G., Della Rovere, F., Costantino, P., ... & Altamura, M. M. (2011). Two SERK genes are markers of pluripotency in Cyclamen persicum Mill. Journal of experimental botany, 63(1), 471-488.

Schallau, A., Arzenton, F., Johnston, A. J., Hähnel, U., Koszegi, D., Blattner, F. R., ... & Bäumlein, H. (2010). Identification and genetic analysis of the APOSPORY locus in Hypericum perforatum L. The Plant Journal, 62(5), 773-784.

Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, Grossniklaus U. Transcriptome analysis of the Arabidopsis megaspore mother cell uncovers the importance of RNA helicases for plant germline development. PLoS Biol. 9(9):e1001155. doi: 10.1371/journal.pbio.1001155.

Schranz, M. E., Dobeš, C., Koch, M. A., & Mitchell-Olds, T. (2005). Sexual reproduction, hybridization, apomixis, and polyploidization in the genus Boechera (Brassicaceae). American Journal of Botany, 92(11), 1797-1810.

Schranz, M. E., Lysak, M. A., & Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends in plant science, 11(11), 535-542.

Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevalier, D., Wisman, E., & Schneitz, K. (1999). Molecular analysis of NOZZLE, a gene involved in pattern formation and early sporogenesis during sex organ development in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 96(20), 11664-11669.

Scott, R. J., Spielman, M., Bailey, J., & Dickinson, H. G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. Development, 125(17), 3329-3341.

Seeliger, K., Dukowic-Schulze, S., Wurz-Wildersinn, R., Pacher, M., & Puchta, H. (2012). BRCA2 is a mediator of RAD51-and DMC1-facilitated homologous recombination in Arabidopsis thaliana. New Phytologist, 193(2), 364-375.

Serivichyaswat, P., Ryu, H. S., Kim, W., Kim, S., Chung, K. S., Kim, J. J., & Ahn, J. H. (2015). Expression of the floral repressor miRNA156 is positively regulated by the AGAMOUS-like proteins AGL15 and AGL18. Molecules and cells, 38(3), 259.

Serra, H., Da Ines, O., Degroote, F., Gallego, M. E., & White, C. I. (2013). Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination. PLoS genetics, 9(11), e1003971.

Sezer, F., Yüzbaşioğlu, G., Özbilen, A., & Taşkin, K. M. (2016). Genome-wide identification and expression analysis of SWI1 genes in Boechera species. Computational biology and chemistry, 62, 75-81.

Sharbel, T. F., Voigt, M. L., Corral, J. M., Thiel, T., Varshney, A., Kumlehn, J., ... & Rotter, B. (2009). Molecular signatures of apomictic and sexual ovules in the Boechera holboellii complex. The Plant Journal, 58(5), 870-882.

Sharbel, T. F., Voigt, M. L., Corral, J. M., Galla, G., Kumlehn, J., Klukas, C., ... & Rotter, B. (2010). Apomictic and sexual ovules of Boechera display heterochronic global gene expression patterns. The Plant Cell, 22(3), 655-671.

Sharp, N. P., & Agrawal, A. F. (2018). An experimental test of the mutation-selection balance model for the maintenance of genetic variance in fitness components. Proceedings of the Royal Society B, 285(1890), 20181864.

Shibukawa, T., Yazawa, K., Kikuchi, A., & Kamada, H. (2009). Possible involvement of DNA methylation on expression regulation of carrot LEC1 gene in its 5′-upstream region. Gene, 437(1-2), 22-31.

Shimotohno, A., Ohno, R., Bisova, K., Sakaguchi, N., Huang, J., Koncz, C., ... & Umeda, M. (2006). Diverse phosphoregulatory mechanisms controlling cyclin-dependent kinase-activating kinases in Arabidopsis. The Plant Journal, 47(5), 701-710x.

Singh, M., Burson, B. L., & Finlayson, S. A. (2007). Isolation of candidate genes for apomictic development in buffelgrass (Pennisetum ciliare). Plant molecular biology, 64(6), 673-682.

Singh, M., Goel, S., Meeley, R. B., Dantec, C., Parrinello, H., Michaud, C., ... & Grimanelli, D. (2011). Production of viable gametes without meiosis in maize deficient for an ARGONAUTE protein. The Plant Cell, 23(2), 443-458.

Singla, B., Khurana, J. P., & Khurana, P. (2009). Structural characterization and expression analysis of the SERK/SERL gene family in rice (Oryza sativa). International journal of plant genomics, 2009.

Smyth, G. K., Thorne, N., & Wettenhall, J. (2004). Limma: Linear Models for Microarray, User's Guide. Software manual available from http://bioinf. wehi. edu. au/limma.

Soll, S. J., Arenas, C. D., & Lehman, N. (2007). Accumulation of deleterious mutations in small abiotic populations of RNA. Genetics, 175(1), 267-275.

Sonnhammer, E. L., & Östlund, G. (2014). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic acids research, 43(D1), D234-D239.

Sorrell, D. A., Marchbank, A., McMahon, K., Dickinson, R. J., Rogers, H. J., & Francis, D. (2002). A WEE1 homologue from Arabidopsis thaliana. Planta, 215(3), 518-522.

Spillane, C., MacDougall, C., Stock, C., Köhler, C., Vielle-Calzada, J. P., Nunes, S. M., ... & Goodrich, J. (2000). Interaction of the Arabidopsis polycomb group proteins FIE and MEA mediates their common phenotypes. Current Biology, 10(23), 1535-1538.

Spillane, C., Curtis, M. D., & Grossniklaus, U. (2004). Apomixis technology development—virgin births in farmers' fields?. Nature biotechnology, 22(6), 687.

Srinivasan, C., Liu, Z., Heidmann, I., Supena, E. D. J., Fukuoka, H., Joosen, R., ... & Boutilier, K. (2007). Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (Nicotiana tabacum L.). Planta, 225(2), 341.

Stacey, N. J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., ... & Sugimoto-Shirasu, K. (2006). Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. The Plant Journal, 48(2), 206-216.

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., ... & Harada, J. J. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proceedings of the National Academy of Sciences, 98(20), 11806-11811.

Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., ... & Harada, J. J. (2008). Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: implications for somatic embryogenesis. Proceedings of the National Academy of Sciences, 105(8), 3151-3156.

Székely, T., Freckleton, R. P., & Reynolds, J. D. (2004). Sexual selection explains Rensch's rule of size dimorphism in shorebirds. Proceedings of the National Academy of Sciences, 101(33), 12224-12227.

Taylor, J. S., & Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet., 38, 615-643.

Tedeschi, F., Rizzo, P., Rutten, T., Altschmied, L., & Bäumlein, H. (2017). RWP-RK domain-containing transcription factors control cell differentiation during female gametophyte development in Arabidopsis. New Phytologist, 213(4), 1909-1924.

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., ... & Scott, R. J. (2008). MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. The Plant Cell, 20(9), 2387-2398.

Triglia, T. (2000). Inverse PCR (IPCR) for obtaining promoter sequence. In Transcription Factor Protocols (pp. 79-84). Humana Press.

Tsuchiya, Y., Nambara, E., Naito, S., & McCourt, P. (2004). The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis. The Plant Journal, 37(1), 73-81.

Uanschou, C., Ronceret, A., Von Harder, M., De Muyt, A., Vezon, D., Pereira, L., ... & Grelon, M. (2013). Sufficient amounts of functional HOP2/MND1 complex promote interhomolog DNA repair but are dispensable for intersister DNA repair during meiosis in Arabidopsis. The Plant Cell, 25(12), 4924-4940.

UniProt Consortium. (2014). UniProt: a hub for protein information. Nucleic acids research, 43(D1), D204-D212.

Vijverberg, K., Milanovic-Ivanovic, S., Bakx-Schotman, T., & van Dijk, P. J. (2010). Genetic fine-mapping of DIPLOSPOROUS in Taraxacum (dandelion; Asteraceae) indicates a duplicated DIP-gene. BMC plant biology, 10(1), 154.

Vitulo, N., Vezzi, A., Galla, G., Citterio, S., Marino, G., Ruperti, B., ... & Barcaccia, G. (2007). Characterization and evolution of the cell cycle-associated Mob domain-containing proteins in eukaryotes. Evolutionary Bioinformatics, 3, 117693430700300007.

Völz, R., von Lyncker, L., Baumann, N., Dresselhaus, T., Sprunck, S., & Groß-Hardt, R. (2012). LACHESIS-dependent egg-cell signaling regulates the development of female gametophytic cells. Development, 139(3), 498-502.

Wang, H., Caruso, L. V., Downie, A. B., & Perry, S. E. (2004). The embryo MADS domain protein AGAMOUS-Like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. The Plant Cell, 16(5), 1206-1219.

Wang, Y., Xiao, R., Wang, H., Cheng, Z., Li, W., Zhu, G., ... & Ma, H. (2014). The Arabidopsis RAD51 paralogs RAD51B, RAD51D and XRCC2 play partially redundant roles in somatic DNA repair and gene regulation. New Phytologist, 201(1), 292-304.

Watanabe, Y., & Nurse, P. (1999). Cohesin Rec8 is required for reductional chromosome segregation at meiosis. Nature, 400(6743), 461.

Wei, B., Zhang, J., Pang, C., Yu, H., Guo, D., Jiang, H., ... & Qu, L. J. (2015). The molecular mechanism of SPOROCYTELESS/NOZZLE in controlling Arabidopsis ovule development. Cell research, 25(1), 121.

Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., & Marks, M. D. (2015). Gene duplication and divergence affecting drug content in Cannabis sativa. New Phytologist, 208(4), 1241-1250.

Welch, D. B. M., & Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science, 288(5469), 1211-1215.

Welch, J. L. M., Welch, D. B. M., & Meselson, M. (2004). Cytogenetic evidence for asexual evolution of bdelloid rotifers. Proceedings of the National Academy of Sciences, 101(6), 1618-1621.

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. L. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications, 18(1), 315-322.

Wilkie, G. S., Dickson, K. S., & Gray, N. K. (2003). Regulation of mRNA translation by 5′-and 3′-UTR-binding factors. Trends in biochemical sciences, 28(4), 182-188.

Wójcikowska, B., Jaskóła, K., Gąsiorek, P., Meus, M., Nowak, K., & Gaj, M. D. (2013). LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. Planta, 238(3), 425-440.

Wojcikowska, B., & Gaj, M. D. (2015). Auxin-related functions of LEAFY COTYLEDON2 gene in the induction of somatic embryogenesis in Arabidopsis. BioTechnologia. Journal of Biotechnology Computational Biology and Bionanotechnology, 96(1).

Xie, Z., Li, X., Glover, B. J., Bai, S., Rao, G. Y., Luo, J., & Yang, J. (2008). Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. Molecular biology and evolution, 25(8), 1581-1592.

Yadegari, R., Kinoshita, T., Lotan, O., Cohen, G., Katz, A., Choi, Y., ... & Fischer, R. L. (2000). Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. The plant cell, 12(12), 2367-2381.

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., & Hattori, T. (2009). Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. The Plant Journal, 58(5), 843-856.

Yamamoto, A., Kagaya, Y., Usui, H., Hobo, T., Takeda, S., & Hattori, T. (2010). Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. Plant and cell physiology, 51(12), 2031-2046.

Yang, W. C., Ye, D., Xu, J., & Sundaresan, V. (1999). The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes & Development, 13(16), 2108-2117.

Yazawa, K., & Kamada, H. (2007). Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1. Journal of experimental botany, 58(13), 3819-3828.

Yona, A. H., Alm, E. J., & Gore, J. (2018). Random sequences rapidly evolve into de novo promoters. Nature communications, 9(1), 1530.

Yuan, L., & Sundaresan, V. (2015). Spore formation in plants: Sporocyteless and more. Cell research, 25(1), 7.

Zuo, J., Niu, Q. W., Frugis, G., & Chua, N. H. (2002). The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. The Plant Journal, 30(3), 349-359.

Zhang, C., Ma, R., Xu, J., Yan, J., Guo, L., Song, J., ... & Yu, M. (2018). Genome-wide identification and classification of MYB superfamily genes in peach. PloS one, 13(6), e0199192.

Zhang, L., Tao, J., Wang, S., Chong, K., & Wang, T. (2006). The rice OsRad21-4, an orthologue of yeast Rec8 protein, is required for efficient meiosis. Plant molecular biology, 60(4), 533-554.

Zhang, H., Bishop, B., Ringenberg, W., Muir, W. M., & Ogas, J. (2012). The CHD3 remodeler PICKLE associates with genes enriched for trimethylation of histone H3 lysine 27. Plant physiology, 159(1), 418-432.

Zheng, Q., Zheng, Y., Ji, H., Burnie, W., & Perry, S. E. (2016). Gene regulation by the AGL15 transcription factor reveals hormone interactions in somatic embryogenesis. Plant Physiology, 172(4), 2374-2387.

Zimmer, C. (2015). Is most of our DNA garbage. The New York Times Magazine. New York, NY: The New York Times Company.