NOVEL INSIGHTS INTO THE PARTHENOGENETIC POTENTIAL OF NATURAL APOMICTIC BUFFELGRASS (*CENCHRUS CILIARIS*) FROM THE PERSPECTIVE OF EGG APPARATUS TRANSCRIPTOME

by

YUJI KE

(Under the Direction of Peggy Ozias-Akins)

ABSTRACT

Apomixis, a type of asexual reproduction in angiosperms, results in progenies that are genetically identical to the mother plant. It is a highly desirable trait in agriculture due to its potential to preserve heterosis of F_1 hybrids through subsequent generations. However, no major crops are apomictic. Deciphering mechanisms underlying apomixis becomes one of the alternatives to engineer self-producing capability into major crops. Parthenogenesis, a major component of apomixis commonly described as the ability to initiate embryo formation from the egg cell without fertilization, also can be valuable in plant breeding for doubled haploid production. Discovery of genes driving parthenogenesis in the natural apomict, *Cenchrus ciliaris*, has been challenging due to limited genomic resources and technical difficulties in accessing the egg cell expressing parthenogenesis. By conducting laser capture microdissection-based RNA-seq on sexual and apomictic egg apparatus on the day of anthesis, we created a de novo transcriptome with sequence information for the *Cenchrus ciliaris* egg apparatus, identified transcriptional profiles that distinguish apomictic egg from its sexual counterpart, and suggested functional roles for a few transcription factors in promoting natural parthenogenesis. Our

transcriptome data significantly complemented previous gene expression studies and will be an important resource for future research on natural parthenogenesis, as well as inquiries using sexual and apomictic egg cells that extend beyond those of parthenogenesis.

INDEX WORDS: *Cenchrus ciliaris*, egg apparatus, parthenogenesis, laser capture microdissection-based RNA-seq, apomixis, differential expression

NOVEL INSIGHTS INTO THE PARTHENOGENETIC POTENTIAL OF NATURAL APOMICTIC BUFFELGRASS (*CENCHRUS CILIARIS*) FROM THE PERSPECTIVE OF EGG APPARATUS TRANSCRIPTOME

by

YUJI KE

B.S., China Agricultural University, P.R.China, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

NOVEL INSIGHTS INTO THE PARTHENOGENETIC POTENTIAL OF NATURAL APOMICTIC BUFFELGRASS (*CENCHRUS CILIARIS*) FROM THE PERSPECTIVE OF EGG APPARATUS TRANSCRIPTOME

by

YUJI KE

Major Professor:	Peggy Ozias-Akins
Committee:	Chung-Jui Tsai
	Dayton Wilde
	Joann Conner
	Zenglu Li

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
December 2019

DEDICATION

For my parents, sister, uncle and the whole big families who encouraged and helped me along this journey.

For my dream, to pursue the highest science diploma in the United States.

ACKNOWLEDGEMENTS

First and foremost, many thanks to my advisor Dr. Peggy Ozias-Akins for bringing me in such an exciting project. Also, many thanks to her for kindly allowing me to freely explore, think and act under her guidance, for which I will always be grateful. To Dr. Maricel Podio, from whom I learnt a lot about how to do experiments in my early years of PhD, and her pioneer work and experience have always been valuable for me to complete this project. To my committee members, Dr. Joann Conner, much appreciation to her for sharing the molecular biology expertise and research experience with me that helped me to make progress with my experiments and data analyses; Dr. CJ Tsai, I can't do without her provision of a laser capture microscope, and many thanks for her patience and help with my relentless questioning on the homework assignments throughout her class; Dr. Dayton Wilde, thanks to him for allowing me to perform my preliminary research in his lab and greenhouse; Dr. Zenglu Li, I enjoyed his breeding class and his critical comments and suggestions throughout the committee meetings have always been useful for me.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Apomixis is a naturally occurring phenomenon in angiosperms where maternal cells within the female reproductive organ form embryos without meiosis and fertilization, leading to the generation of clonal seeds (Nogler 1984; Bicknell *et al.,* 2004; Ozias-Akins *et al.,* 2007). The very first documented discovery of apomixis in angiosperms could be traced back to the notice that a female plant of *Alchornea ilicifolia* produced seeds without any proven action of pollen (Smith 1841). Rare as such botanical observation was back then, apomixis has now been described in 148 genera as a form of adventitious embryony, 110 genera as aposporous and 68 genera as diplosporous among angiosperms (Hojsgaard *et al.,* 2014). While apomicts are by nature clonal and can produce seeds of identical genotype to the mother plant, no major crops have this self-producing capability. For crops like maize and pearl millet with proven heterosis (Moll *et al.,* 1965, Ouendeba *et al.,* 1993), breeding attempts have been made to transfer apomixis from their apomictic wild relatives *Tripsacum dactyloides* and *Pennisetum squamulatum*, respectively (De Wet *et al.,* 1973; Dujardin and Hanna 1983, 1985). However, it turned out not to be very successful in terms of apomixis-reproducing maize and pearl millet having agronomically unacceptable low seed set (Morgan *et al.,* 1998; Sokolov *et al.,* 1998; Savidan 2000; Savidan 2001). Unsuccessful as it is in generating agronomically acceptable genotypes of apomictic crops, further studies of these populations with contrasting reproduction mode helped identify and characterize genomic regions co-segregating with apomixis, and allowed further elucidation of the genetic control of apomixis (Ozias-Akins *et al.,* 1998). Further

understanding of the molecular mechanisms of apomixis using natural apomicts and their sexual counterparts becomes a vital and a more flexible alternative to enable major crops to self-clone through seed.

Sexual reproduction through seed in angiosperms

Angiosperm is a term etymologically derived from the Greek words angeion ("case") and sperma ("seed"), meaning plants that bear seeds within an enclosed structure. As the most diverse and numerous terrestrial plants, angiosperms comprise about 416 families, 13,164 genera and 295,383 species (Christenhusz *et al.,* 2016).

Also known as flowering plants, angiosperms develop flowers, the most prominent feature that distinguishes them from other seed plants, as the reproductive organ (Eames 1961). The flower usually consists of male (anther) and female (ovule) parts within which diploid precursor cells form, with their main function to guarantee the formation of seeds through fertilization of the ovule. Multiple microspore mother cells (male diploid gamete precursor cells) form in the anther, while only a single megaspore mother cell (female gamete precursor cell) forms in the ovule. The diploid microspore mother cell undergoes meiosis and forms four haploid unicellular microspores. The nucleus within the microspore undergoes its first mitosis resulting in a large vegetative cell and a small generative cell each containing a haploid nucleus. Subsequently, the generative cell undergoes the second mitosis and form two haploid sperm cells, which leads to a mature pollen grain containing a vegetative cell and two sperm cells. The diploid megaspore mother cell undergoes meiosis producing four haploid megaspores. Three of the four megaspores degenerate, with only one surviving to become the functional megaspore. The surviving functional megaspore undergoes three mitotic divisions bypassing cytokinesis and forms an embryo sac containing eight nuclei, which is also called *Polygonum*-type embryo sac

that exists in more than 70% angiosperms (Drews *et al.,* 2011). The eight nuclei become compartmentalized to form seven cells with three antipodal cells sitting at the chalazal end, a large central cell containing two polar nuclei in the middle of the embryo sac, two synergid cells and a single egg cell at the micropylar end.

Sexual seed formation initiates following double fertilization where two sperm cells fertilize the egg cell and central cell in the ovule. The process begins with a pollen grain adhering to the stigma of a pistil, followed by germination and pollen tube growth from stigma, down the style and into the micropylar end of the ovary under the guidance of synergids (Higashiyama *et al.,* 2001). One of the sperm cells released from the pollen tube fertilizes the egg cell, producing a diploid zygote that later develops into an embryo. The other sperm cell fuses with central cell nuclei to form a triploid endosperm providing essential nutrition to the developing embryo (Berger *et al.,* 2008). As the embryo and endosperm develop within the embryo sac, the surrounding nucellus and integuments in the ovule contribute to the seed coat, of which altogether comprise a mature seed.

Mechanisms of apomixis

Apomixis in angiosperms is generally composed of apomeiosis (bypass of meiosis during female gametophyte formation), parthenogenesis (development of an embryo from the egg cell without fertilization) and pseudogamous or autonomous (fertilization dependent or independent) endosperm formation (Koltunow *et al.,* 2003). Depending on whether the embryo is formed through a gametophyte or directly from somatic cells in ovule, apomixis can be subdivided into gametophytic apomixis and sporophytic apomixis (adventitious embryony) (Nogler 1984; Koltunow 1993).

Gametophytic apomixis could be further divided into apospory and diplospory depending on the type of precursor diploid cell from which the embryo sac arises: in diplospory, the precursor cell is the megaspore mother cell, which may go through restitutional meiosis and abort or directly go through mitosis to form an embryo sac. Species that has been observed to reproduce through diplospory include *Taraxacum officinale* (dandelion) (Bergman 1950; Richards 1973; Nogler 1984), *Tripascum dactyloides* (Leblanc *et al.,* 1995), and *Erigeron annuus* (Tahara 1915; Bergman 1944). In apospory, the precursor cell are somatic cells nearby the megaspore mother cell called aposporous initials, which directly undergo mitosis and form aposporous embryo sacs. Species that reproduce by apospory include, but are not limited to, *Brachiaria brizantha* (Lutts *et al.,* 1994), *Cenchrus ciliaris* (Fisher *et al.,* 1954; Snyder *et al.,* 1955), *Pennisetum squamulatum* (Dujardin *et al.,* 1984) and *Hieracium* (Koltunow *et al.,* 1998). Both aposporous and diplosporous embryo sacs contain diploid eggs that can develop into embryos without fertilization through parthenogenesis, while the endosperm develops pseudogamously or autonomously (with or without the fertilization of polar nuclei). In adventitious embryony, somatic cells in the ovule differentiate and initiate an embryogenic fate along with the development of a sexual embryo sac from the megaspore mother cell. These embryogenic cells start to divide to form mature embryos only if the sexual embryo sac is fertilized, as the development of apomictic embryos relies on nutrition from endosperm. Thus, sporophytic apomixis often results in polyembryonic seed, which occurs commonly in *Citrus* species (Bacchi 1943).

Since meiosis is normal in the pollen of most of the natural apomicts allowing the possibility for fertilization of the unreduced egg cell, how unreduced eggs successfully initiate parthenogenesis without pollen contribution is also of great importance in understanding the

mechanisms of apomixis. Through an ultrastructural study of the early fertilization events in the apomictic buffelgrass (*Cenchrus ciliaris*) female gametophyte, a cell wall was observed to cover the apomictic egg cell several hours before pollen tubes reach the embryo sac physically preventing fertilization of the egg (Vielle *et al.,* 1995), suggesting a mechanism for the unreduced egg to avoid fertilization.

Genetic basis of apomixis

As previously described, apomixis exists in various forms across a wide range of species but invariably circumvents meiosis (apomeiosis) and enables the fertilization-independent development of an embryo from the egg cell (parthenogenesis), avoiding genetic recombination and fusion with male gamete and keeping its unchanged genetics in the form of a seed. The apomixis community has made significant strides in expanding our knowledge on the genetic control of apomixis in a wide range of species with the help of molecular markers (Ozias-Akins *et al.,* 2007), allowing us to further study the molecular mechanisms underlying apomixis.

Buffelgrass (*Cenchrus ciliaris*) is a highly polymorphic, perennial warm-season grass that was observed to reproduce by apospory and pseudogamy (endosperm development with fertilization) (Fisher *et al.,* 1954; Snyder *et al.,* 1955). The subsequent discovery and cytological identification of a sexual individual (Bashaw 1962) offered the opportunity to study the genetic inheritance of apospory in buffelgrass with mapping populations. Selfing or intercrossing among different sexual tetraploid (2n = 36) buffelgrass accessions resulted in only sexual progenies, while the crossing of sexual females with male obligate apomictic genotypes resulted in segregation for sexuality to apospory varying from 1:1 to 1:3, suggested that apospory in buffelgrass is heterozygous and dominant (Sherwood *et al.,* 1994). In *Pennisetum squamulatum*, an apomictic relative of buffelgrass, a mapping study of a F_1 population segregating for apospory

and sexuality derived from a cross between *Pennisetum glaucum* (pearl millet) and *Pennisetum squamulatum* revealed 12 markers that strictly cosegregated with apospory, defining a contiguous apospory-specific genomic region (ASGR) (Ozias-Akins *et al.,* 1998). The linkage of many of these markers to apospory was further proven in two small segregating populations derived from crosses between sexual and apomictic *Cenchrus ciliaris*, indicating that ASGR is conserved between these two species (Roche *et al.,* 1999). Further characterization of the ASGR in *Pennisetum squamulatum* and *Cenchrus ciliaris* was carried out through Bacterial Artificial Chromosome (BAC) clone sequencing, and genes including AP2-domain containing transcription factor *ASGR-BBM-like* genes were found in ASGR-linked BACs (Gualtieri *et al.,* 2006; Conner *et al.,* 2008). The *PsASGR-BABY BOOM-like* (*PsASGR-BBML*) transgene is able to induce embryo formation in unfertilized egg cells of sexual pearl millet, maize and rice making it the first experimentally verified parthenogenesis gene (Conner *et al.,* 2015; Conner *et al.,* 2017). A rare recombinant plant (A8) that lost a portion of the ASGR was identified from a cross between sexual and apomictic buffelgrass and shown to form aposporous embryo sacs but not express parthenogenesis, suggesting that apospory and parthenogenesis could be separately expressed (Conner *et al.,* 2013).

As Mendel was verifying his laws of inheritance in *Hieracium*, he crossed two assumed-to-be "true breeding" strains during the summer of 1866 and examined the traits of the resulting F_1 hybrids. To his consternation, he observed a very perplexing result where one out of four F_1 hybrid plants was a true hybrid while the remaining three were mother-like plants. Further crossing trials still led to mostly mother-like plants and rarely true hybrid plants, which apparently stood in contradiction to his laws of inheritance (Nogler 2006). Now we know, which, was unbeknownst to him, that he actually made the first crosses between sexual and facultative

apomictic *Hieracium* in the summer of 1866. Extensive mapping efforts have now been made in *Hieracium* to decipher the genetic control of apomixis. Instead of genetic mapping, deletion mapping was used in *Hieracium caespitosum* (2n = 4X), which reproduces by autonomous apospory, to determine chromosome segments conferring apomixis (Catanach *et al.,* 2006). Chromosome deletions were induced in ~ 5000 viable seeds through irradiation, ~ 2400 M_1 plants survived and grew to maturity. Chimeric M_1 plants were morphologically screened into three categories: 1) Loss-of-apomeiosis, 2) Loss-of-parthenogenesis, 3) Loss-of-both, and were made homogeneous through tissue culture. Seventy-nine viable loss-of-apomixis regenerants were used for AFLP mapping. Two genomic regions were identified to align well with the mutant phenotypes, which were named as *Loss of Apomeiosis* (*LOA*) and *Loss of Parthenogenesis* (*LOP*), respectively. Markers linked to *LOA* and *LOP* were tested in a segregating population, and 100% co-segregation was found between the *LOP* marker and parthenogenesis while the *LOA* marker was estimated to be ~5 cM away from apospory. In all, two major genomic regions *LOA* and *LOP* have been identified in *Hieracium caespitosum* that may genetically control apomeiosis and parthenogenesis, respectively.

Moreover, an AFLP study on obligate sexual and obligate apomictic genotypes of *Hypericum perforatum* identified fragments that only exist in apomicts. Further association study of a CAPS (cleaved amplified polymorphic sequence) marker derived from the AFLP fragment revealed that it co-segregated with apospory but not parthenogenesis in a segregating population obtained from a cross between diploid sexuals (2n = 16) and tetraploid apomicts (2n = 32), indicating that two independent loci are responsible for these two components of apomixis (Schallau *et al.,* 2010). In another species, a cytohistological investigation was carried out on a segregating population of *Poa pratensis* created from a cross between a sexual and an apomictic

parent, where apospory was found in two non-parthenogenetic individuals, indicating that apospory and parthenogenesis may be genetically uncoupled in *Poa pratensis* (Albertini *et al.,* 2001).

In the case of diplospory, triploid non-apomictic dandelions hybrids were produced in crosses between diploid sexual as the mother and triploid apomicts as the pollen donor. The progenies from a following cross between the non-apomictic triploid hybrids as the maternal parent and diploid sexual as the paternal parent showed that one type of triploid hybrids produced exclusively 2n + n hybrids progenies. This result indicated that the non-apomictic triploid hybrids were diplosporous but lack parthenogenesis, suggestive of the independent nature of the loci controlling diplospory and parthenogenesis (Van Dijk *et al.,* 1999). Genetics studies on a segregating population produced in a cross between triploid diplosporous *Erigeron annuus* and sexual diploid *Erigeron strigosus* showed that four closely-linked AFLP markers were significantly associated with parthenogenesis and eleven co-segregating markers were completely associated with diplospory, indicating that diplospory and parthenogenesis are also independently inherited in the case of *Erigeron annuus* (Noyes *et al.,* 2000).

Over the past two decades, multiple independent mapping studies in different apomictic species and their sexual relatives have more or less revealed the inheritance of components of apomixis (Ozias-Akins *et al.,* 2007). However, up to now only the *PsASGR-BBML* gene located in *Pennisetum squamulatum* ASGR (Ozias-Akins *et al.,* 1998) and identified through ASGR-linked bacterial artificial clone sequencing (Gualtieri *et al.,* 2006; Conner *et al.,* 2008) is an experimentally verified parthenogenesis gene (Conner *et al.,* 2015; Conner *et al.,* 2017). This is an example of successful practice of map-based cloning method to identify an apomixis gene and may encourage a similar approach to be applied in other species to identify other apomixis-

controlling genes. However, such an approach is greatly hindered by low recombination in many apomixis-linked genomic regions.

Apomeiosis vs Meiosis

Apomeiosis, a major component of the gametophytic apomixis (apospory and diplospory), essentially results in the production of unreduced gametes from a somatic cell in the ovule maintaining the 2n chromosomal number. While this is not unique to natural apomicts, sexual species can also produce unreduced gametes via various mechanisms mainly but not limited to meiotic defects (Brownfield *et al.,* 2010).

Meiosis is an essential process for sexual reproduction with the end-products being daughter cells with half of the chromosomes from its mother cell (Mercier *et al.,* 2015). During meiosis I prophase, homologous chromosomes pair and crossover between non-sister chromatids occurs, and this is a stage where natural genetic variations are generated. During meiosis I anaphase, homologous chromosomes move towards two opposite poles of the cells and eventually separate into two haploid cells in telophase. During meiosis II, which resembles mitosis, sister chromatids of each chromosome separate into two cells retaining the haploid chromosomal number. Two most prevalent mechanisms of meiotic defects resulting in unreduced gametes production are first division restitution (FDR) and second division restitution (SDR) where FDR retains all homologous chromosomes during meiosis I and SDR keeps all sister chromatid during meiosis II. For FDR-derived unreduced gametes, the majority of chromosome constitution would be unchanged except that half of the unreduced gametes may lose heterozygosity in the cross-over region, and this may resemble what was formed during apomeiosis. For SDR-derived unreduced gametes, normal meiosis I but omission of sister chromatid separation in meiosis II permits chromosomal recombination and independent

assortment of homologous chromosomes, giving rise to unreduced, however, recombined gametes (Bretagnolle *et al.,* 1995).

In maize, the unreduced female gametes production was observed in maize meiotic mutant *elongate1* due to the absence of female meiosis II, but no corresponding gene has been identified yet (Barrell *et al.,* 2005). In the *dyad* mutant in *Arabidopsis thaliana*, the megaspore mother cell either meiotically divides once producing a dyad of two cells or bypasses meiosis giving rise to unreduced female gamete (Siddiqi *et al.,* 2000; Agashe *et al.,* 2002). The homozygous *dyad* mutants can sexually produce triploid progenies with full retention of parental heterozygosity resembling what was found in apomeiosis (Ravi *et al.,* 2008). However, with only 0.2% of the *dyad* ovules producing viable gametes, it is not agronomically practical to employ *dyad* into apomeiosis engineering. Still in *Arabidopsis*, the observation that *osd1* (*omission of second division*) mutants fail to undergo the meiosis II, resulting in the production of ~ 85% female unreduced gametes, together with the fact that the *recombination8* (*rec8*) and *sporulation11* (*spo11*) double mutant abolished recombination, has led to the creation of a triple (*osd1*, *rec8* and *spo11*) *Mitosis instead of Meiosis* (*MiMe*) mutant (d'Erfurth *et al.,* 2009). This triple *Mitosis instead of Meiosis* (*MiMe*) mutant produces nearly 100% unreduced and unrecombined female gametophytes that are genetically identical to the megaspore mother cell, resembling the apomeiosis, thus can be a competent alternative for engineering apomixis.

Apomeiosis now could be replaced by *MiMe* for the production of synthetic apomicts, however, the genetic basis of apomeiosis is still unclear as no apomeiosis gene has been identified yet from natural apomicts. The disadvantage of *MiMe* is that it requires three recessive mutations to function, whereas apomeiosis in natural apomixis may only require a single

dominant gene. However, simultaneously editing multi-genes may not be a problem anymore with CRISPR/Cas9 technology.

Parthenogenesis vs Zygotic embryogenesis

Parthenogenesis, a major component of apomixis, is described as the development of an embryo from the egg cell without fertilization. Autonomous embryo formation also occurs in many sexual species, eventually giving rise to progeny having half the chromosome number of its mother cell from meiotically reduced gametes, which could be valuable for doubled haploids (DHs) production in plant breeding.

Obtaining haploid plants could be achieved in many ways such as wide hybridization or tissue culture from anthers, microspores, ovaries and ovules (Forster *et al.,* 2007). In major crops, high frequency haploid embryo production was first achieved in barley through interspecific crosses between cultivated barley (*Hordeum vulgare*) and *Hordeum bulbosum* followed by embryo culture (Kasha *et al.,* 1970). This has now become a routine practice to produce haploids in barley (Devaux 2003). Haploid wheat embryos could also be generated in a cross between the wheat variety Chinese Spring and the maize variety Seneca 60, where the zygote was hybrid but the paternal chromosomes were eliminated through subsequent zygotic divisions (Laurie *et al.,* 1990). The common sexual wheat (*Triticum aestivum*) line 'Salmon' (2n = 42), whose long arm of chromosome 1B is substituted by the short arm of the rye chromosome 1R, can be parthenogenetic and produce haploids when its nucleus is transferred into *Aegilops cuudata* cytoplasma (Tsunewaki *et al.,* 1990; Hsam *et al.,* 1993; Matzk 1996). This method of generating haploids has been applied and named as 'Salmon method' in haploid wheat production.

Haploid embryo induction is valuable in plant breeding for its use in the production of doubled haploids (DHs) by artificially doubling its chromosome number, reaching complete homozygosity in just one generation (Forster *et al.,* 2005). However, the molecular mechanisms of haploid embryo production or haploid induction is still not fully revealed.

In maize, the discovery of the haploid inducing line, Stock6, which produces maternal haploid progenies through self-pollination at relatively high frequencies (~3.23%) affords the opportunity to study genetic inheritance of haploid induction (Coe Jr 1959). Two Stock6-derived inducer lines, CAU^{B} containing cytogenetic marker B chromosome and CAU^{YFP} containing centromeric histone H3 variant-yellow fluorescent protein were used to pollinate a commercial hybrid ZhengDan958 to trace the molecular markers throughout haploid induction process. The following cytogenetic analysis revealed that most of the inducer chromosomes were eliminated at the start of haploid embryogenesis (Zhao *et al.,* 2013). The *MATRILINEAL* (*MTL*) gene, a pollen-specific phospholipase only expressed in the cytoplasm of the sperm cell, was discovered to control haploid induction in the Stock6 where genome editing in this gene can lead to a 6.7% haploid induction rate (Kelliher *et al.,* 2017). These findings suggest that the potential mechanism of Stock6 maternal haploid induction in maize could be the paternal chromosome elimination following fertilization.

Apart from maize, transgenic studies in *Arabidopsis* have also deepened our understanding of molecular mechanisms of haploid induction. In *Arabisopsis thaliana*, haploid plants can be easily generated by crossing a wild-type *Arabidopsis* with *cenh3* null mutant expressing *GFP-tailswap* CENH3. The *GFP-tailswap CENH3* transgene was constructed by fusing GFP reporter with CENH3 that has its hypervariable N-terminal tail sequences replaced by the tail of a conventional histone H3 (Ravi *et al.,* 2010). The resulting haploids exclusively

retained the chromosomes from the wild-type parent regardless of who served as the pollen donor. Further crosses of GEM (Genome Elimination) line co-expressing GFP–CENH3 and GFP-tailswap variants in the *cenh3-1* mutant with *MiMe* or *dyad* have led to synthetic viable clonal seeds mimicking what was produced in natural apomicts, although full penetrance was not achieved (Marimuthu *et al.,* 2011). This method theoretically could be applied to any major genetically modifiable crops as the CENH3 is universally present in all plants.

Even though the genome elimination-based autonomous embryo formation can induce haploid embryo production by nature or by transgenetics, it is by definition different than those achieved in natural apomicts through parthenogenesis, as parthenogenetic embryos are achieved without contribution from a paternal genome. Maternal haploids were first achieved from a sexual tetraploid pearl millet through transformation of the gene *PsASGR-BABYBOOM-like* (*PsASGR-BBML*), which was isolated from a natural apomictic grass (*Pennisetum squamulatum*) (Gualtieri *et al.,* 2006; Conner *et al.,* 2008; Conner *et al.,* 2015). The transgenic expression of *PsASGR-BBML* was also shown to induce parthenogenesis in maize and rice (Conner *et al.,* 2017), meaning a single gene of natural apomict origin is sufficient to induce parthenogenesis in the sexual monocot relatives. However, the molecular mechanisms of natural parthenogenesis are still not fully understood with respect to its relationship to sexual zygotic embryogenesis, which is also the embryo development of the egg cell but after fertilization. Is *PsASGR-BBML* a factor in natural apomicts that initiates a unique program leading to embryogenesis, or is it a factor that can substitute for fertilization to initiate the same program as zygotic embryogenesis?

The first *BBM* gene was discovered through subtractive hybridization on upregulated genes in the microspore-derived embryo in *Brassica napus*, and the ectopic expression of *BBM*

in both *Arabidopsis* and *Brassica* induced somatic embryogenesis (Boutilier *et al.,* 2002). The natural *BBM1* expression in rice was detected in zygotes at 2.5 h after pollination (HAP) (corresponding to karyogamy) in a male-origin-specific manner, and its ectopic expression in egg cells under an egg-cell-specific promoter induced parthenogenesis bypassing fertilization (Anderson *et al.,* 2017; Khanday *et al.,* 2019). This evidence supports that not only the *BBM* of natural apomict origin but also a sexually-derived *BBM* gene can induce parthenogenesis when expressed in the egg before fertilization. This leads to the hypothesis that parthenogenesis in apomicts might be largely similar to sexual zygotic embryogenesis but differs by a heterochronic shift toward expression of embryogenesis-initiating genes prior to fertilization.

To test this hypothesis, a gene or pathway commonly employed by the natural parthenogenesis in apomicts and zygotic embryogenesis in their sexual relatives needs to be discovered. If natural parthenogenesis and zygotic embryogenesis do share common pathways, then what are the interacting factors or microenvironments that cause these commonly employed genes or pathways to express differently temporally or spatially?

RNA sequencing

Innovative technologies have always been a huge force driving scientific discovery, and RNA sequencing is such a technology that has basically revolutionized the way biology research is done. One of the earliest versions of RNA-seq-like research was done in maize for more than a decade ago (Emrich *et al.,* 2007), and now RNA-seq has become one of the most routine tools used in modern molecular-level plant research (Martin *et al.,* 2013).

A standard workflow in RNA-seq studies starts from the wet-lab work with major steps including the collection of samples of different genotypes or conditions, RNA extraction, poly-A selected mRNA enrichment or ribosomal RNA depletion, cDNA synthesis, DNA fragmentation

and adaptor-ligation-based sequencing library preparation. Libraries can be sequenced to a depth of 4 million to 20 billion 150-300 bp long paired-end reads per run using Illumina platforms (https://www.illumina.com/systems/sequencing-platforms.html). The following steps would be computational including data preprocessing, which depends on the quality of raw data, genome-guided assembly or de novo, aligning and quantifying reads from each sample to the total assembly, case-to-case normalization, and statistical analyses of significantly differentially expressed genes or transcripts between different genotypes or conditions (Wang *et al.,* 2009). While this is the mostly applied short-read RNA-seq workflow for differential gene expression study, RNA-seq workflow has also been extensively revised and optimized to adapt to newly-emerging sequencing methods like long-read sequencing and direct RNA-sequencing (Sharon *et al.,* 2013; Cartolano *et al.,* 2016; Byrne *et al.,* 2017; Garalde *et al.,* 2018) to provide novel solutions.

Moreover, single-cell RNA-seq and spatial transcriptomics would provide unprecedentedly high-resolution in answering biological questions that are intrinsically not answerable through bulk RNA-seq or other traditional molecular biology methods. The inherent limitations of RNA-seq from bulk tissues is the incapability to reflect or represent transcripts derived from the biologically relevant cells or tissues, by bulking all cell types from the tissue and subsequent loss of resolved spatial information. Both specific cell types and resolved spatial information are extremely critical to truly understand a complex biological trait like apomixis, of which the expression of its components is essentially restricted to few cell types. Bulking the ovary or ovule transcripts would definitely dilute what is truly relevant to apomixis.

The first single cell RNA-seq (scRNA-seq) work was published in 2009, where individual mouse oocytes were manually isolated under a microscope, processed for RNA

extraction, cDNA synthesis, library prep with a modified protocol, leading to the identification of previously unknown splice junctions and novel isoforms (Tang *et al.,* 2009). Mechanical disaggregation and enzymatic dissociation of tissues of interests into single cell-suspension is generally the first step in single cell RNA-seq. Once single cell suspension is achieved, several methods can be used to separate individual cells depending on the requirement of the throughput and instrument accessibility. Several single cell separation techniques for animal cells have been quite successful such as flow sorting (Ramsköld *et al.,* 2012), microfluids Fluidigm C1 system (Brennecke *et al.,* 2013), nanowell-based ICELL8 microchip system (Goldstein *et al.,* 2017), Drop-seq (Macosko *et al.,* 2015), InDrop (Klein *et al.,* 2015) and in situ sequence-based barcode sorting (Cao *et al.,* 2017; Rosenberg *et al.,* 2018). After single cells are isolated, RNA extraction, cDNA synthesis, library prep are similar to bulk RNA-seq workflow but require case-to-case modifications. Very recently, single cell RNA-seq was firstly applied in plants where more than 10,000 single-cell protoplast transcriptomes from *Arabidopsis* root cells were isolated and sequenced, from which not only all sorts of expected and rare cell types were identified but also questions regarding the developmental dynamics of cell differentiation could be studied (Ryu *et al.,* 2019).

However, neither bulk nor single cell RNA-seq could resolve the spatial information when cellular and spatial context is critical to the traits of interests. One of the alternatives is laser-capture microdissection (LCM) RNA-seq, which has been used to isolate spatially-restricted cells or tissues from cryosections or paraffin sections (Kerk *et al.,* 2003; Nakazono *et al.,* 2003; Zhan *et al.,* 2015; Chen *et al.,* 2017). The limitations for LCM-based RNA-seq are the requirement of a laser capture microdissection instrument, and a laborious and tedious process to scale as well as a destructive nature that cause RNA to degrade. Besides, cryosections may not

work well with highly vacuolated cells or tissues as ice crystals may form and negatively affect cell or tissue structure. Paraffin sections preserve the spatial information very well, however, the damage imposed to the tissues by fixation and embedding affect the RNA quality, and thus increases the difficulty of all downstream experimental procedures including sequencing library preparation and computational analyses of the resulting sequencing data.

Another alternative could be directly capturing mRNA spatial information from tissue sections. This can be done on a specialized slide with oligo-dT primer, UMI (unique molecule identifier), and spatial barcode (spatial coordinates) on its surface, to which mRNA from tissue sections could hybridize and reverse-transcribe to cDNA. All cDNAs reverse-transcribed on the slides are pooled for library preparation and sequencing. Each sequencing read produced in this way contains not only the mRNA sequence information, but also the corresponding spatial coordinates. Subsequent computational analyses enable visualization of transcripts, abundance and localization of expression (Ståhl *et al.,* 2016; Rodriques *et al.,* 2019), which could be very useful at a scalable level for obtaining spatially resolved gene expression for medical research or plant disease diagnose.

Goals

The goal of this study was to identify genes or pathways expressed in unpollinated sexual and apomictic egg cells at the stage of parthenogenesis initiation to understand transcriptomic control of parthenogenesis in the context of natural apomixis. To accomplish this goal we first needed to overcome technical challenges in capturing the egg cell that is deeply embedded in the ovaries without contamination from other tissues or severe physical damage, protocols to isolate RNA of sufficient quality and quantity, conversion to cDNA, and preparation of sequencing library. We also generated a reference ovule transcriptome assembly for *Cenchrus ciliaris* to

properly analyze the egg cell RNA-seq data generated in this study, since there currently is no assembled and annotated *Cenchrus ciliaris* genome available. Finally, we carried out computational analysis and in situ hybridization of a subset of differentially expressed genes to test the validity of the RNA-seq data generated.

References

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

Albertini, E., A. Porceddu, F. Ferranti, L. Reale, G. Barcaccia, B. Romano and M. Falcinelli (2001). "Apospory and parthenogenesis may be uncoupled in *Poa pratensis*: a cytological investigation." Sexual Plant Reproduction 14(4): 213-217.

Anderson, S. N., C. S. Johnson, J. Chesnut, D. S. Jones, I. Khanday, M. Woodhouse, C. Li, L. J. Conrad, S. D. Russell and V. Sundaresan (2017). "The zygotic transition is initiated in unicellular plant zygotes with asymmetric activation of parental genomes." Developmental cell 43(3): 349-358.

Bacchi, O. (1943). "Cytological observations in *Citrus*: III. Megasporogenesis, fertilization, and polyembryony." Botanical Gazette 105(2): 221-225.

Barrell, P. J. and U. Grossniklaus (2005). "Confocal microscopy of whole ovules for analysis of reproductive development: the elongate1 mutant affects meiosis II." The Plant Journal 43(2): 309-320.

Bashaw, E. C. (1962). "Apomixis and Sexuality in Buffelgrass 1." Crop Science 2(5): 412-415.

Berger, F., Y. Hamamura, M. Ingouff and T. Higashiyama (2008). "Double fertilization–caught in the act." Trends in plant science 13(8): 437-443.

Bergman, B. (1944). "A contribution to the knowledge of the embryo sac mother cell and its development in two apomicts." Svensk Bot. Tidskr 38: 249-259.

Bergman, B. (1950). "Meiosis in two different clones of the apomictic *Chondrilla juncea*." Hereditas 36(3): 297-320.

Bicknell, R. A. and A. M. Koltunow (2004). "Understanding apomixis: recent advances and remaining conundrums." The Plant Cell 16(suppl 1): S228-S245.

Boutilier, K., R. Offringa, V. K. Sharma, H. Kieft, T. Ouellet, L. Zhang, J. Hattori, C.-M. Liu, A. A. M. van Lammeren and B. L. A. Miki (2002). "Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth." The Plant Cell 14(8): 1737-1749.

Brennecke, P., S. Anders, J. K. Kim, A. A. Kołodziejczyk, X. Zhang, V. Proserpio, B. Baying, V. Benes, S. A. Teichmann and J. C. Marioni (2013). "Accounting for technical noise in single-cell RNA-seq experiments." Nature methods 10(11): 1093.

Bretagnolle, F. a. and J. D. Thompson (1995). "Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants." New Phytologist 129(1): 1-22.

Brownfield, L. and C. Köhler (2010). "Unreduced gamete formation in plants: mechanisms and prospects." Journal of experimental botany 62(5): 1659-1668.

Byrne, A., A. E. Beaudin, H. E. Olsen, M. Jain, C. Cole, T. Palmer, R. M. DuBois, E. C. Forsberg, M. Akeson and C. Vollmers (2017). "Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells." Nature communications 8: 16027.

Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan and F. J. Steemers (2017). "Comprehensive single cell transcriptional profiling of a multicellular organism by combinatorial indexing." BioRxiv: 104844.

Cartolano, M., B. Huettel, B. Hartwig, R. Reinhardt and K. Schneeberger (2016). "cDNA library enrichment of full length transcripts for SMRT long read sequencing." PloS one 11(6): e0157779.

Catanach, A. S., S. K. Erasmuson, E. Podivinsky, B. R. Jordan and R. Bicknell (2006). "Deletion mapping of genetic regions associated with apomixis in *Hieracium*." Proceedings of the National Academy of Sciences 103(49): 18650-18655.

Chen, J., S. Suo, P. P. L. Tam, J.-D. J. Han, G. Peng and N. Jing (2017). "Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq." Nature protocols 12(3): 566.

Christenhusz, M. J. M. and J. W. Byng (2016). "The number of known plants species in the world and its annual increase." Phytotaxa 261(3): 201-217.

Coe Jr, E. H. (1959). "A line of maize with high haploid frequency." The American Naturalist 93(873): 381-382.

Conner, J. A., S. Goel, G. Gunawan, M.-M. Cordonnier-Pratt, V. E. Johnson, C. Liang, H. Wang, L. H. Pratt, J. E. Mullet and J. DeBarry (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., G. Gunawan and P. Ozias-Akins (2013). "Recombination within the apospory specific genomic region leads to the uncoupling of apomixis components in *Cenchrus ciliaris*." Planta 238(1): 51-63.

Conner, J. A., M. Mookkan, H. Huo, K. Chae and P. Ozias-Akins (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.

Conner, J. A., M. Podio and P. Ozias-Akins (2017). "Haploid embryo production in rice and maize induced by PsASGR-BBML transgenes." Plant reproduction 30(1): 41-52.

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

De Wet, J. M. J., J. R. Harlan, L. M. Engle and C. A. Grant (1973). "Breeding Behavior of Maize-*Tripsacum* Hybrids 1." Crop Science 13(2): 254-256.

Devaux, P. (2003). The *Hordeum bulbosum* (L.) method. Doubled Haploid Production in Crop Plants, Springer: 15-19.

Drews, G. N. and A. M. G. Koltunow (2011). "The female gametophyte." The Arabidopsis book/American Society of Plant Biologists 9.

Dujardin, M. and W. W. Hanna (1983). "Apomictic and sexual pearl millet x *Pennisetum squamulatum* hybrids." Journal of Heredity 74(4): 277-279.

Dujardin, M. and W. Hanna (1984). "Microsporogenesis, reproductive behavior, and fertility in five *Pennisetum* species." Theoretical and Applied Genetics 67(2-3): 197-201.

Dujardin, M. and W. Hanna (1985). "Cytology and Reproduction of Reciprocal Backcrosses Between Pearl Millet and Sexual and Apomictic Hybrids of Pearl Millet× *Pennisetum squamulatum* 1." Crop Science 25(1): 59-62.

Eames, A. J. (1961). "Morphology of the angiosperms." Morphology of the angiosperms.

Emrich, S. J., W. B. Barbazuk, L. Li and P. S. Schnable (2007). "Gene discovery and annotation using LCM-454 transcriptome sequencing." Genome research 17(1): 69-73.

Fisher, W. D., E. C. Bashaw, E. C. Holt, L. M. Soffer and E. Carpenter (1954). "2456201. Evidence for apomixis in *Pennisetum ciliare* and *Cenchrus setigerus*." Agronomy journal 46(9): 401-404.

Forster, B. P., E. Heberle-Bors, K. J. Kasha and A. Touraev (2007). "The resurgence of haploids in higher plants." Trends in plant science 12(8): 368-375.

Forster, B. P. and W. T. B. Thomas (2005). "Doubled haploids in genetics and plant breeding." Plant Breed Rev 25: 57-88.

Garalde, D. R., E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James and A. Warland (2018). "Highly parallel direct RNA sequencing on an array of nanopores." Nature methods 15(3): 201.

Goel, S., Z. Chen, Y. Akiyama, J. A. Conner, M. Basu, G. Gualtieri, W. W. Hanna and P. Ozias-Akins (2006). "Comparative physical mapping of the apospory-specific genomic region in two apomictic grasses: *Pennisetum squamulatum* and *Cenchrus ciliaris*." Genetics 173(1): 389-400.

Goel, S., Z. Chen, J. A. Conner, Y. Akiyama, W. W. Hanna and P. Ozias-Akins (2003). "Physical evidence that a single hemizygous chromosomal region is sufficient to confer aposporous embryo sac formation in *Pennisetum squamulatum* and *Cenchrus ciliaris*." Genetics 163: 1069-1082.

Goldstein, L. D., Y.-J. J. Chen, J. Dunne, A. Mir, H. Hubschle, J. Guillory, W. Yuan, J. Zhang, J. Stinson and B. Jaiswal (2017). "Massively parallel nanowell-based single-cell gene expression profiling." BMC genomics 18(1): 519.

Gualtieri, G., J. A. Conner, D. T. Morishige, L. D. Moore, J. E. Mullet and P. Ozias-Akins (2006). "A segment of the apospory-specific genomic region is highly microsyntenic not only between the apomicts *Pennisetum squamulatum* and buffelgrass, but also with a rice chromosome 11 centromeric-proximal genomic region." Plant physiology 140(3): 963-971.

Higashiyama, T., S. Yabe, N. Sasaki, Y. Nishimura, S.-y. Miyagishima, H. Kuroiwa and T. Kuroiwa (2001). "Pollen tube attraction by the synergid cell." Science 293(5534): 1480-1483.

Hojsgaard, D., S. Klatt, R. Baier, J. G. Carman and E. Hörandl (2014). "Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics." Critical Reviews in Plant Sciences 33(5): 414-427.

Hsam, S. L. K. and F. J. Zeller (1993). "Haploid production in durum wheat by the interaction of *Aegilops kotschyi* cytoplasm and 1BL/1RS chromosomal interchange." Theoretical and Applied Genetics 86(8): 951-954.

Kasha, K. J. and K. N. Kao (1970). "High frequency haploid production in barley (*Hordeum vulgare* L.)." Nature 225(5235): 874.

Kelliher, T., D. Starr, L. Richbourg, S. Chintamanani, B. Delzer, M. L. Nuccio, J. Green, Z. Chen, J. McCuiston and W. Wang (2017). "MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction." Nature 542(7639): 105.

Kerk, N. M., T. Ceserani, S. L. Tausta, I. M. Sussex and T. M. Nelson (2003). "Laser capture microdissection of cells from plant tissues." Plant physiology 132(1): 27-35.

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

Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner (2015). "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." Cell 161(5): 1187-1201.

Koltunow, A. M. (1993). "Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules." The Plant Cell 5(10): 1425.

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

Koltunow, A. M., S. D. Johnson and R. A. Bicknell (1998). "Sexual and apomictic development in *Hieracium*." Sexual Plant Reproduction 11(4): 213-230.

Laurie, D. A., L. S. O'Donoughue and M. D. Bennett (1990). Wheat x maize and other wide sexual hybrids: their potential for genetic manipulation and crop improvement. Gene Manipulation in Plant Improvement II, Springer: 95-126.

Leblanc, O., M. D. Peel, J. G. Carman and Y. Savidan (1995). "Megasporogenesis and megagametogenesis in several *Tripsacum* species (Poaceae)." American Journal of Botany 82(1): 57-63.

Lutts, S., J. Ndikumana and B. P. Louant (1994). "Male and female sporogenesis and gametogenesis in apomictic *Brachiaria brizantha*, *Brachiaria decumbens* and F 1 hybrids with sexual colchicine induced tetraploid *Brachiaria ruziziensis*." Euphytica 78(1-2): 19-25.

Macosko, E. Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki and E. M. Martersteck (2015). "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets." Cell 161(5): 1202-1214.

Marimuthu, M. P. A., S. Jolivet, M. Ravi, L. Pereira, J. N. Davda, L. Cromer, L. Wang, F. Nogué, S. W. L. Chan and I. Siddiqi (2011). "Synthetic clonal reproduction through seeds." Science 331(6019): 876-876.

Martin, L., Z. Fei, J. Giovannoni and J. K. C. Rose (2013). "Catalyzing plant science research with RNA-seq." Frontiers in plant science 4: 66.

Matzk, F. (1996). "The 'Salmon System'of wheat—a suitable model for apomixis research." Hereditas 125(2-3): 299-304.

Mercier, R., C. Mézard, E. Jenczewski, N. Macaisne and M. Grelon (2015). "The molecular biology of meiosis in plants." Annual review of plant biology 66: 297-327.

Moll, R. H., J. H. Lonnquist, J. V. Fortuno and E. C. Johnson (1965). "The relationship of heterosis and genetic divergence in maize." Genetics 52(1): 139.

Morgan, R. N., P. Ozias-Akins and W. W. Hanna (1998). "Seed set in an apomictic BC3 pearl millet." International journal of plant sciences 159(1): 89-97.

Nakazono, M., F. Qiu, L. A. Borsuk and P. S. Schnable (2003). "Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize." The Plant Cell 15(3): 583-596.

Nogler, G. A. (1984). Gametophytic apomixis. Embryology of angiosperms, Springer: 475-518.

Nogler, G. A. (2006). "The lesser-known Mendel: his experiments on *Hieracium*." Genetics 172(1): 1-6.

Noyes, R. D. and L. H. Rieseberg (2000). "Two independent loci control agamospermy (apomixis) in the triploid flowering plant *Erigeron annuus*." Genetics 155(1): 379-390.

Ouendeba, B., G. Ejeta, W. E. Nyquist, W. W. Hanna and A. Kumar (1993). "Heterosis and combining ability among African pearl millet landraces." Crop Science 33(4): 735-739.

Ozias-Akins, P., D. Roche and W. W. Hanna (1998). "Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes." Proceedings of the National Academy of Sciences 95(9): 5127-5132.

Ozias-Akins, P. and P. J. van Dijk (2007). "Mendelian genetics of apomixis in plants." Annu Rev Genet 41: 509-537.

Ramsköld, D., S. Luo, Y.-C. Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtukova, J. F. Loring and L. C. Laurent (2012). "Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells." Nature biotechnology 30(8): 777.

Ravi, M. and S. W. L. Chan (2010). "Haploid plants produced by centromere-mediated genome elimination." Nature 464(7288): 615.

Ravi, M., M. P. A. Marimuthu and I. Siddiqi (2008). "Gamete formation without meiosis in *Arabidopsis*." Nature 451(7182): 1121.

Richards, A. J. (1973). "The origin of *Taraxacum* agamospecies." Botanical Journal of the Linnean Society 66(3): 189-211.

Roche, D., P. Cong, Z. Chen, W. W. Hanna, D. L. Gustine, R. T. Sherwood and P. Ozias-Akins (1999). "An apospory-specific genomic region is conserved between Buffelgrass (*Cenchrus ciliaris* L.) and *Pennisetum squamulatum* Fresen." The Plant Journal 19(2): 203-208.

Rodriques, S. G., R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen and E. Z. Macosko (2019). "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution." Science 363(6434): 1463-1467.

Rosenberg, A. B., C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Graybuck, D. J. Peeler, S. Mukherjee and W. Chen (2018). "Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding." Science 360(6385): 176-182.

Ryu, K. H., L. Huang, H. M. Kang and J. Schiefelbein (2019). "Single-cell RNA sequencing resolves molecular relationships among individual plant cells." Plant physiology 179(4): 1444-1456.

Savidan, Y. (2000). "Apomixis, the way of cloning seeds." Biofutur (France).

Savidan, Y. (2001). "Transfer of apomixis through wide crosses." Flowering of Apomixis: From Mechanisms to Genetic Engineering, Y. Savidan, JG Carman, and T. Dresselhaus, eds (Mexico: CIMMYT, IRD, European Commission DG VI): 153-167.

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

Sharon, D., H. Tilgner, F. Grubert and M. Snyder (2013). "A single-molecule long-read survey of the human transcriptome." Nature biotechnology 31(11): 1009.

Sherwood, R. T., C. C. Berg and B. A. Young (1994). "Inheritance of apospory in buffelgrass." Crop Science 34(6): 1490-1494.

Siddiqi, I., G. Ganesh, U. Grossniklaus and V. Subbiah (2000). "The dyad gene is required for progression through female meiosis in *Arabidopsis*." Development 127(1): 197-207.

Smith, J. (1841). "XXXII. Notice of a Plant which produces perfect Seeds without any apparent Action of Pollen." Transactions of the Linnean Society of London(4): 509-512.

Snyder, L. A., A. R. Hernandez and H. E. Warmke (1955). "The mechanism of apomixis in *Pennisetum ciliare*." Botanical Gazette 116(3): 209-221.

Sokolov, V. A., B. Kindiger and I. V. Khatypova (1998). "Apomictically reproducing 39-chromosome maize-*Tripsacum* hybrids." Генетика 34(4): 499-506.

Ståhl, P. L., F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm and M. Huss (2016). "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics." Science 353(6294): 78-82.

Tahara, M. (1915). "Parthenogenesis in *Erigeron annuus*." Bot Mag Tokyo 29: 245-254.

Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch and A. Siddiqui (2009). "mRNA-Seq whole-transcriptome analysis of a single cell." Nature methods 6(5): 377.

Tsunewaki, K. and Y. Mukai (1990). Wheat haploids through the Salmon method. Wheat, Springer: 460-478.

Van Dijk, P. J., I. C. Q. Tas, M. Falque and T. Bakx-Schotman (1999). "Crosses between sexual and apomictic dandelions (*Taraxacum*). II. The breakdown of apomixis." Heredity 83(6): 715.

Vielle, J. P., B. L. Burson, E. C. Bashaw and M. A. Hussey (1995). "Early fertilization events in the sexual and aposporous egg apparatus of *Pennisetum ciliare* (L.) Link." The Plant Journal 8(2): 309-316.

Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nature Reviews Genetics 10(1): 57.

Zhan, J., D. Thakare, C. Ma, A. Lloyd, N. M. Nixon, A. M. Arakaki, W. J. Burnett, K. O. Logan, D. Wang and X. Wang (2015). "RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation." The Plant Cell 27(3): 513-531.

Zhao, X., X. Xu, H. Xie, S. Chen and W. Jin (2013). "Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers." Plant physiology 163(2): 721-731.

CHAPTER 2

A REFERENCE OVULE TRANSCRIPTOME FOR UNPOLLINATED SEXUAL AND APOMICTIC *CENCHRUS CILIARIS* ON THE DAY OF ANTHESIS[1]

[1] Y.Ke, J.Conner and P. Ozias-Akins. To be submitted to *BMC genomics*

Abstract

Apomixis intrinsically circumvents meiosis (apomeiosis) and fusion of the egg cell with a male gamete to develop embryo (parthenogenesis), preventing maternal genetics from changing but maintaining the ability to disperse in the form of a seed. This phenomenon of asexual seed formation is seen in a wide range of plant families but not in major crops. The engineering of such self-producing ability into major crops would revolutionize agriculture and the hybrid-based seed industry, if successfully synthesized in major crops, would be economically more efficient than current ways to produce hybrid seeds. Buffelgrass (*Cenchrus ciliaris*) can reproduce sexually or by obligate apomixis, and is a model species for studying genetic inheritance and molecular mechanisms of apomixis. However, no sequenced and annotated *Cenchrus ciliaris* genome is currently available, limiting the analyses of previously identified apomixis-associated gDNA markers or cDNA fragments identified through genetic mapping or differential display as well as restricting the ability to properly assemble or functionally annotate single cell RNA-seq data. By conducting RNA-seq on high-quality, abundant RNA collected from the sexual and apomictic ovules on the day of anthesis, we created a de novo transcriptome with near-complete sequence information for the *Cenchrus ciliaris* ovules, and identified putative pathways that distinguish apomictic ovules from their sexual counterpart at the stage where parthenogenesis is being expressed. Our transcriptome data served as a reference unpollinated ovule transcriptome for *Cenchrus ciliaris* on the day of anthesis, and is a valuable resource for further research into female reproduction and buffelgrass breeding.

Introduction

Apomixis is commonly described as the asexual reproduction through seed, which means the embryo within the seed actually arises from the maternal cells in ovule without genetic contribution from pollen (Nogler 1984; Bicknell *et al.,* 2004; Ozias-Akins *et al.,* 2007). Apomixis was first reported in *Alchornea ilicifolia* as a reproductive abnormality where the seeds were produced without any proven action of pollination (Smith 1841). Based on more than a century of germplasm collection and cytogenetic analyses, apomixis is considered as a conserved reproductive strategy widely employed in angiosperms instead of a rare phenomenon (Carman 1997; Hojsgaard *et al.,* 2014). However, it is still rare in the gene pools of major crops. Introgression-based breeding attempts have been made to transfer apomixis to major crops such as maize and pearl millet from their apomictic wild relatives *Tripsacum dactyloides* and *Pennisteum squamulatum*, respectively (Morgan *et al.,* 1998; Savidan 2001). However, it has proved difficult to recover agronomically acceptable lines as the resulting apomictic hybrids have a high degree of seed abortion. Understanding the molecular mechanisms underlying apomixis becomes essential for the genetic engineering of apomixis into major crops.

Buffelgrass (*Cenchrus ciliaris*) is a highly polymorphic, perennial warm-season grass widely grown in low latitude arid region due to its high drought tolerance (Skerman *et al.,* 1990). It is also a popular pasture grass in rangeland for its capacity to tolerate heavy grazing (Humphreys 1980). Buffelgrass (*Cenchrus ciliaris*) was first observed to reproduce by apospory and pseudogamy (endosperm development with fertilization) (Fisher *et al.,* 1954; Snyder *et al.,* 1955). Most of the native buffelgrass accessions reproduce by obligate apospory (Hignight *et al.,* 1991), and the discovery of a rare sexual buffelgrass (Bashaw 1962) offered the opportunity to study the genetic inheritance of apospory. Several molecular markers associated with apomixis

have been mapped in buffelgrass and its relative species *Pennisetum squamulatum* (Roche *et al.,* 1999; Jessup *et al.,* 2002; Ozias-Akins *et al.,* 2003; Dwivedi *et al.,* 2007). It is noteworthy that in the natural apomictic grass *Pennisetum squamulatum*, the genetic mapping study revealed 12 molecular markers that strictly cosegregate with apospory in a population comprised of 397 individuals that segregate for sexuality and apospory, which defines the apospory-specific genomic region (ASGR) (Ozias-Akins *et al.,* 1998). The 12 molecular markers were tested for linkage to apospory in buffelgrass, and the ASGR was shown to be highly conserved between these two species (Roche *et al.,* 1999; Goel *et al.,* 2003; Goel *et al.,* 2006). The *PsASGR-BabyBoom-Like* gene was then discovered through sequencing the ASGR-linked bacterial artificial clones (Gualtieri *et al.,* 2006; Conner *et al.,* 2008) and demonstrated to induce parthenogenesis in sexual pearl millet, rice and maize (Conner *et al.,* 2015; Conner *et al.,* 2017). Other apomixis candidate genes or apomixis-associated cDNAs were identified through differential display between sexual and apomictic buffelgrass reproductive tissues (Vielle-Calzada *et al.,* 1996; Singh *et al.,* 2007), although further characterization or validation for these candidates was not pursued.

Transcriptomic comparisons between sexual and apomictic reproductive tissues across different development stages have been another strategy to understand the transcriptomic basis of apomixis (Sharbel *et al.,* 2009; Sharbel *et al.,* 2010; Okada *et al.,* 2013; Bräuning *et al.,* 2018). Some apomixis candidate genes have been identified in this way, for example, the *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*PpSERK*) and *APOSTART* were identified in *Poa pratensis* through differential display using cDNA-AFLPs (Albertini *et al.,* 2004; Albertini *et al.,* 2005). However, none of these candidate genes has been experimentally shown to be causal for apomixis.

Here we carried out an RNA-seq study of unpollinated sexual and apomictic buffelgrass ovules on the day of anthesis, allowing us to compare gene expression between sexual and apomictic ovules at the stage of parthenogenesis initiation, and most importantly, to generate reference sequence information for further high-resolution research in an effort to identify parthenogenesis pathway genes.

Materials and Methods

Plant material, Ovary clearing and Ovules collection

The *Cenchrus ciliaris* plants used as a source of materials were vegetatively propagated tillers of the sexual genotype B-2s and natural apomictic genotype B-12-9 (Sherwood *et al.,* 1994). The plants were grown in the greenhouse (24-30 °C) for head collection. The heads were bagged prior to stigma exsertion and stigmas were manually removed with tweezers upon appearance. The heads remained bagged until collection.

For ovary clearing, heads were collected on the day of anthesis (Day0) and two days after (Day2) and immediately fixed in Farmer's fixative (ethanol: glacial acetic acid, 3:1) (Kerk *et al.,* 2003). Ovaries were carefully dissected from the florets, and were dehydrated in a graded ethanol series for two hours in each step as follows: 70% ethanol, 85% ethanol, and 100% ethanol, followed by another 100% ethanol incubation for overnight. Ovaries were then transferred to a methyl salicylate series for 2 hours in each step as follows: ethanol: methyl salicylate (3:1), ethanol: methyl salicylate (1:1), ethanol: methyl salicylate (1:3), and followed by a 100% methyl salicylate incubation for overnight (Young *et al.,* 1979). Cleared ovaries were then observed under a DIC (differential interference contrast) microscope.

For ovules collection, florets with fresh anthers half or fully exserted were collected from 8:00 – 11:00 am on the day of anthesis, and immediately fixed in Farmer's fixative (ethanol:

glacial acetic acid, 3:1) and stored at 4 °C for overnight. Fixed florets were then transferred and stored in 70% ethanol (DEPC-treated water) at 4 °C prior to dissection. Intact ovaries were carefully dissected from the florets on ice using fine tweezers to avoid physical damage, and ovules were subsequently isolated by tearing apart the ovary wall from the ovaries. Fifty ovules per sample were thus collected.

RNA extraction, cDNA Synthesis, Amplification, Library Construction and Sequencing

Ovule RNA was extracted using RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's protocol. The quantity and quality of RNA were checked using the Qubit 2.0 Fluorometer RNA assay (Invitrogen) and an Agilent 2100 Bioanalyzer using a RNA 6000 nano kit (Agilent Technologies), respectively. For each of the samples, 500 ng of total RNA was used for poly-A capture, cDNA synthesis and amplification using KAPA stranded RNA-seq Kit. Libraries were quality checked with the Qubit 2.0 Fluorometer ds DNA high sensitivity assay (Invitrogen) and Fragment AnalyzerTM Automated CE System (Agilent Technologies), and were sequenced on a NextSeq (300 Cycles) PE150 Mid Output flow cell on which three biological replicates of two genotypes were pooled.

Trinity De Novo RNA-Seq Assembly

Trinity-2.8.4.simg (Grabherr *et al.,* 2011; Haas *et al.,* 2013) was used for de novo transcriptome assembly. Raw reads were cleaned by removing adapter sequences and sequences of poor-quality using trimmomatic (Bolger *et al.,* 2014) (parameters LEADING:3TRAILING:3 MINLEN:36). A few bases from 5' and 3' end of the trimmed reads were further cut using TrimGalore (Krueger 2015) to avoid potential sequence bias. The ~291 million cleaned RNA-seq reads were combined into a single fastq file, normalized using the *in silico* normalization step contained in Trinity. The resulting normalized reads were then

assembled using Trinity (parameters --seqType fq --single --run_as_paired --max_memory 150G --CPU 8 --no_version_check --normalize_reads).

BLAST Annotation and Gene Ontology Analyses

Trinity contigs were first annotated by a BLASTN search against the latest NCBI nt database (e-value cutoff of 1e-10), then were applied to BLASTX (Camacho *et al.,* 2009) search against (e-value = 1e-5) a comprehensive protein database comprised of the Swiss-Prot (Boeckmann *et al.,* 2005) and UniRef90 protein databases. Longest possible open reading frames and putative coding regions were predicted within each trinity transcripts using TransDecoder (http://transdecoder.github.io). The predicted coding sequences were further annotated through BLASTP (e-value = 1e-5) search against the comprehensive protein database mentioned above and for protein domains search using hmmer (http://hmmer.org/) and PFam (Finn *et al.,* 2013). Potential signal peptides were also predicted using SignalP (Petersen *et al.,* 2011). All results were integrated by Trinotate, stored in a SQLite database, and then reported as a tab-delimited excel file. The gene ontology functional analyses were based on Trinotate-assigned GO annotations.

BUSCO completeness analyses

The completeness of the transcriptome assembly was examined by subjecting the Trinity.fasta file to BUSCO (Simão *et al.,* 2015) analyses using the command run_busco -i Trinity.fasta -l liliopsida_odb10 -m tran.

Transcript Abundance Estimation and Differential Expression Analyses

Kallisto (Bray *et al.,* 2016) software was used to quantitate the expression level for the transcripts, and DEseq2 (Love *et al.,* 2014) was used to identify significantly differentially expressed transcripts.

Results and Discussions

Parthenogenetic embryo frequency

Twelve percent of the unpollinated B-12-9 ovaries on Day 2 were shown to contain parthenogenetic embryo (Figure 2.1; Table 2.1) as evidenced by the presence of a polar nucleus structure, and no parthenogenetic embryo was observed in B-2s. This result showed that only the apomictic genotype of buffelgrass B-12-9 is able to express parthenogenesis.

A de novo assembled ovule transcriptome

A total number of ~ 331 million 150-base paired-end reads (Table 2.2) were generated. After preprocessing, about 291 million cleaned reads (Table 2.2) remained and were used for sexual, apomictic and combined assembly, using Trinity (Grabherr *et al.,* 2011; Haas *et al.,* 2013). For the combined assembly, *in silico* normalization retained 27.02% of reads for assembly (78,702,132 reads). We then used Trinity to generate an assembly of 204,909 transcript contigs clustered into 104,039 ‘‘gene’’ groupings (median transcript length: 690 bases, N50 of 1964 bases). Although the *in silico* normalization filtered out the vast majority of the reads prior to assembly, 98.51 % of the original reads still mapped back to the assembly (Table 2.3).

Functional Annotation with BLAST and Assessment of Transcriptome Completeness

For 204,909 transcripts assembled, 126,928 (61.94%) hit the NCBI nt database (e-value < $1e^{-10}$), and 98,922 (77.93%) of them hit a close relative, *Setaria italica* (foxtail millet), suggestive of a close evolutionary relationship between these two species, consistent with previous studies (Donadío *et al.,* 2009; Chemisquy *et al.,* 2010; Kellogg 2017). We then applied BLASTX to the 204,909 transcripts in our assembly, 76,580 transcripts were reported to match 14,077 Swiss-Prot proteins. 6,457 proteins are represented by nearly full-length transcripts, having over 80% alignment coverage, and 4671 of them are covered by more than 90% of their

protein lengths. We also subjected our transcriptome to BUSCO analysis (Simão *et al.,* 2015) that is based on the examination of target transcriptome for its representation in a group of evolutionarily conserved near-universal single-copy orthologs, and identified by this measure that the transcriptome contains near-complete gene sequence information for 93.4% of the genes in the *Cenchrus* genome (Table 2.4).

Functional landscape of the *Cenchrus ciliaris* ovule transcriptome

Trinotate-assigned gene ontology terms revealed that in unpollinated *Cenchrus ciliaris* ovules on the day of anthesis, the potential protein products predicted from the transcriptome are likely to be located in 915 different cellular components (cellular anatomy), playing 2,452 molecular functions and participating in 4,477 biological processes. The top 30 most represented cellular components (Figure 2.2), molecular functions (Figure 2.3) and biological processes (Figure 2.4) suggest that the unpollinated ovule may have an active transcription state on the day of anthesis. These data would be valuable for future research on gene identification and transcriptional basis elucidation of the natural parthenogenesis, and can also be beneficial for buffelgrass breeding programs. For example, there are 72 transcripts in the category of cold acclimation, which could potentially be used to search for cold-related genes for breeding cold-tolerant buffelgrass. There are 7 transcripts that belong to embryo development ending in birth or egg hatching, 7 transcripts that belong to somatic embryogenesis; 21 transcripts that belong to embryo sac cellularization as well as 72 transcripts related to embryo sac egg cell differentiation. These may be of great interests to identify genes or pathways related to apomictic egg cell and embryo development.

Differential Expression Analysis Identified Apomictic-upregulated Gene Expression and Splicing

We estimated the expression profiles for all transcripts across all samples, and found excellent correlation between biological replicates (Figure 2.5). As expected, ovule samples derived from the same genotype have more highly correlated expression. We next identified differentially expressed transcripts by DESeq2. We identified 8,725 transcripts corresponding to 5,811 genes at least 4-fold differentially expressed (False Discovery Rate = 0.05) between sexual and apomictic ovules, with 5,487 transcripts upregulated in sexual ovule and 3,238 upregulated in apomictic ovules.

In order to view common plant pathways enriched among apomictic upregulated genes, we extracted all apomictic upregulated ($Log_2FC>2$, FDR < 0.05) transcripts and their corresponding orthologous protein ID in *Arabidopsis* by doing BLASTX (cutoff 1e -5) search against Araport11_genes.201606.pep.fasta (https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FProteins%2FAraport11_protein_lists). We obtained the top 200 most apomictic upregulated transcripts (with lowest FDR) that matched 169 *Arabidopsis* orthologs ID and ran a functional protein association networks STRING (https://string-db.org/). Among the top apomictic upregulated transcripts, enriched GO terms or KEGG and reactome pathways included but were not limited to response to stress, spliceosome, transcriptional regulation by small RNAs, gene expression (transcription), and signaling by nuclear receptors. Since the spliceosome was the only enriched KEGG pathway, further analyses were focused on different isoform usages among differentially expressed transcripts between sexual and apomictic ovules.

Of the 8,725 differentially expressed transcripts (5,811 genes), there were 1,709 genes (4,623 transcripts) with putative alternatively spliced isoforms. For 714 genes (2,194 transcripts), alternative isoforms have at least one isoform enriched in either sexual or apomictic ovules, with 1,165 isoforms enriched in sexual and 1,029 enriched in apomictic groups, suggestive of alternate functional roles of this group of genes between sexual and apomictic ovules.

By examining the differentially expressed genes list and a few previously identified apomixis candidate or proven genes, *Pca21* (Singh *et al.,* 2007) but not *ASGR-BBML* (Gualtieri *et al.,* 2006; Conner *et al.,* 2008; Conner *et al.,* 2015; Conner *et al.,* 2017) was found to be upregulated in apomictic ovules. *Pca21* was identified through a subtractive hybridization method from ovary libraries at the initial stage of megasporogenesis or aposporous initial formation. Since its upregulation is not specific to either apomeiosis or parthenogenesis stages and the expression of these two components are separated (Conner *et al.,* 2013), *Pca21* might play a more general role in apomictic embryo sac development. No *ASGR-BBML* contig was present in the assembly, but expression was detected through bowtie alignment (Langmead *et al.,* 2013) to an *ASGR-BBML* gene deposited in NCBI database at a frequency of 6 out of ~291 million cleaned reads. This suggests the whole ovule RNA-seq is not effective in reflecting the detailed transcriptional structure of apomixis, as the tissues expressing the components of apomixis are basically at a single-cell level, with aposporous initial expressing apomeiosis and egg cell expressing parthenogenesis. The transcripts expressed from critical single cells expressing apomixis are diluted by irrelevant tissues or cells (nucellus, integuments etc.) in ovules. To efficiently study the transcriptional essence of apomixis, laser-capture microdissection-based RNA-seq on cells across different stages describing the full trajectory of apomeiosis and parthenogenesis development would be essential.

Understanding the molecular basis of apomixis would be an essential step towards utilization of apomixis in plant breeding programs. We created a de novo ovule transcriptome for sexual and apomictic *Cenchrus ciliaris* ovules on the day of anthesis when the development of parthenogenesis is underway, providing near-complete coding sequences information for *Cenchrus ciliaris*. Putative transcriptional pathways differentially expressed between apomictic and sexual ovules at the stage where parthenogenesis is being expressed were identified, providing us the hints to pathways underlying parthenogenesis development. Most importantly, the data and analyses generated from this study will serve as a reference for further research aiming at the transcriptomic determinants of parthenogenesis.

Figures and Tables

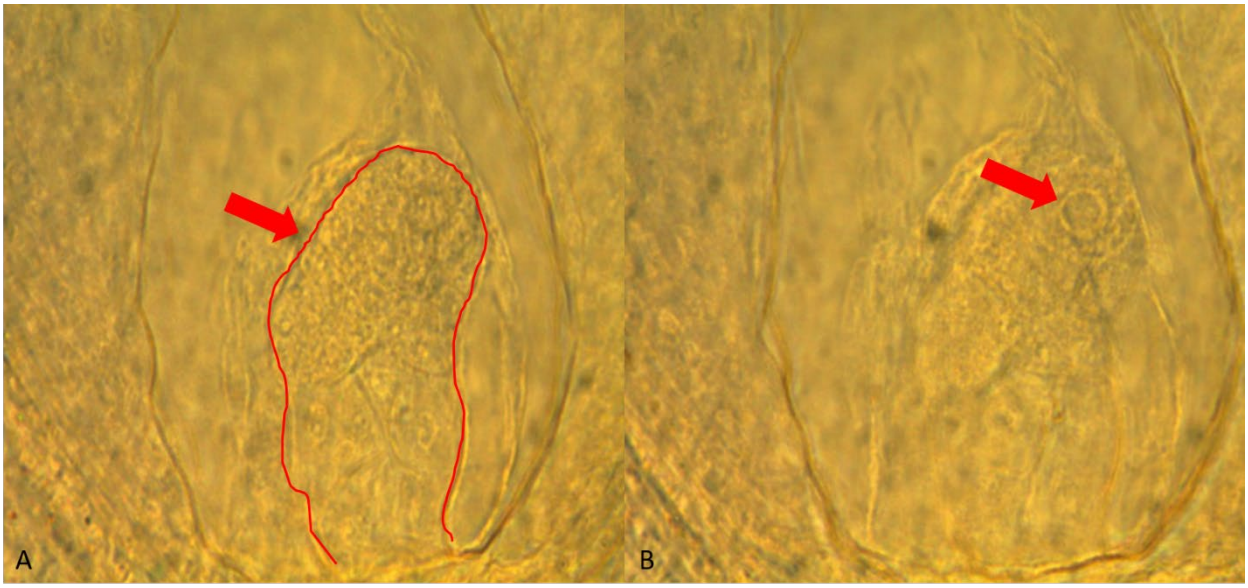

Figure 2.1: (A) a parthenogenetic embryo structure in B-12-9 ovary indicated by a red curve and arrow; (B) a polar nucleus structure indicated by the red arrow.

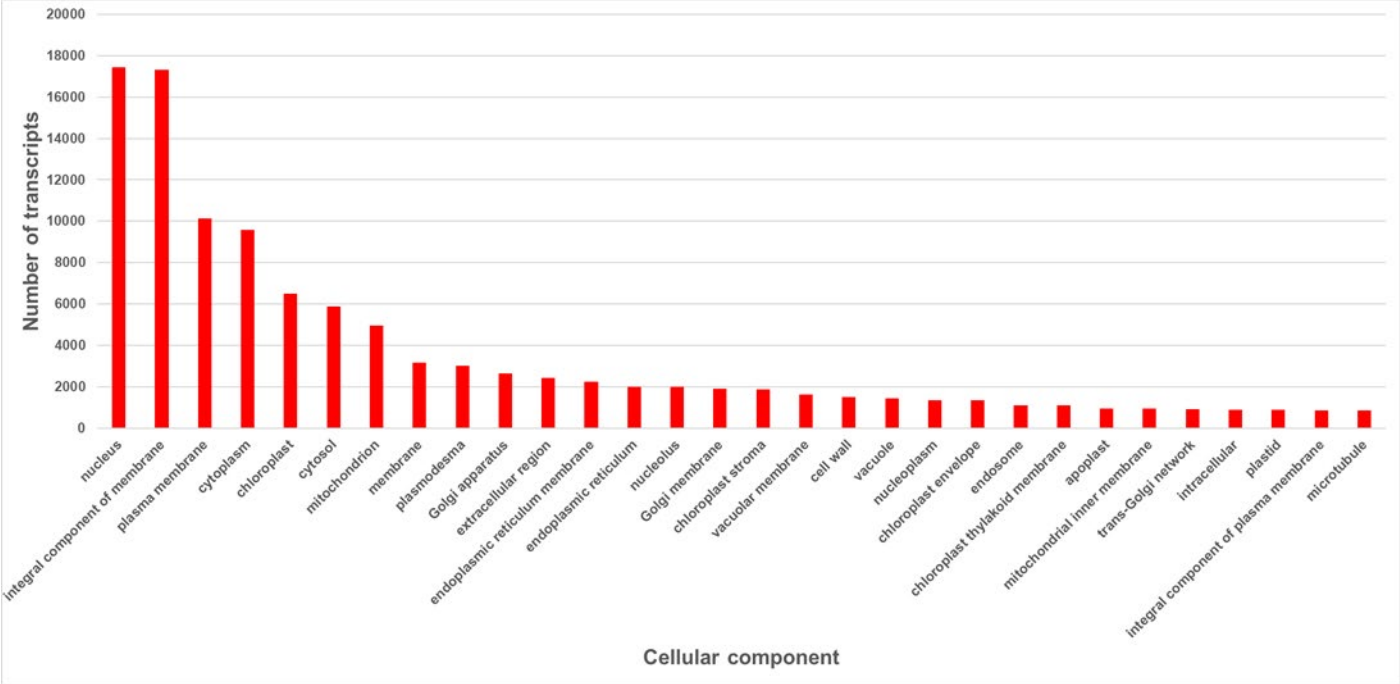

Figure 2.2: Top 30 most represented cellular component GO terms in the *Cenchrus ciliaris* ovule transcriptome and the corresponding number of transcripts.

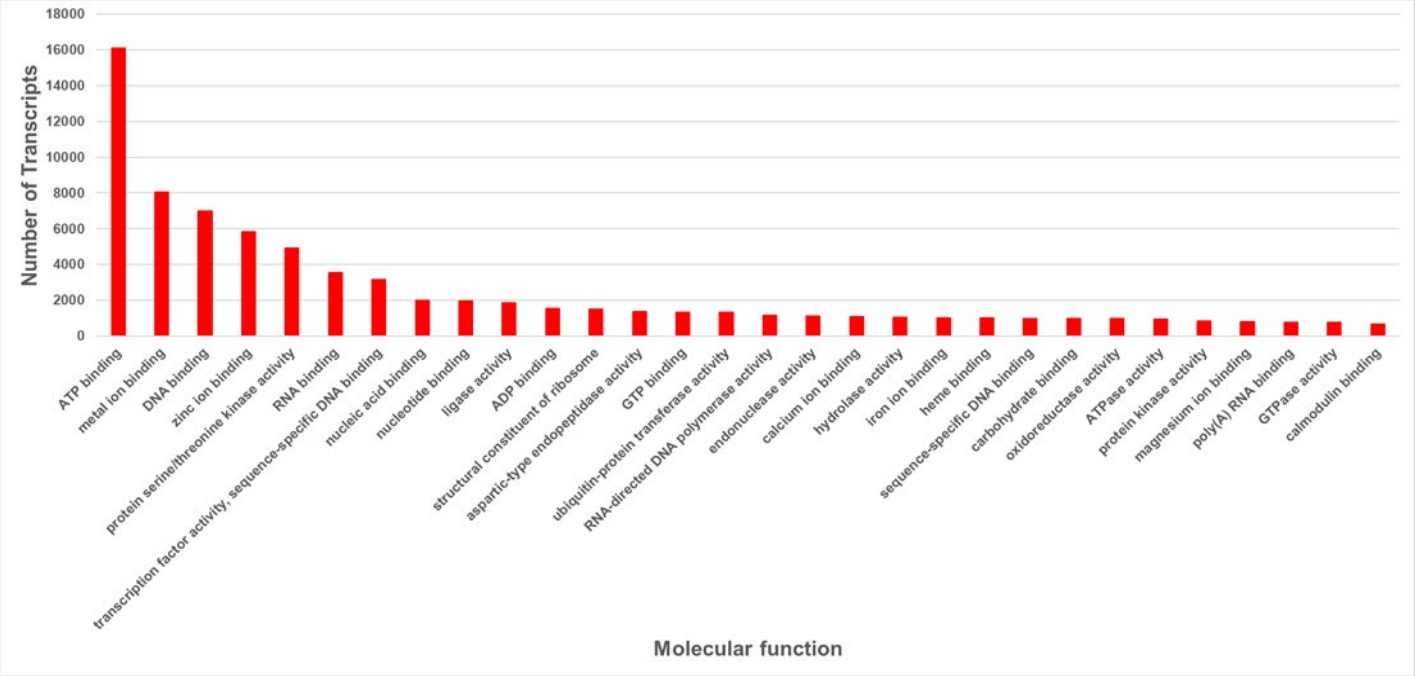

Figure 2.3: Top 30 most represented molecular function GO terms in the *Cenchrus ciliaris* ovule transcriptome and the corresponding number of transcripts.

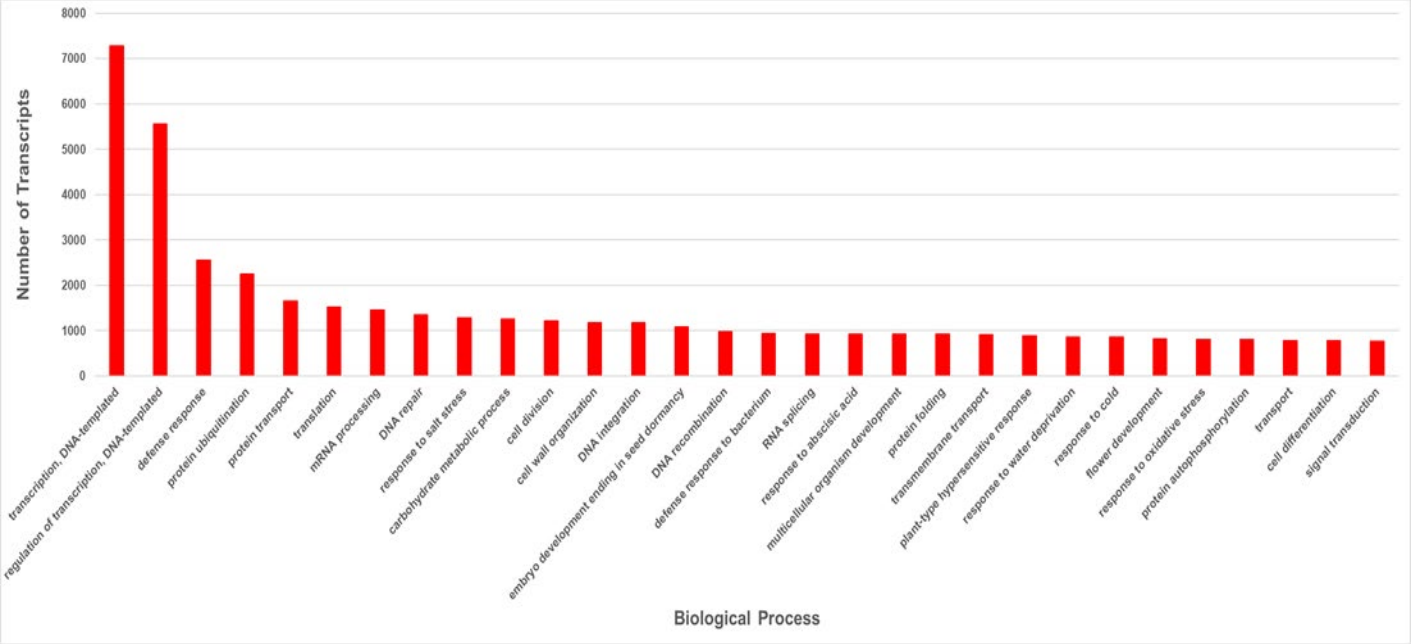

Figure 2.4: Top 30 most represented biological process GO terms in the *Cenchrus ciliaris* ovule transcriptome and the corresponding number of transcripts.

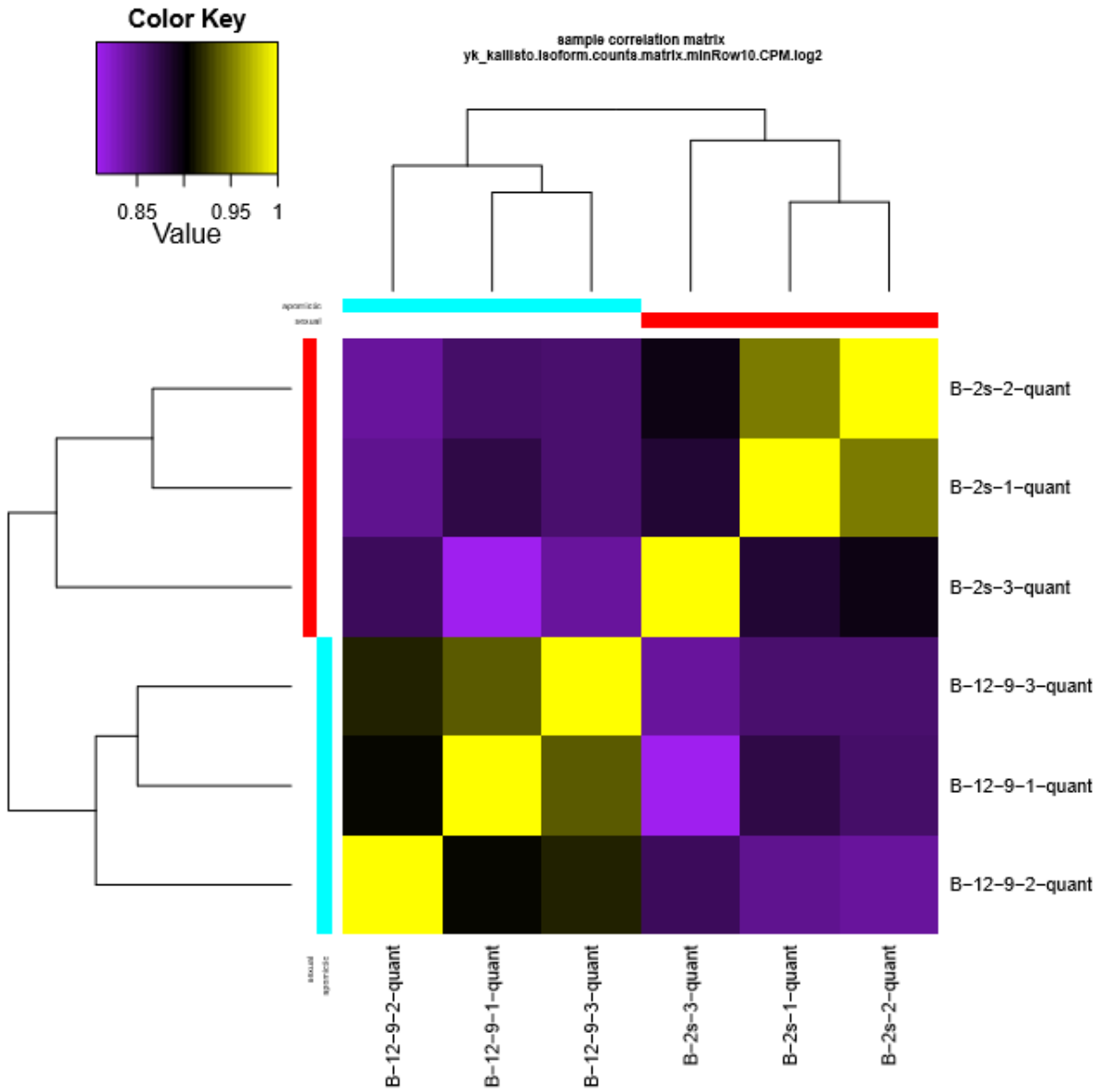

Figure 2.5: Pearson correlation heatmap between replicates.

Table 2.1: Parthenogenetic embryo frequency in sexual and apomictic ovaries on Day0 and Day2.

Genotype/date	Number of ovaries observed	Number of ovary containing parthenogenetic embryo	Frequency of parthenogenetic embryo
B-2s Day0	50	0	0
B-2s Day2	50	0	0
B-12-9 Day0	50	0	0
B-12-9 Day2	50	6	12%

Table 2.2: Raw data and cleaned data information

Sample	Number of Raw reads	Reads Survived from trimmomatic and trim_galore cleaning
B-2s-1-R1	26,090,932	24,570,807
B-2s-1-R2	26,090,932	24,570,807
B-2s-2-R1	25,252,490	23,966,674
B-2s-2-R2	25,252,490	23,966,674
B-2s-3-R1	25,651,180	23,035,544
B-2s-3-R2	25,651,180	23,035,544
B-12-9-1-R1	22,676,570	21,499,865
B-12-9-1-R2	22,676,570	21,499,865
B-12-9-2-R1	29,717,947	27,840,316
B-12-9-2-R2	29,717,947	27,840,316
B-12-9-3-R1	26,187,818	24,718,686
B-12-9-3-R2	26,187,818	24,718,686

Table 2.3: Sexual, apomictic and combined trinity assembly information

	Total trinity 'genes'	Total trinity transcripts	Contig N50 (bases)	Average contig (bases)	Median contig length (bases)	Alignment rate
Sexual	79,484	152,191	1,911	1,153.03	742	98.46%
Apomictic	80,784	152,237	1,951	1,162.82	737	98.39%
Combined	104,039	204,909	1,964	1,146.35	690	98.51%

Table 2.4: BUSCO analyses of Transcriptome Completeness

	Combined assembly
Complete BUSCOs (C)	3,062
Complete and single-copy BUSCOs (S)	917
Complete and duplicated BUSCOs (D)	2,145
Fragmented BUSCOs (F)	98
Missing BUSCOs (M)	118
Total BUSCO groups searched	3278
% of complete BUSCOs	93.4%
% of fragmented BUSCOs	3.0%
% of missing BUSCOs	3.6%

References

Albertini, E., G. Marconi, G. Barcaccia, L. Raggi and M. Falcinelli (2004). "Isolation of candidate genes for apomixis in *Poa pratensis* L." Plant molecular biology 56(6): 879-894.

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

Bashaw, E. C. (1962). "Apomixis and Sexuality in Buffelgrass 1." Crop Science 2(5): 412-415.

Bicknell, R. A. and A. M. Koltunow (2004). "Understanding apomixis: recent advances and remaining conundrums." The Plant Cell 16(suppl 1): S228-S245.

Boeckmann, B., M.-C. Blatter, L. Famiglietti, U. Hinz, L. Lane, B. Roechert and A. Bairoch (2005). "Protein variety and functional diversity: Swiss-Prot annotation in its biological context." Comptes rendus biologies 328(10-11): 882-899.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics 30(15): 2114-2120.

Bräuning, S., A. Catanach, J. M. Lord, R. Bicknell and R. C. Macknight (2018). "Comparative transcriptome analysis of the wild-type model apomict *Hieracium praealtum* and its loss of parthenogenesis (lop) mutant." BMC plant biology 18(1): 206.

Bray, N. L., H. Pimentel, P. Melsted and L. Pachter (2016). "Near-optimal probabilistic RNA-seq quantification." Nature biotechnology 34(5): 525.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden (2009). "BLAST+: architecture and applications." BMC bioinformatics 10(1): 421.

Carman, J. G. (1997). "Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony." Biological Journal of the Linnean Society 61(1): 51-94.

Chemisquy, M. A., L. M. Giussani, M. A. Scataglini, E. A. Kellogg and O. Morrone (2010). "Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): a combined nuclear, plastid and morphological analysis, and nomenclatural combinations in Cenchrus." Annals of botany 106(1): 107-130.

Conner, J. A., S. Goel, G. Gunawan, M.-M. Cordonnier-Pratt, V. E. Johnson, C. Liang, H. Wang, L. H. Pratt, J. E. Mullet and J. DeBarry (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., G. Gunawan and P. Ozias-Akins (2013). "Recombination within the apospory specific genomic region leads to the uncoupling of apomixis components in *Cenchrus ciliaris*." Planta 238(1): 51-63.

Conner, J. A., M. Mookkan, H. Huo, K. Chae and P. Ozias-Akins (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.

Conner, J. A., M. Podio and P. Ozias-Akins (2017). "Haploid embryo production in rice and maize induced by PsASGR-BBML transgenes." Plant reproduction 30(1): 41-52.

Donadío, S., L. M. Giussani, E. A. Kellogg, F. O. Zuolaga and O. Morrone (2009). "A preliminary molecular phylogeny of *Pennisetum* and *Cenchrus* (Poaceae-Paniceae) based on the trnL-F, rpl16 chloroplast markers." Taxon 58(2): 392-404.

Dwivedi, K. K., S. R. Bhat, V. Bhat, B. V. Bhat and M. G. Gupta (2007). "Identification of a SCAR marker linked to apomixis in buffelgrass (*Cenchrus ciliaris* L.)." Plant Science 172(4): 788-795.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm and J. Mistry (2013). "Pfam: the protein families database." Nucleic acids research 42(D1): D222-D230.

Fisher, W. D., E. C. Bashaw, E. C. Holt, L. M. Soffer and E. Carpenter (1954). "2456201. Evidence for apomixis in *Pennisetum ciliare* and *Cenchrus setigerus*." Agronomy journal 46(9): 401-404.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury and Q. Zeng (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nature biotechnology 29(7): 644.

Gualtieri, G., J. A. Conner, D. T. Morishige, L. D. Moore, J. E. Mullet and P. Ozias-Akins (2006). "A segment of the apospory-specific genomic region is highly microsyntenic not only between the apomicts *Pennisetum squamulatum* and buffelgrass, but also with a rice chromosome 11 centromeric-proximal genomic region." Plant physiology 140(3): 963-971.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li and M. Lieber (2013). "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." Nature protocols 8(8): 1494.

Hignight, K. W., E. C. Bashaw and M. A. Hussey (1991). "Cytological and morphological diversity of native apomictic buffelgrass, *Pennisetum ciliare* (L.) Link." Botanical Gazette 152(2): 214-218.

Hojsgaard, D., S. Klatt, R. Baier, J. G. Carman and E. Hörandl (2014). "Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics." Critical Reviews in Plant Sciences 33(5): 414-427.

Humphreys, L. R. (1980). "A guide to better pastures for the tropics and subtropics. rev." Ermington: W. Stepheson.

Jessup, R. W., B. L. Burson, G. B. Burow, Y. W. Wang, C. Chang, Z. Li, A. H. Paterson and M. A. Hussey (2002). "Disomic inheritance, suppressed recombination, and allelic interactions govern apospory in buffelgrass as revealed by genome mapping." Crop Science 42(5): 1688-1694.

Kellogg, E. A. (2017). Evolution of *Setaria*. Genetics and Genomics of Setaria, Springer: 3-27.

Kerk, N. M., T. Ceserani, S. L. Tausta, I. M. Sussex and T. M. Nelson (2003). "Laser capture microdissection of cells from plant tissues." Plant physiology 132(1): 27-35.

Krueger, F. (2015). Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Bioinformatics, Cambridge, United Kingdom.

Langmead, B. and S. L. Salzberg (2013). "Langmead. 2013. Bowtie2." Nature methods 9: 357-359.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome biology 15(12): 550.

Morgan, R. N., P. Ozias-Akins and W. W. Hanna (1998). "Seed set in an apomictic BC3 pearl millet." International journal of plant sciences 159(1): 89-97.

Nogler, G. A. (1984). Gametophytic apomixis. Embryology of angiosperms, Springer: 475-518.

Okada, T., Y. Hu, M. R. Tucker, J. M. Taylor, S. D. Johnson, A. Spriggs, T. Tsuchiya, K. Oelkers, J. C. M. Rodrigues and A. M. G. Koltunow (2013). "Enlarging cells initiating apomixis in *Hieracium praealtum* transition to an embryo sac program prior to entering mitosis." Plant physiology 163(1): 216-231.

Ozias-Akins, P., Y. Akiyama and W. W. Hanna (2003). "Molecular characterization of the genomic region linked with apomixis in *Pennisetum/Cenchrus*." Functional & integrative genomics 3(3): 94-104.

Ozias-Akins, P., D. Roche and W. W. Hanna (1998). "Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes." Proceedings of the National Academy of Sciences 95(9): 5127-5132.

Ozias-Akins, P. and P. J. van Dijk (2007). "Mendelian genetics of apomixis in plants." Annu Rev Genet 41: 509-537.

Petersen, T. N., S. Brunak, G. Von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nature methods 8(10): 785.

Roche, D., P. Cong, Z. Chen, W. W. Hanna, D. L. Gustine, R. T. Sherwood and P. Ozias-Akins (1999). "An apospory-specific genomic region is conserved between Buffelgrass (*Cenchrus ciliaris* L.) and *Pennisetum squamulatum* Fresen." The Plant Journal 19(2): 203-208.

Savidan, Y. (2001). "Transfer of apomixis through wide crosses." Flowering of Apomixis: From Mechanisms to Genetic Engineering, Y. Savidan, JG Carman, and T. Dresselhaus, eds (Mexico: CIMMYT, IRD, European Commission DG VI): 153-167.

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

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

Sherwood, R. T., C. C. Berg and B. A. Young (1994). "Inheritance of apospory in buffelgrass." Crop Science 34(6): 1490-1494.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov (2015). "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." Bioinformatics 31(19): 3210-3212.

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

Skerman, P. J. and F. Riveros (1990). "FAO plant production and protection series." FAO and UN, Rome, Italy.

Smith, J. (1841). "XXXII. Notice of a Plant which produces perfect Seeds without any apparent Action of Pollen." Transactions of the Linnean Society of London(4): 509-512.

Snyder, L. A., A. R. Hernandez and H. E. Warmke (1955). "The mechanism of apomixis in *Pennisetum ciliare*." Botanical Gazette 116(3): 209-221.

Vielle-Calzada, J. P., M. L. Nuccio, M. A. Budiman, T. L. Thomas, B. L. Burson, M. A. Hussey and R. A. Wing (1996). "Comparative gene expression in sexual and apomictic ovaries of *Pennisetum ciliare* (L.) Link." Plant molecular biology 32(6): 1085-1092.

Young, B. A., R. T. Sherwood and E. C. Bashaw (1979). "Cleared-pistil and thick-sectioning techniques for detecting aposporous apomixis in grasses." Canadian Journal of Botany 57(15): 1668-1672.

Appendix 2.A: Supplemental Experimental Procedures

Ovule RNA-seq Library Preparation

KAPA stranded RNA-seq library preparation

First-time Poly-A capture

1. Pipette 500 ng of total RNA per reaction in less than a total volume of 50 µl to a 96-well plate, add nuclease-free water to allow each reaction to reach a total volume of 50 µl.
2. Let Dynabeads®, washing buffer B, and binding buffer equilibrate to room temperature and re-suspend the Dynabeads® thoroughly by vortexing for at least 1 minute.
3. Pipette 25 µl of re-suspended Dynabeads® into each reaction.
4. Place the 96-well plate on the magnetic stand and incubate at room temperature until the solution is clear, discard the supernatant without disturbing the beads.
5. Remove the 96-well plate from the magnetic stand, add 50 µl of binding buffer to each reaction and thoroughly re-suspend the beads by pipetting up and down for at least 10 times.
6. Place the 96-well plate on the magnetic stand and incubate at room temperature until the solution is clear, discard the supernatant without disturbing the beads.
7. Remove the 96-well plate from the magnetic stand, add 50 µl of binding buffer to each reaction and thoroughly re-suspend the beads by pipetting up and down for at least 10 times.
8. Add 50 µl of total RNA (500 ng) to the 50 µl Dynabeads® suspension and mix thoroughly by pipetting up and down for at least 10 times. Seal the 96-well plate.

9. Incubate the 96-well plate at 65 °C for 2 minutes, then incubate at room temperature for 5 minutes. During the 5-minute room temperature incubation, mix the sample at the time point of 2 minutes and 4 minutes by gently pipetting up and down for at least 10 times.
10. Place the total RNA and Dynabeads® mixtures on the magnetic stand and incubate at room temperature until the solution is clear. Discard the supernatant.
11. Remove the 96-well plate from the magnetic stand, add 200 µl of washing buffer B to each reaction and thoroughly re-suspend the beads by gently pipetting up and down for at least 10 times.
12. Place the 96-well plate on the magnetic stand and incubate at room temperature until the solution is clear. Discard the supernatant.
13. Remove the 96-well plate from the magnetic stand, add 50 µl of 10 mM Tris-HCl to each reaction and thoroughly re-suspend the beads by gently pipetting up and down for at least 10 times. Seal the 96-well plate.

Second-time Poly-A capture

14. Incubate the 96-well plate at 70°C for 2 minutes, and then incubate at room temperature at room temperature for 5 minutes. During the 5-minute room temperature incubation, mix the sample at the time point of 2 minutes and 4 minutes by gently pipetting up and down for at least 10 times.
15. Add 50 µl of binding buffer to RNA and Dynabeads® mixture and mix by gently pipetting up and down for at least 10 times.
16. Incubate the 96-well plate at room temperature for 5 minutes.

17. Place the 96-well plate on the magnetic stand and incubate at room temperature until the solution is clear. Discard the supernatant.
18. Remove the 96-well plate from the magnetic stand, add 200 µl of washing buffer B to each reaction and thoroughly re-suspend the beads by pipetting up and down for at least 10 times.
19. Place the 96-well plate on the magnetic stand and incubate at room temperature until the solution is clear. Discard the supernatant.
20. RNA Fragmentation
21. Prepare the 1X Fragment, Prime and Elute Buffer (FPE Buffer).

RNA fragmentation master mix		With 10% more
	1 rxn (µl) original amt.	2 rxns (µl)
Water (RNase-free water)	11	24.2
Fragment Prime and Elute Buffer (2x)	11	24.2
Total Volume	22	48.4
Distribute 22 µl of RNA fragmentation master mix per vial/well		

22. Add 22 µl of 1X FPE Buffer to each reaction, and re-suspend the Dynabeads® thoroughly by pipetting up and down for at least 10 times.
23. Place the 96-well plate in a thermal cycler and incubate at 85°C for 6 minutes.
24. Upon completion of fragmentation step, immediately incubate the fragmented RNA on ice for 1 minute.

25. Place the 96-well plate on the magnetic stand and incubate at room temperature until the solution is clear.
26. Carefully transfer 20 µl of supernatant for each reaction to a new 96-well plate. Proceed immediately.

First Strand cDNA synthesis

27. Prepare the First Strand cDNA synthesis master mix on ice.

1st strand synthesis master mix	Volume	
	1 rxn (µl) original amt.	2 rxns (µl)
1st strand synthesis buffer	11	22
KAPA Script	1	2
Total Volume	12	24
Distribute 10 µl of 1st strand synthesis master mix into 20 µl fragmented RNA		

28. Keep the 96-well plate on ice, and mix thoroughly by pipetting up and down for at least 10 times.
29. Incubate the 96-well plate as follows:

Temperature (^{0}C)	Time
25	10 min
42	15 min
70	15 min
4	Hold

30. Second Strand cDNA synthesis

2nd strand marking buffer master mix	Volume	
	1 rxn (µl) original amt.	2 rxns (µl)
2nd strand marking buffer	31	62
2nd strand synthesis enzyme mix	2	4
Total Volume	33	66
Distribute 30 µl of 2nd strand marking buffer master mix into 30 µl 1st strand synthesis rxn		

31. Mix thoroughly by pipetting up and down for at least 10 times.

32. Incubate the reactions as follows:

Temperature (^{0}C)	Time
16	1 hr
4	Hold

33. Keep the plate on ice and proceed immediately.

Cleanup

34. Take out the AMPure XP beads and equilibrate to room temperature, and thoroughly vortex the beads for 1 minute.

35. Add 108 µl AMPure XP beads to the reactions above. Pipette up and down for more than 10 times to thoroughly re-suspend the mixture.

36. Incubate the plate at room temperature for 10 minutes for DNA to bind to the beads.
37. Place the plate on the magnetic stand to capture the beads. Incubate 5 minutes until the liquid is clear.
38. Remove and discard 160 µl of supernatant carefully without disturbing the beads.
39. Keep the plate on the magnetic stand, and add 200 µl of 80% ethanol.
40. Incubate the plate on the magnetic stand at room temperature for 30 seconds.
41. Remove and discard the ethanol carefully without disturbing the beads.
42. Keep the plate on the magnetic stand, and add 200 µl of 80% ethanol.
43. Incubate the plate on the magnetic stand at room temperature for 30 seconds.
44. Remove and discard the ethanol. Try to remove all the residual ethanol without disturbing the beads.
45. Air-dry the beads for about 5 minutes at room temperature until the pellet is no longer shiny, but before it cracks.
46. Proceed immediately to A-Tailing

A-Tailing

47. Assemble the A-tailing mix as follows:

A-Tailing master mix	Original amount	With 10% more
	1 rxn (µl)	1 rxn (µl)
Water	24	26.4
10x KAPA A-Tailing buffer	3	3.3
KAPA A-tailing enzyme	3	3.3
Total Volume	30	33

48. Add 30 µl of the master mix to the beads and mix thoroughly by pipetting up and down for more than 10 times.

49. Incubate the plate according to the following protocol:

Step	Temperature (^{0}C)	Time
A-tailing	30	30 min
Enzyme Inactivation	60	30 min
Hold	4	-

50. Proceed to Adapter ligation immediately after the incubation.

Adapter Ligation

51. Assemble the ligation reaction mix:

Adapter ligation master mix	original amount	With 10% more
	1 rxn (µl)	1 rxn (µl)
Water	16	17.6
5x ligation buffer	14	15.4
KAPA T4 DNA ligase	5	5.5
Total Volume	35	38.5
Dispense 35 µl per rxn		

52. Add 5 µl of adapter at 350 nM to each sample as follows:

Samples	TruSeq Adapter
B-2s-1	AD006
B-2s-2	AD012
B-2s-3	AD002
B-12-9-1	AD004
B-12-9-2	AD016
B-12-9-3	AD007

Total ligation reaction volume is 70 µl.

53. Mix thoroughly by pipetting up and down several times to re-suspend the beads.

54. Incubate the plate at 20^0C for 15 minutes. Proceed immediately.

First Post-ligation Cleanup

55. Perform a 1X AMPure Beads-based by mixing the following:

Component	Volume
Beads bound with adapter-ligated DNA	70 µl
PEG/NaCl Solution	70 µl
Total Volume	140µl

56. Mix thoroughly by pipetting up and down several times.

57. Incubate the plate at room temperatures for 10 minutes to allow the DNA to bind to the beads.

58. Place the plate on a magnetic stand to capture the beads. Incubate 5 minutes or until the liquid is clear.
59. Remove and discard 135 µl of supernatant without disturbing the beads.
60. Keep the plate on the magnetic stand, add 200 µl of 80% ethanol.
61. Incubate the plate at room temperature for 30 seconds.
62. Remove and discard the ethanol carefully without disturbing the beads.
63. Keep the plate on the magnetic stand, add 200 µl of 80% ethanol.
64. Incubate the plate on the magnetic stand for 30 seconds.
65. Remove and discard the ethanol carefully without disturbing the beads, and try to remove all residual ethanol as thoroughly as possible.
66. Air-dry the beads at room temperature for ~ 5 minutes until the pellet is no longer shiny, but before it cracks.
67. Remove the plate from the magnetic stand.
68. Add 50 µl of 10 mM Tris-HCl (pH 8.0 – 8.5) to the beads, and thoroughly re-suspend by pipetting up and down several times.
69. Incubate the plate at room temperature for 2 minutes to elute DNA off the beads.

Second-time Post Ligation Cleanup

70. Perform a 1X AMPure Beads-based by mixing the following:

Component	Volume
Beads bound with adapter-ligated DNA	50 µl
PEG/NaCl Solution	50 µl
Total Volume	140 µl

71. Thoroughly re-suspend the beads by pipetting up and down several times.
72. Incubate the plate at room temperature for 10 minutes to allow the DNA to bind to the beads.
73. Place the plate on a magnetic stand to capture the beads. Incubate 5 minutes until the liquid is clear.
74. Remove and discard 95 µl of supernatant without disturbing the beads.
75. Keep the plate on the magnetic stand and add 200 µl of 80% ethanol.
76. Incubate the plate at room temperature for 30 seconds.
77. Remove and discard the ethanol without disturbing the beads.
78. Keep the plate on the magnetic stand and add 200 µl of 80% ethanol.
79. Incubate the plate at room temperature for 30 seconds.
80. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads.
81. Air-dry the beads at room temperature for 5 minutes until the pellet is no longer shiny, but before it cracks.
82. Remove the plate from the magnetic stand.
83. Thoroughly re-suspend the beads in 22.5 µl of 10 mM Tris-HCl (pH8.0-8.5).
84. Incubate the plate at room temperature for 2 minutes to allow the DNA to elute off the beads.

85. Place the plate on a magnetic stand to capture the beads. Incubate 5 minutes until the liquid is clear.

86. Transfer 20 µl of the clear supernatant to a new plate and proceed.

Library Amplification.

87. Assemble the library amplification master mix as follows:

Library amplification master mix	Original amount	With 10% more
	1 rxn (µl)	1 rxn (µl)
2x KAPA HiFl hotStart ready mix	25	27.5
10x KAPA library amplification primer mix	5	5.5
Total Volume	30	33
Dispense 30 µl of the master mix in each of the 20 µl ligated DNA.		

Total rxn volume is 50 µl.

88. Mix well by pipetting up and down several times.

89. Amplify the library using the following thermal cycling profile.

Temperature (^{0}C)	Time	Cycles

98	45 secs	-
98	15 secs	14
60	30 secs	
72	30 secs	
72	5 mins	-
4	Hold	-

90. Place the plates on ice and proceed

Library amplification clean up

91. 1xSPRI clean up, add 50 µl Agencourt® AmPure XP reagent to each of the 50 µl PCR rxn.

92. Thoroughly re-suspend the beads by pipetting up and down multiple times.

93. Incubate the plate at RT for 10 mins to allow the DNA to bind to the beads.

94. Place the plate on a magnet to capture the beads. Incubate (5 mins) until the liquid is clear.

95. Carefully remove and discard 95 µl of supernatant.

96. Keep the plate on the magnet and add 200 µl of 80% ETOH.

97. Incubate the plate at RT for greater than 30 secs (1 min).

98. Carefully remove and discard the ethanol.

99. Keep the plate on the magnet and add 200 µl of 80% ETOH.

100. Incubate the plate at RT for greater than 30 secs (1 min).

101. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads. (Note: Use 10 µl pipette to remove the remaining ethanol)

102. Allow the beads to air dry (5 mins) at RT sufficient for ETOH to evaporate

103. Remove the plate from the magnet.

104. Thoroughly re-suspend the beads in 24 µl of 10 mM Tris-HCl (pH8.0).

105. Incubate the plate at RT for 2 min to allow the DNA to elute from the beads.

106. Place the plate on a magnet to capture the beads. Incubate (5 mins) until the liquid is clear.

107. Transfer 20 µl of the clear supernatant to a new plate or 0.5ml tube and proceed.

108. Library QC

CHAPTER 3

NOVEL INSIGHTS INTO THE PARTHENOGENETIC POTENTIAL OF BUFFELGRASS (*CENCHRUS CILIARIS*) FROM THE PERSPECTIVE OF EGG APPARATUS TRANSCRIPTOME[2]

[2] Y.Ke, M.Podio, J.Conner and P. Ozias-Akins. To be submitted to *Molecular Plant*

Abstract

Apomixis, a type of asexual reproduction in angiosperms, results in progenies that are genetically identical to the mother plant. It is a highly desirable trait in agriculture due to its potential to preserve heterosis of F_1 hybrids through subsequent generations. However, no major crops are apomictic. Deciphering mechanisms underlying apomixis becomes one of the alternatives to engineer self-producing capability into major crops. Parthenogenesis, a major component of apomixis, commonly described as the ability to initiate embryo formation from the egg cell without fertilization, also can be valuable in plant breeding for doubled haploid production. Discovery of genes driving parthenogenesis in the natural apomictic buffelgrass (*Cenchrus ciliaris*), has been challenging due to limited genomic resources and technical difficulties in accessing the egg cell expressing parthenogenesis. By conducting laser capture microdissection-based RNA-seq on sexual and apomictic egg apparatus on the day of anthesis, we created a de novo transcriptome with sequence information for the *Cenchrus ciliaris* egg apparatus, identified transcriptional profiles that distinguish apomictic egg from its sexual counterpart, and suggested functional roles for a few transcription factors in promoting natural parthenogenesis. Our transcriptome data significantly complement previous gene expression studies and will be an important resource for future research on natural parthenogenesis, as well as for inquiries using sexual and apomictic egg cells that extend beyond those of parthenogenesis.

Introduction

Apomixis is commonly defined as a reproductive phenomenon in angiosperms where embryos form from maternal cells in the ovule without meiosis and syngamy, resulting in

asexual seed formation (Nogler 1984; Bicknell *et al.,* 2004; Ozias-Akins *et al.,* 2007). Apomixis is widespread in flowering plants and has now been described in 148 genera as a form of adventitious embryony, 110 genera as aposporous and 68 genera as diplosporous (Hojsgaard *et al.,* 2014). While apomicts are by nature clonal and can produce seeds of identical genotype to the mother plant, no major crops have this self-producing capability. Elucidating molecular mechanisms of apomixis will significantly enhance the plant breeding toolbox for preserving genetic composition of elite cultivars.

Significant strides have been made in deciphering the genetic control of apomixis in a wide range of natural apomicts (Ozias-Akins *et al.,* 2007), although genes underlying the multiple components comprising apomixis are still not fully elucidated. In the natural apomict *Pennisetum squamulatum*, mapping in a F_1 population segregating for apospory and sexuality derived from a cross between *Pennisetum glaucum* (pearl millet) X *Pennisetum squamulatum* revealed 12 apospory-linked markers, which defined a contiguous apospory-specific genomic region (ASGR) (Ozias-Akins *et al.,* 1998). The linkage of many of these markers to apospory were further shown in *Cenchrus ciliaris* (Roche *et al.,* 1999). Sequencing the ASGR-linked bacterial artificial chromosome clones in *Pennisetum squamulatum* and *Cenchrus ciliaris* led to the discovery of genes including AP2-domain containing transcription factor *ASGR-BBM-like* (Gualtieri *et al.,* 2006; Conner *et al.,* 2008). The *PsASGR-BBML* transgene can also induce parthenogenesis in maize and rice but not in *Arabidopsis* (Conner *et al.,* 2017).

Unbiased comparative genomics and transcriptomics study between apomictic plants and their sexual relatives or siblings often can help unlock the functional molecular components of apomixis that have not been genetically tractable; however, most natural apomictic genomes remain unsequenced and targeting the critical tissues where components of apomixis are

expressed for RNA-seq has been biologically and technically challenging. Recent expression studies have focused on and advanced our understanding of apomeiosis by investigating differentially expressed molecular signatures between sexual and apomictic reproductive tissues at the stage of apomixis initiation (Sharbel *et al.,* 2009; Sharbel *et al.,* 2010; Okada *et al.,* 2013). An important missing component for current expression study is the transcriptomic comparison between sexual egg and apomictic egg expressing parthenogenesis, which should serve as a reference template for discovering genes and pathways underlying natural parthenogenesis in the frame of apomixis.

Discovery of genes driving parthenogenesis in the natural apomict *Cenchrus ciliaris*, has been challenging due to limited genomic resources and technical difficulties in accessing the egg cell expressing parthenogenesis. By conducting laser capture microdissection (LCM)-based RNA-seq on unfertilized sexual and apomictic egg apparatus on the day of anthesis, combined with de novo transcriptome assembly and computational analyses, we created a de novo transcriptome with sequence information for *Cenchrus ciliaris* egg apparatus. Transcriptional profiles that distinguish apomictic egg from its sexual counterpart were identified, and suggested functional roles for a few key transcription factors and pathways in promoting natural parthenogenesis. Our transcriptome data significantly complemented previous gene expression studies and will be an important resource for future research on natural parthenogenesis, as well as for inquiries using sexual and apomictic egg cells that extend beyond those of parthenogenesis and eventually contribute to doubled haploid production and heterosis fixation.

Materials and Methods

Plant Material and Florets Collection

The *Cenchrus ciliaris* plants used as a source of materials were vegetatively propagated

tillers of the sexual genotype B-2s and natural apomictic genotype B-12-9 (Sherwood *et al.,* 1994). The plants were grown in the greenhouse (24-30 °C) for head collection from July to September in 2016. The heads were bagged prior to stigma exsertion and stigmas were manually removed with tweezers upon appearance. The heads remained bagged until the day of anther exsertion (anthesis) at which time florets were collected. From 8:00 – 11:00 am on the day of anthesis, florets with fresh anthers half or fully exserted were collected and immediately fixed in ethanol (ethyl alcohol 200 proof, catalog number:111000200, PHARMCO) : acetic acid (acetic acid, glacial, catalog number: A38S-212, Fisher Chemical), 3:1 (Kerk *et al.,* 2003) and stored at 4 °C for overnight. Fixed florets were then transferred and stored in 70% ethanol (DEPC-treated water) at 4 °C.

Tissue Preparation

Intact ovaries were carefully dissected from the florets on ice using fine tweezers to avoid physical damage, and were dehydrated in a graded ethanol series on ice with gentle shaking for 20 minutes in each step as follows: 80% ethanol, 90% ethanol, and three changes of 100% ethanol. Ovaries were stored overnight in 100% ethanol at 4 °C, and transferred the next day to a xylene series with gentle shaking for 20 minutes in each step as follows: ethanol: xylene (3:1), ethanol: xylene (1:1), ethanol: xylene (1:3), and three changes of 100% xylene. Xylene-cleared ovaries were then transferred to a Paraplast series in an incubator (54 °C) with shaking for 9 hours in each step as follows: xylene and Paraplast (1:1) mixture, and seven changes of 100 % paraplast. Paraplast-infiltrated ovaries were embedded in Paraplast blocks using a tissue embedding center, and 8-µm sections were cut on a rotary microtome (Leica RM2145, Germany). The sections were floated on methanol and mounted on metal frame PET foil slides (Leica) that had been UV-irradiated (DNA transfer lamp, Fotodyne) for 30 minutes. Slides were

heated on a slide warmer at 42 °C for overnight to air-dry and stretch the Paraplast ribbons. Prior to egg identification and laser capture microdissection, PET foil slides with ovary sections were de-paraffinized in xylene for two changes of 5 minutes each with gentle shaking and air-dried for 1 hour in the fume hood. A few ovary sections were stained with Safranin and FastGreen to check tissue quality, and egg apparatus were identified under a microscope (Figure 3.1). Unstained sections were used for LCM.

Laser Capture Microdissection and Egg Apparatus Collection

A Leica LMD6000 laser microdissection system was used to capture the egg apparatus from tissue sections prepared as above. Before loading PCR tubes with cap into the collection device, 8 µl of RNAlater was pipetted into the lid of the cap. Microdissections were then performed by drawing a line or circle around the egg apparatus along which the laser beam cut (Figure 3.2), with the laser setting as follows: Power 45, Aperture 4, Speed 9 and Specimen balance 15. After the LCM, the egg apparatus sections fell into the lid by gravity, were preserved by the RNAlater, and stored at -80 °C until use. Around 1000 egg apparatus sections per biological replicate (4 B-2s and 4 B-12-9) were collected.

RNA extraction, cDNA Synthesis, Amplification, Library Construction and Sequencing

Egg apparatus RNA was extracted using RNeasy Plus Micro Kit (QIAGEN) following the manufacturer's protocol with minor modifications. The quantity and quality of RNA were checked using the Qubit 2.0 Fluorometer RNA assay (Invitrogen) and an Agilent 2100 Bioanalyzer using a RNA 6000 pico kit (Agilent Technologies), respectively. For each of the samples, 2 ng of total RNA was used for cDNA synthesis and amplification using SMART-Seq (Zhu *et al.,* 2001) v4 Ultra Low Input RNA Kit (Clontech) with modifications. These included increasing the elution buffer volume, resuspension time, and constantly mixing the cDNA-bound

beads as well as increasing cDNA amplification cycles to improve the cDNA recovery and yield. The concentration and profile of the amplified cDNA were checked using the Qubit 2.0 Fluorometer ds DNA high sensitivity assay (Invitrogen) and the Fragment AnalyzerTM Automated CE System (Agilent Technologies), respectively. For each sample, 100 ng of cDNA was sheared using a Covaris E220 Evolution with 350 bp insert setting and was used to library construction with the TruSeq DNA Nano LT Kit (Illumina) following the manufacturer's suggested protocol with modifications mainly by doing two-time Enrich DNA Fragments step. Libraries were quality checked with the Qubit 2.0 Fluorometer ds DNA high sensitivity assay (Invitrogen) and Fragment AnalyzerTM Automated CE System (Agilent Technologies), and were sequenced in two separate runs of NextSeq (300 Cycles) PE150 Mid Output flow cell on which two biological replicates of each genotype were pooled. Further details of modified protocols can be found in Supplemental Experimental Procedures (Appendix 3.B).

Data Preprocessing

Raw reads were cleaned by removing adapter sequences, over-represented technical sequences detected by FASTQC (Andrews 2010), and sequences of poor-quality using trimmomatic (Bolger *et al.,* 2014). A few bases from 5' and 3' end of the trimmed reads were further cut using TrimGalore (Krueger 2015) to avoid potential sequence bias. After trimming, comprehensive rRNA removal was done using SortMeRNA (Kopylova *et al.,* 2012).

To further identify and remove potential biological contaminants, we constructed de novo assembly using Trinity (Grabherr *et al.,* 2011; Haas *et al.,* 2013) with cleaned reads from all 8 libraries. Trinity contigs were first annotated by a BLASTN (Camacho *et al.,* 2009) search against the NCBI nt database (e-value cutoff of 1e-10). Using these annotations, a custom non-plant contamination database was constructed by extracting the aligned portion of the subject

sequences from the contigs that hit (< 1e-10) bacterial, fungal and animal sequences stored in the NCBI nt database. The non-plant contamination database was indexed as the reference to clean each library till no significant amount of non-plant hit was seen in the final assembly. Further details of data preprocessing can be found in Supplemental Experimental Procedures (Appendix 3.B). All cleaned reads were used for downstream analyses.

Trinotate Annotation and Gene Ontology Enrichment Analyses

We applied all of the trinity transcripts to Trinotate by doing BLASTX (Camacho *et al.,* 2009) search (e-value = 1e-5) against a comprehensive protein database comprised of the Swiss-Prot (Boeckmann et al., 2005) and UniRef90 (UniProt Consortium, 2015) protein databases. Putative coding regions within each trinity transcript were predicted using TransDecoder (http://transdecoder.github.io), and the predicted coding sequences were further annotated through BLASTP (e-value = 1e-5) search against the comprehensive protein database mentioned above and for protein domains search using hmmer (http://hmmer.org/) and PFam (Finn *et al.,* 2013). SignalP (Petersen *et al.,* 2011) was used to predict the potential signal peptides in the transcripts. All results were integrated by Trinotate, stored in an SQLite database, and then reported as a tab-delimited excel file. The functional enrichment analyses was done on significantly differentially expressed transcripts set against the whole expressed transcripts using Trinotate-assigned GO annotations and GOseq (Young *et al.,* 2012), and enriched GO terms of apomictic up-regulated transcripts were reported as tab-delimited file.

BUSCO completeness analyses

The completeness of the transcriptome assembly was examined by subjecting the Trinity.fasta file to BUSCO (Simão *et al.,* 2015) analyses using the command run_busco -i Trinity.fasta -l liliopsida_odb10 -m tran.

Transcript Abundance Estimation and Differential Expression Analysis

Kallisto (Bray *et al.,* 2016) software was used to quantitate the expression level for transcripts, and DEseq2 (Love *et al.,* 2014) was used to identify significantly differentially expressed transcripts. Further details can be found in Supplemental Experimental Procedures (Appendix 3.B).

Plant transcription factors identification

The potential plant transcription factors in the trinity transcripts were identified by doing BLASTX (Camacho *et al.,* 2009) search (e-value < 1e-10) against a close relative *Setaria italica* transcription factor database downloaded from http://planttfdb.cbi.pku.edu.cn/download.php (Jin *et al.,* 2016).

Results and Discussions

Egg apparatus Identification and Collection

Cenchrus ciliaris B-2s serial sections show a single sexual embryo sac containing one egg apparatus and two polar nuclei in a sexual ovary. In aposporous B-12-9, one egg apparatus and a polar nucleus can be observed in the aposporous embryo sac near the micropylar end of the ovule (Figure 3.1). The egg cell is located at micropylar end of the embryo sac and cannot be visually distinguished from synergids and thus was collectively captured as egg apparatus (Figure 3.2).

De Novo Egg, Ovule, Ovule+Egg Transcriptome Assemblies

Assembling and annotating the de novo assembled *Cenchrus ciliaris* egg apparatus transcriptome is challenging for several reasons. No sequenced and annotated *Cenchrus ciliaris* genome is available to which reads can be aligned to guide assembly and eliminate contaminants. The ultra-low input quantity of egg apparatus RNA requires multiple cycles of

cDNA amplification for the downstream library preparation, which may introduce unknown PCR errors in the assembled contigs resulting in an excess of allelic variants. The compromised quality of LCM-derived RNA as well as a poly-A based cDNA synthesis strategy may result in the loss of detectable reads at the 5'mRNA ends causing annotation ambiguities for those transcripts that are partially sequenced and assembled.

To partially address these challenges, high-quality, abundant ovule RNA was used for cDNA amplification-free library construction. The resulting assembly was used to guide the assembly and annotation of egg apparatus reads based on the hypothesis that egg apparatus, as a biological subset of ovule tissues, should be represented in the whole ovule transcriptome with more complete transcripts than in LCM egg apparatus transcriptome. We therefore constructed de novo egg, ovule and egg+ovule combined assembly using Trinity and examined the statistics, the completeness and the number of unique plant annotations for each assembly (Table 3.1).

As expected, significant longer N50 (1747 vs 433 base) and higher overall alignment (96.1% vs 77.6%) were achieved in the ovule assembly compared with the egg alone, while the combined assembly was slightly less than ovule alone. By subjecting our transcriptome assemblies to BUSCO analysis (Simão *et al.,* 2015), we found the ovule transcriptome by this measure is near-complete while the egg is fragmented and the combined again is slightly less than ovule alone. With a BLASTN (e-value < $1e^{-10}$) search against NCBI nt database, we found the greatest number of unique plant hit descriptions (36,861) in the combined assembly and slightly fewer (33,826) in the ovule assembly. The fewest unique plant hit descriptions (23,049) were found in the egg assembly. These observations indicate that most egg apparatus transcripts are represented in the ovule transcriptome, but the combined assembly not only properly captured the diversity from both individual assemblies but also captured the transcripts that

failed to be assembled and annotated solely by egg or ovule data. For example, *PsASGR-BBML* reads were detected in the apomictic ovule RNA-seq data but no contig was present in the ovule assembly, whereas there was a contig in the combined assembly.

We next examined the number of contigs assembled and annotated in the egg assembly and the number of contigs with detectable egg expression (total count >0 across all egg libraries) in the combined assembly (Table 3.2). With the combined assembly, there was a 9.8% increase in the number of assembled egg contigs and a 26.8% of increase in the number of annotations. These results support our hypothesis that the egg is a biological subset of the ovule and combining reads from ovule improves the assembly and annotation of egg cell sequences in the context of egg+ovule combined assembly. The egg+ovule combined assembly was thus used for the downstream analyses.

Cell-type Specific Expression Patterns

The top 50 most highly abundant egg-expressing contigs accounted for 12.7% of the total and include constitutively expressed transcripts from mitochondrion, chloroplast, and ribosomal protein genes (Table S1), which is consistent with the expected biology. Among 165,998 egg-expressing contigs, 91,730 (55.26%) of them hit the NCBI nt database (e-value < $1e^{-10}$), and 71,631 (78.1%) of them hit a close relative, *Setaria italica* (foxtail millet), consistent with the known evolutionary relationship between these species (Donadío *et al.,* 2009; Chemisquy *et al.,* 2010; Kellogg 2017).

To further check the validity of our transcriptome data, we examined the expression level of egg cell-specific gene *EC1* (*egg cell 1*) (Sprunck *et al.,* 2012), synergid predominant genes *MYB98* (Kasahara *et al.,* 2005) and a previously experimentally verified parthenogenesis gene *ASGR-BBM-like* (Conner *et al.,* 2015). We found three potential *EC1* orthologs (*EC1.2-like*,

EC1.3 and *EC1.4*) that were expressed in all of the apomictic and sexual egg apparatus libraries. The *MYB98* was shown among top 5 most synergid-enriched transcripts in rice synergid data, and its expression level is significantly higher in synergid than in the egg with Log_2 (synergid/egg cell) = 7.23 suggesting *MYB98* is significantly more abundant in synergid than in the egg (Ohnishi *et al.,* 2011). *MYB98-like* (Figure 3.3) was also detected at extremely low abundance, which suggests that we have primarily captured the egg cells with few synergids included (Figure 3.3). *CcASGR-BBM-like* expression was detected in all of the apomictic egg cell libraries, and was completely absent from all of the sexual libraries indicating that we correctly collected apomictic eggs expressing parthenogenesis and sexual eggs without parthenogenesis expression. The expression patterns of *EC1, MYB98-like* and *PsASGR-BBM-like* genes in our data are as expected compared with previously published expression patterns (Ohnishi *et al.,* 2011; Conner *et al.,* 2015) and support the validity of our transcriptome data.

Differentially Expressed Embryonic Transcription Factors Reveal the Major Functional Distinctions between Apomictic and Sexual Eggs

We examined the expression profiles across all samples through principle component analyses, and found excellent correlations among biological replicates as they cluster tightly according to genotypes and tissue types (Figure 3.4). We next performed pairwise comparisons of apomictic egg to sexual egg and defined differentially expressed (DE) genes as those with $log_2FC > 2$, false discovery rate < 0.05 using DESeq2 (Love *et al.,* 2014). We identified 4,625 differentially expressed trinity contigs (Table S2) with 2,571 of those up-regulated in the apomictic egg. We further subjected apomictic up-regulated DEs to gene ontology enrichment analyses and found 175 GO terms over-represented (Table S3) including embryo development (GO: 0009790), reproductive process (GO: 0022414), transcription factor activity, transcription

factor binding (GO: 0000989) etc. To estimate major functional distinctions between apomictic and sexual egg, we subjected trinity contigs to blastx (e-value < 1e-10) search against plantTFDB (Jin *et al.,* 2016) and found 72 potential transcription factors spanning 27 families that were de novo and up-regulated in apomictic egg while 21 potential transcription factors spanning 11 families in sexual egg (Figure 3.6;Table S4). AP2 and WOX are among the 16 TF families that are exclusively de novo or up-regulated in apomictic egg. These gene families have been shown to play major roles in embryogenesis and embryo development (Zuo *et al.,* 2002; El Ouakfaoui *et al.,* 2010).

As predicted, we found an *ASGR-BabyBoom-like* (*ASGR-BBML*) gene de novo expressed in apomictic eggs. Expression of this gene in unfertilized eggs of sexual pearl millet, rice and maize induced embryo formation in the absence of pollination (Conner *et al.,* 2015; Conner *et al.,* 2017). Notably, DE analysis also provided evidence for expression of a potentially novel, previously unidentified *BBM2* gene in unfertilized ovules. As is known, ectopic overexpression of *Brassica BBM* induces somatic embryogenesis in *Arabidopsis* (Boutilier *et al.,* 2002) and the rice *BBM1* transgene, expressed from an egg-cell-specific promoter, induces embryogenesis in rice eggs without fertilization (Khanday *et al.,* 2019) suggesting a potential role of this novel *BBM2* in promoting natural parthenogenesis. The natural *BBM1* expression in rice was detected in zygote at 2.5 h after pollination (HAP) (corresponding to karyogamy) in a male-origin-specific manner, and its ectopic expression in egg cells under egg-cell-specific promoter induced parthenogenesis. The fact that a sexually-derived wild-type *BBM* gene can induce parthenogenesis when expressed in the egg before fertilization suggests that parthenogenesis in apomicts might result from a heterochronic shift in expression of a *BBM* commonly employed by the apomictic and sexual pathway, which is partially supported by the preliminary evidence that

this novel *BBM2* is also present in the sexual buffelgrass genome. If this *BBM2* alone could induce parthenogenesis, then parthenogenesis in apomixis may just be a result of early expression of *BBM2* or early expression of *ASGR-BBML*. If this *BBM2* alone could not induce parthenogenesis, together with the auto-activation nature of BBM leads to the hypothesis that *ASGR-BBML* might activate the common *BBM2* to initiate parthenogenesis in apomictic development. To test this hypothesis, we may want to know the interactions between *ASGR-BBML* and *BBM2* by doing DEX-treatment on a sexual line containing *dd45-ASGR-BBML-GR* construct so we can test if *ASGR-BBML* activates *BBM2* expression (not doable in buffelgrass but maybe in rice) or simply check if *ASGR-BBML* expresses earlier than *BBM2* by doing RT-PCR. If *ASGR-BBML* activates *BBM2*, together with the fact that *ASGR-BBML* is unique to natural apomicts, then *ASGR-BBML* may have a novel function to interact with *BBM2* or other factors to initiate parthenogenesis. A *WUSCHEL-related homeobox 7* is also up-regulated and de novo expressed in apomictic eggs. Since the over-expression of maize *BBM* and *WUSCHEL2* in somatic tissues of sorghum, rice and sugarcane greatly increases their embryogenic potential thereby enhancing transformation frequency (Lowe *et al.,* 2016), it is possible that the de novo-expressed WUSCHEL from apomictic eggs works in concert with *BBM2* to promote parthenogenesis. It is noteworthy that *BBM1* and *WUSCHEL* are also found to be de novo expressed in rice zygote, 2.5 hour after pollination and karyogamy, (Anderson *et al.,* 2017) suggesting a role in promoting zygotic embryogenesis. Besides *BBM* and *WUSCHEL*, we also found an embryonic factor *AINTEGUMENTA-LIKE 5* (*AIL5*) ortholog up-regulated in apomictic eggs. Ectopic expression of *Arabidopsis AIL5* can induce somatic embryo formation in *Arabidopsis* (Tsuwamoto *et al.,* 2010). The identification of embryonic transcription factors up-regulated or de novo expressed in apomictic eggs that previously have been shown to function in

parthenogenesis, somatic embryogenesis, or zygotic embryogenesis, demonstrated the utility of our data and confirmed that embryogenesis potential is one of the major functional distinctions between apomictic and sexual eggs.

ASGR-BBML-mediated transcriptional change in apomictic eggs may confer the core parthenogenesis pathway

Based on prior research that the transcription factor *ASGR-BBML* is the only experimentally verified apomict-derived parthenogenesis gene (Conner *et al.,* 2015; Conner *et al.,* 2017) and *BBM* transcription factors are known to work with other protein complexes to control cell proliferation and somatic embryogenesis (Passarinho *et al.,* 2008; Horstman *et al.,* 2015; Horstman *et al.,* 2017), we hypothesize that the core natural parthenogenesis pathway initiated by *ASGR-BBML* progresses through interactions with other TFs followed by up or down-regulation of their target genes.

We examined candidate BBM-target genes that were directly activated by BBM expression in *Arabidopsis* seedlings and those with DNA sites bound by BBM in *Arabidopsis* somatic embryos through chromatin immunoprecipitation sequencing (ChIP-seq) (Passarinho *et al.,* 2008; Horstman *et al.,* 2015; Horstman *et al.,* 2017), and found considerable overlaps with our differentially expressed genes (Table S5). Specifically, we observed that the parthenogenesis gene *ASGR-BBML* is de novo expressed and at a very low level (average TPM = 3.7) in apomictic eggs, and its potential target gene *BBM2*, is also de novo expressed in apomictic egg with relatively abundant expression (average TPM = 44.6). Based on the observation that *BBM* has a transcriptional autoactivation nature (Horstman *et al.,* 2015), we speculate that the higher abundance of *BBM2* may be caused by *ASGR-BBML* activation and *BBM2* autoactivation. It is possible that part of the parthenogenesis pathway may be initiated by *ASGR-BBML* and proceed

through its activation of the *BBM2* together with *BBM2* autoactivation to further promote cell proliferation and embryogenesis. Furthermore, the *BBM* target gene, *AINTEGUMENTA-LIKE 5* (*AIL5*) that can induce somatic embryo formation in *Arabidopsis* (Tsuwamoto *et al.,* 2010) is also found up-regulated in apomictic egg. It is likely that *BBM2* may activate *AINTEGUMENTA-LIKE 5* (*AIL5*) to control embryogenesis.

In addition, orthologs were found of another *BBM* target gene, *NF-YA9*, that is thought to play a role in inducing embryogenesis when overexpressed in *Arabidopsis* (Mu *et al.,* 2013), Nuclear transcription factor Y subunit A-1 and nuclear transcription factor Y subunit A-10 were up-regulated in apomictic eggs suggesting that *BBM* may also interact with *NF-YA* to control embryogenesis.

Apart from the apomictic egg up-regulated transcripts with similarity to *BBM*-target genes that correspond to genes with proven function to induce embryogenesis, we also found those that may play roles in transcription, signaling, protein-protein interaction and cytoskeleton organization that function in embryogenesis. For example, two members of a well-studied gene family *ACTIN DEPOLYMERIZING FACTOR* that is thought to control actin dynamics (Carlier *et al.,* 1997) and reorganize the actin cytoskeleton under *BBM* activation (Passarinho *et al.,* 2008) were identified. Both actin-depolymerizing factor 2 ($Log_2FC = 6.34$, FDR = 0.03, average TPM in apomictic egg = 40.03) and actin-depolymerizing factor 5 ($Log_2FC = 8.75$, FDR= 0.001, average TPM = 12.35) were up-regulated in apomictic egg suggesting *BBM2* may regulate cytoplasmic cytoskeleton dynamics through activating actin-depolymerizing factors to control cell proliferation.

In all, we find *BBM*-mediated transcriptional change taking place in apomictic eggs and expression evidence that supports the hypothesis that the core natural parthenogenesis pathway is

initiated by *ASGR-BBML*. The pathway progresses through interacting with other embryogenic transcription factors followed by up and down regulation of their target genes eventually leading to embryo formation without fertilization.

The non-*BBM*-target gene, *WUSCHEL*, also was up-regulated in apomictic eggs. *WUSCHEL* is thought to be required in maintaining meristem identity (Laux *et al.,* 1996) and its overexpression can induce organogenesis and somatic embryogenesis in shoot and root tissues (Zuo *et al.,* 2002; Gallois *et al.,* 2004). We found overlapping genes between our DEGs and *WUSCHEL*-target genes (Leibfried *et al.,* 2005) (Table S6) but none of them is known to induce embryogenesis suggesting *WUSCHEL*-mediated transcriptional change may be ancillary to regulating parthenogenesis.

Pathways Beyond Core Parthenogenesis -- A More Dynamic Chromatin State is seen in Apomictic vs Sexual egg as Revealed by Differentially Expressed Chromatin Remodeling Factors

Previous cytological observation showed that 12% of the unpollinated apomictic eggs of B-12-9 will form parthenogenetic embryos two days after anthesis suggesting that the rate of parthenogenesis is somehow governed. Also we noticed from this study that the parthenogenesis gene *ASGR-BBML* is expressed at a low level (average TPM = 3.7) in unpollinated B-12-9 eggs on the day of anthesis. The chromosomal regions associated with apomixis in *Cenchrus ciliaris* are largely heterochromatic (Akiyama *et al.,* 2005), suggestive of suppressed expression. These observations lead us to speculate whether there may be a regulatory role for chromatin states in apomictic eggs and whether chromatin remodelers may specifically contribute to apomixis expression.

Chromatin states are mainly determined via two mechanisms: covalent modifications of

histone and DNA methylation (Vaillant *et al.,* 2007). Major histone modification events can be categorized into acetylation, methylation, phosphorylation and ubiquitination (Strahl *et al.,* 2000), which affect chromatin states by activating or repressing gene expression (Grunstein 1997; Struhl 1998; Ng *et al.,* 2000).

Among DEGs, we see histone deacetylase genes exclusively up-regulated in apomictic eggs while acetyltransferase genes including a histone 3 acetyltransferase gene *increased DNA methylation 1* (Qian *et al.,* 2012) is up-regulated in the sexual egg, suggesting the presence of an increase of histone deacetylation in apomictic eggs as compared to their sexual counterpart. Hypoacetylation is known to be associated with chromatin condensation and transcriptional suppression (Ng *et al.,* 2000), which suggest that there are some genomic regions in apomictic eggs that are hypoacetylated and genes within those regions may be transcriptionally silent.

Another potential role histone deacetylases may play is to serve as a prerequisite for histone methylation to happen in the histone 3 lysine 9 residue (Sobel *et al.,* 1995; Rice *et al.,* 2001) and interact with DNA methyltransferase to repress gene expression (Burgers *et al.,* 2002). We next examined the well-studied histone methylation factors and DNA methylation factors that are differentially expressed, and interestingly, all of them are up-regulated in apomictic eggs. Specifically, two *histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH4 (KRYPTONITE)* orthologs are up-regulated in apomictic eggs, a protein which is thought to methylate histone 3 lysine 9 and is required for the maintenance of cytosine methylation in the CpNpG context and repress retrotransposon activity in *Arabidopsis* (Jackson *et al.,* 2002). Another apomictic up-regulated histone-lysine N-methyltransferase gene, *H3 lysine-9 specific SUVH1*, associated with heterochromatic histone 3 lysine 9 dimethylation, may also play a role in heterochromatin gene silencing in *Arabidopsis* (Naumann *et al.,* 2005). Apart from histone

modification, we also found evidence of DNA methylation that may play a role in determining chromatin states. One of the best understood classes of DNA methylation genes in *Arabidopsis*, *DOMAINS REARRANGED METHYL TRANSFERASE2* (*DRM2*), that is responsible for de novo DNA methylation in all sequence contexts and mediating transgene silencing (Cao *et al.,* 2002), is found up-regulated in apomictic eggs ($log_2FC > 7$, FDR = 0.0077 *).

Hypoacetylation and histone 3 lysine 9 methylation on specific genomic region or heterochromatin may reflect the repressive side of chromatin states in apomictic eggs, which complements the hypothesis that condensed repressive chromatin states of the egg before fertilization may be necessary to acquire totipotency for zygotic embryogenesis (Baroux *et al.,* 2015) in a way that apomictic eggs may also require repression on specific regions of chromatin to gain totipotency before they are competent to initiate parthenogenesis. The upregulation of *DRM2* implied an increasing level of de novo DNA methylation in apomictic eggs with parthenogenetic fate, which is consistent with the fact that embryogenesis is characterized by increased de novo DNA methylation (Jullien *et al.,* 2012). Taken together, the recruitment of histone deacetylase and histone 3 lysine 9 methylase as well as de novo DNA methyltransferse is likely to transcriptionally silence a spectrum of genes that characterize apomictic eggs conferring parthenogenesis fate.

However, Day0 unpollinated apomictic eggs are biologically active in terms of having the potential to form embryos two days after anthesis and are transcriptionally active as indicated by GO enrichment analyses reflecting the active side of chromatin states. As the histone 3 lysine 4 methylation is often positively associated with actively transcribed genes (Martin *et al.,* 2005), we found a COMPASS-like H3K4 histone methylase component *WDR5B* ortholog up-regulated in apomictic eggs that may play a role in promoting homeotic gene expression in *Arabidopsis*

(Jiang *et al.,* 2009). Another H3K4 histone-lysine N-methyltransferase *TRX1* was up-regulated in apomictic eggs and may play a role in activating flowering under LD conditions (Choi *et al.,* 2014). Although the histone-lysine N-methyltransferase *ATXR4* was up-regulated in apomictic eggs, no direct evidence is available to infer its function. In *Arabidopsis ATX1*, *ATX2*, *ATXR3*, and *ATXR7* positively regulate *FLC* via histone 3 lysine 4 methylation (Alvarez-Venegas *et al.,* 2003; Pien *et al.,* 2008; Tamada *et al.,* 2009; Yun *et al.,* 2012) while *ATXR5* and *ATXR6* control the heterochromatin condensation and heterochromatic elements silencing via methylation of H3K27 (Jacob *et al.,* 2009).

Overall, we see more dynamic chromatin states in apomictic eggs indicated by the upregulation of various chromatin remodelers that may transcriptionally activate or suppress a spectrum of genes.

Pathways beyond core Parthenogenesis -- Cell Cycle

The idealized cell cycle is comprised of four successive phases: G1 (Gap 1), S (Synthesis), G2 (Gap 2) and M (Mitosis), through which apomictic egg cells must traverse before it is competent for cell division. To check which cell cycle stage Day 0 apomictic and sexual egg cell are most likely at, we examined the expression profiles of some key cell-cycle regulators.

BREAST CANCER SUSCEPTIBILITY 1 ortholog, *BRCA1*, a key regulator controlling the G2/M checkpoint and involved in regulating the onset of mitosis (Yarden *et al.,* 2002), is expressed in apomictic eggs while barely detectable in the sexual egg suggesting that the apomictic egg might have reached G2/M checkpoint and the sexual egg is at a stage prior to G2. The expression of a key negative regulator of mitosis entry, *wee1-like* protein kinase gene (Russell *et al.,* 1987), in all of the apomictic samples indicates that the apomictic eggs may have reached at most the stage of G2 and arrested prior to mitosis. *Licensing factor MCM 7*, required

for S-phase checkpoint activation, is substantially more expressed in sexual egg with a fold change of sex/apo $\log_2$FC > 5, FDR = 0.016 suggesting sexual eggs might be most likely pre S-phase.

Pathways beyond Core Parthenogenesis -- Auxin and Calcium

Auxin plays a role in virtually every aspect of plant growth and development including early embryogenesis (Friml *et al.,* 2003) and gamete specification in the female gametophyte (Pagnussat *et al.,* 2009). In relation to apomixis, auxin treatment on the inflorescence of apomictic *Poa pratensis* was used to rapidly and reliably test if the egg cell was competent for parthenogenesis showing that exogenously supplied auxin promoted the expression of apomictic parthenogenesis (Matzk 1991).

Among DEGs, we found an auxin receptor *TRANSPORT INHIBITOR RESPONSE 1* (*TIR1*) ortholog (Ruegger *et al.,* 1998; Dharmasiri *et al.,* 2005; Kepinski *et al.,* 2005) up-regulated in apomictic eggs indicating that the apomictic egg may differ from the sexual egg in auxin perception. Interestingly, we found auxin negative regulators *AUX/IAA* (*IAA17*, *IAA2*, *IAA31*) are exclusively up-regulated in apomictic egg. *AUX/IAAs* are thought to interact with SCF^{TIR1} and led to ubiquitination and degradation in the presence of high auxin concentrations, derepressing auxin response factors (ARFs) and allow ARFs to transcriptionally activate or repress downstream auxin responsive genes (Chapman *et al.,* 2009). This may explain that apomictic eggs at this stage are more responsive to auxin treatment, and parthenogenesis expression in the apomictic egg can potentially be triggered through auxin stimulation via interaction with *AUX/IAA*.

Calcium ionophore is also reported to stimulate parthenogenesis in mouse oocyte (Uranga *et al.,* 2002) but calcium alone is not sufficient to induce parthenogenesis in plants. As

for plants, Ca^{2+} changes in egg cells are widely thought to be associated with successful fertilization and egg activation (Digonnet *et al.,* 1997; Antoine *et al.,* 2000; Denninger *et al.,* 2014). In our data, we found intracellular calcium receptor proteins calmodulin (Means *et al.,* 1980) and its effector calcium/calmodulin-dependent protein kinases (Braun *et al.,* 1995) are exclusively up-regulated in apomictic eggs compared to the sexual eggs, suggesting the presence of an internal Ca^{2+} increase and calcium-triggered signaling pathway in apomictic eggs. This may reflect some aspects of apomictic egg activation but its correlation to parthenogenesis is still unknown.

With the differentially expressed genes and their annotations generated in this study, we discussed the potential core pathways conferring natural parthenogenesis and pathways that may play an ancillary role in promoting parthenogenesis based on sequence similarities to previously identified genes and pathways. However, to experimentally test our hypotheses, chromatin immunoprecipitation-seq (ChiP-seq) on ASGR-BBML-bound DNA or BBM2-bound DNA would be essential to study *ASGR-BBML* or *BBM2* target genes or protein-DNA interactions; yeast two hybrid experiments may be needed to experimentally investigate physical interactions between ASGR-BBML and BBM2 or other embryogenic factors identified in this study; bisulfite-sequencing on sexual and apomictic eggs is effective in identifying methylation patterns and could potentially be used to ultimately check the chromatin states difference between sexual and apomictic egg cells. In all, this study has provided numerous sequence and computational information for sexual and apomictic eggs, which is valuable for studying natural parthenogenesis. Among future research goals, functional validation of the novel *BBM2* identified in this study through transformation should be of the top priority.

Figures and Tables

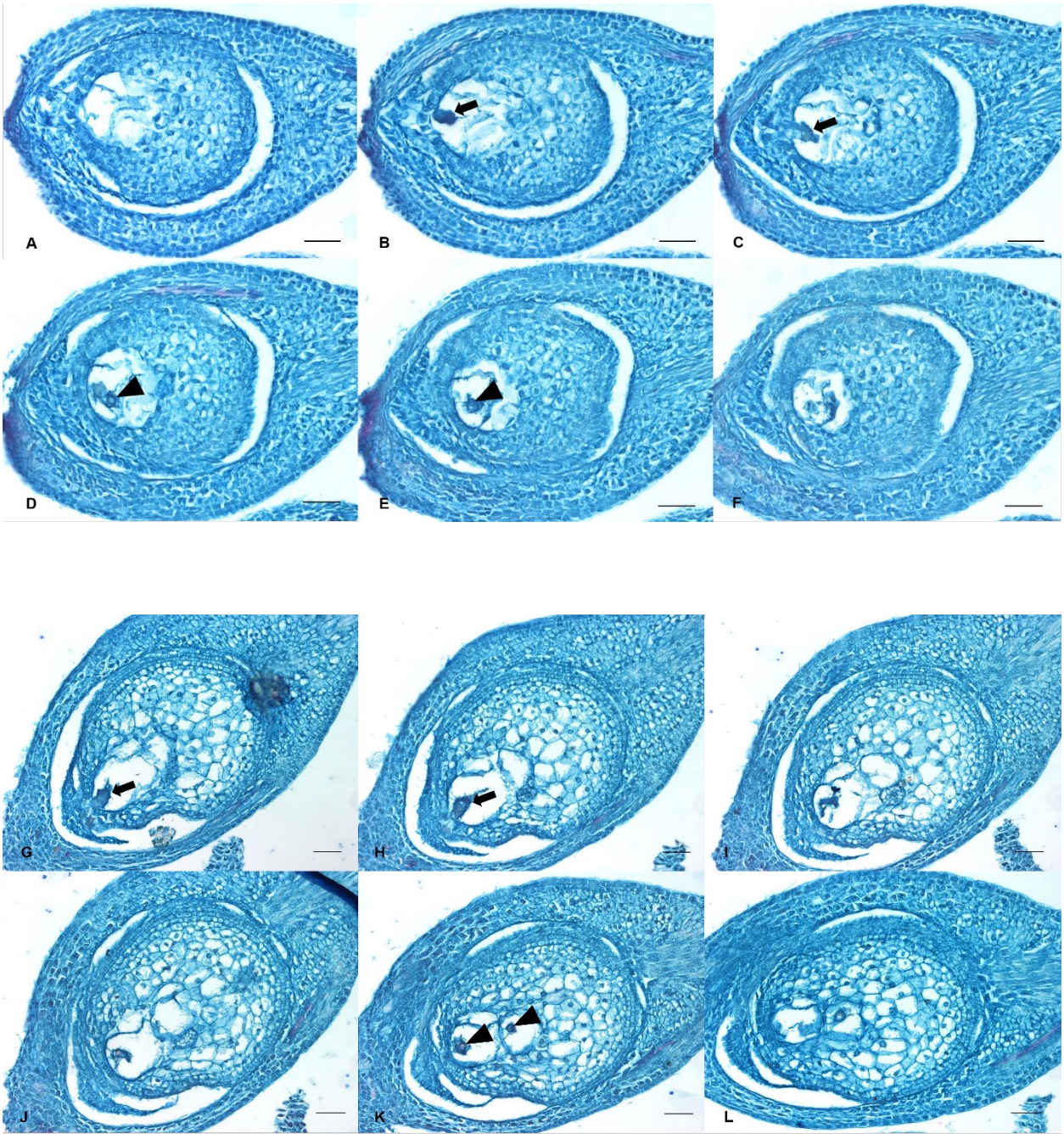

Figure 3.1: Buffelgrass ovary sections (8 µm) stained with Safranin & FastGreen, egg apparatus indicated by arrows, and polar nuclei indicated by triangles, scale bar = 50 µm. (A-F). Serial sections of B-2s showing a single sexual embryo sac containing egg apparatus and polar nucleus.

(G-L). Serial sections of B-12-9 showing egg apparatus in one aposporous embryo sac and polar nucleus in more than one aposporous embryo sac.

Figure 3.2: Laser capture microdissection under a Leica LMD6000 microscope, scale bar = 50 µm. (A & B) Ovary section before & after egg apparatus microdissection. (C) Hundreds of egg apparatus collected in a PCR tube lid.

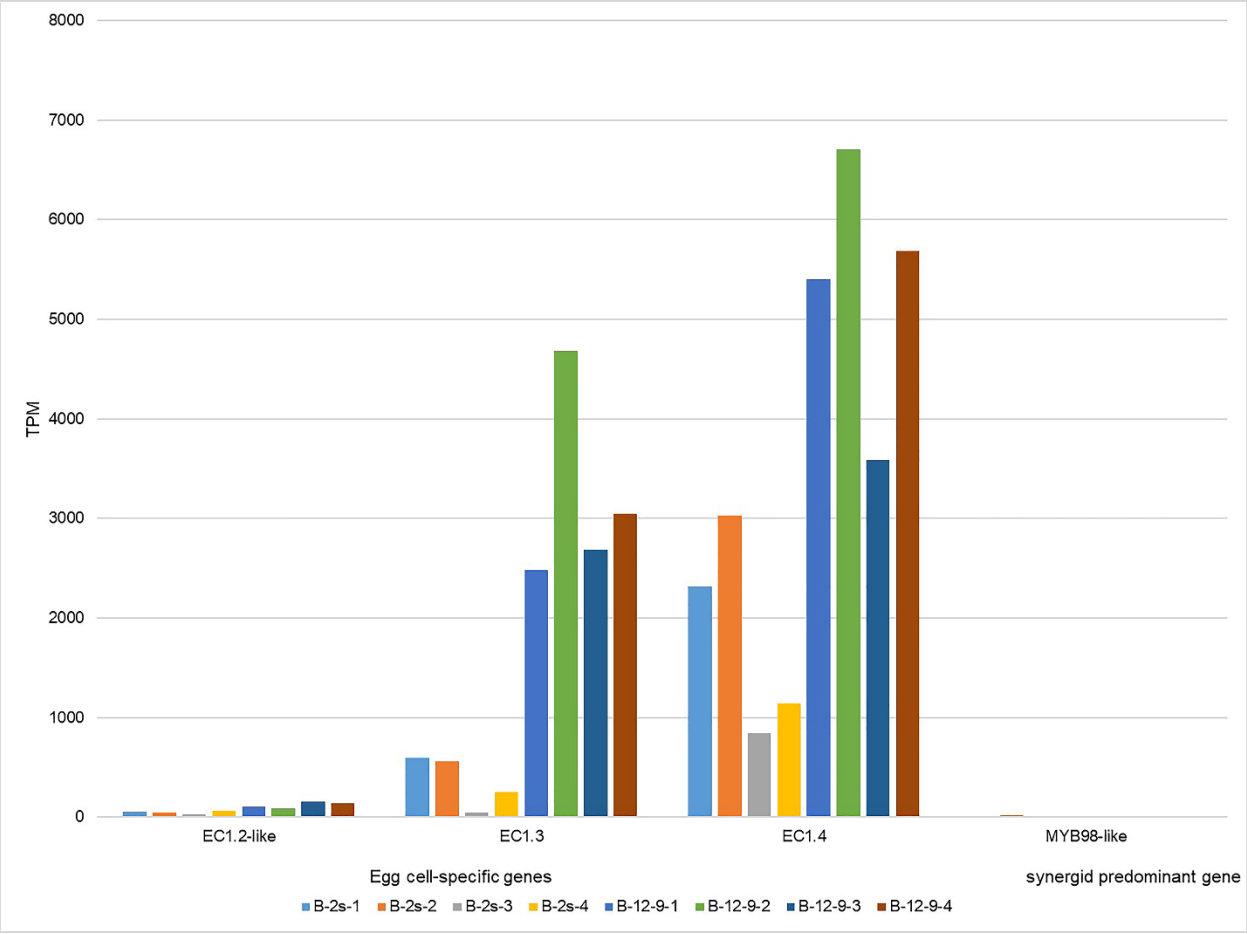

Figure 3.3: Cell type-specific gene expression of sexual and apomictic egg apparatus.

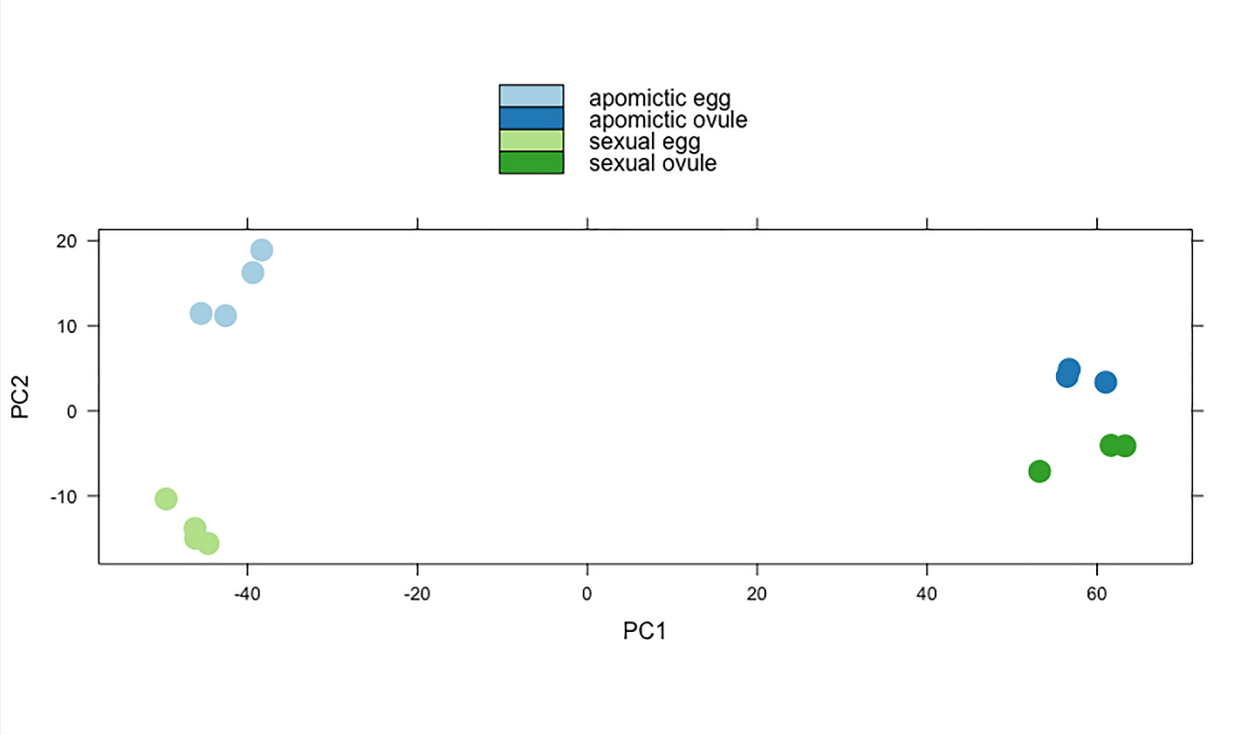

Figure 3.4: PCA plot of expression profile across all samples (PC1 = 0.4137, PC2 = 0.1032).

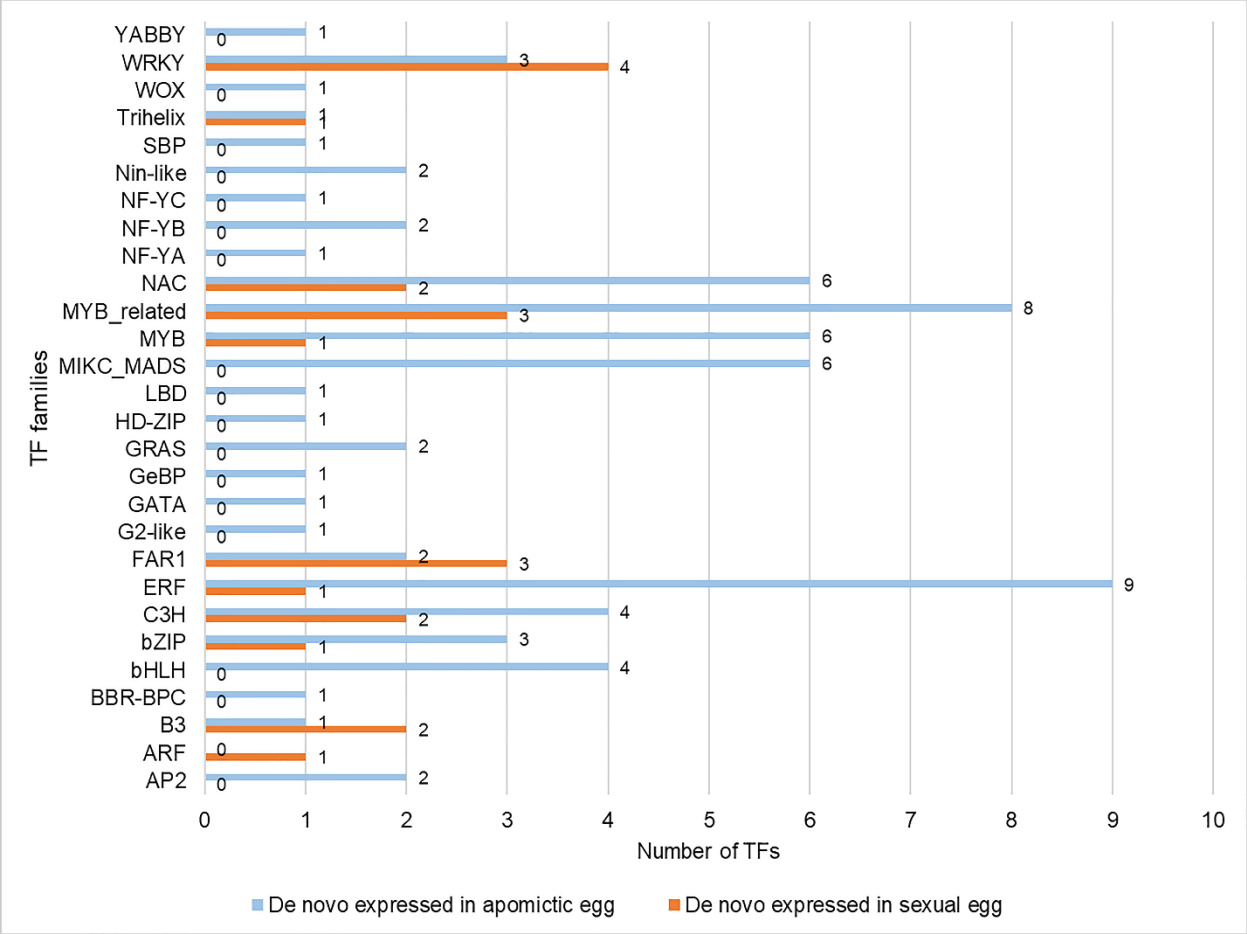

Figure 3.5: De novo and differentially expressed transcription factor families between sexual and apomictic egg.

Table 3.1: Basic statistics, the completeness and the number of unique plant hit descriptions for each assembly.

	De novo egg apparatus assembly	De Novo ovule assembly	Combined assembly
Total Trinity transcripts	151,167	214,836	285,510
Total trinity 'genes[3]':	91,742	87,787	130,601
Contig N50 based on transcripts (bases)	433	1,747	1,527
Contig N50 based on 'genes' (bases)	454	1,595	1,163
Overall alignment rate[4]	77.6%	96.1%	92.9%
# of plant hit descriptions (annotation through NCBI nt database)	70,650	137,724	157,523
# of unique plant hit descriptions	23,049	33,826	36,861
% of complete BUSCOs	10.2%	88.6%	85.1%

[3] Transcripts that were grouped into clusters based on shared sequence content.
[4] The percentage of RNA-seq reads that mapped back to the transcriptome assembly.

% of fragmented BUSCOs	17.8%	6.7%	9.3%
% of missing BUSCOs	72.0%	4.7%	5.6%

Table 3.2: Assembly and annotation statistics for de novo egg assembly and egg cell-expressing transcripts in egg+ovule combined assembly.

	De Novo Egg apparatus assembly	Combined assembly
Number of trinity contigs expressed in egg apparatus	151,167	165,998
With annotation by BLASTN against NCBI nt database	47.87%	55.26%

Reference

Akiyama, Y., W. W. Hanna and P. Ozias-Akins (2005). "High-resolution physical mapping reveals that the apospory-specific genomic region (ASGR) in *Cenchrus ciliaris* is located on a heterochromatic and hemizygous region of a single chromosome." Theoretical and Applied Genetics 111(6): 1042-1051.

Alvarez-Venegas, R., S. Pien, M. Sadder, X. Witmer, U. Grossniklaus and Z. Avramova (2003). "ATX-1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes." Current Biology 13(8): 627-637.

Anderson, S. N., C. S. Johnson, J. Chesnut, D. S. Jones, I. Khanday, M. Woodhouse, C. Li, L. J. Conrad, S. D. Russell and V. Sundaresan (2017). "The zygotic transition is initiated in unicellular plant zygotes with asymmetric activation of parental genomes." Developmental cell 43(3): 349-358.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Antoine, A. F., J. E. Faure, S. Cordeiro, C. Dumas, M. Rougier and J. A. Feijo (2000). "A calcium influx is triggered and propagates in the zygote as a wavefront during in vitro fertilization of flowering plants." Proceedings of the National Academy of Sciences 97(19): 10643-10648.

Baroux, C. and U. Grossniklaus (2015). The maternal-to-zygotic transition in flowering plants: evidence, mechanisms, and plasticity. Current topics in developmental biology, Elsevier. 113: 351-371.

Bicknell, R. A. and A. M. Koltunow (2004). "Understanding apomixis: recent advances and remaining conundrums." The Plant Cell 16(suppl 1): S228-S245.

Boeckmann, B., M.-C. Blatter, L. Famiglietti, U. Hinz, L. Lane, B. Roechert and A. Bairoch (2005). "Protein variety and functional diversity: Swiss-Prot annotation in its biological context." Comptes rendus biologies 328(10-11): 882-899.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics 30(15): 2114-2120.

Boutilier, K., R. Offringa, V. K. Sharma, H. Kieft, T. Ouellet, L. Zhang, J. Hattori, C.-M. Liu, A. A. M. van Lammeren and B. L. A. Miki (2002). "Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth." The Plant Cell 14(8): 1737-1749.

Braun, A. P. and H. Schulman (1995). "The multifunctional calcium/calmodulin-dependent protein kinase: from form to function." Annual review of physiology 57(1): 417-445.

Bray, N. L., H. Pimentel, P. Melsted and L. Pachter (2016). "Near-optimal probabilistic RNA-seq quantification." Nature biotechnology 34(5): 525.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden (2009). "BLAST+: architecture and applications." BMC bioinformatics 10(1): 421.

Cao, X. and S. E. Jacobsen (2002). "Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing." Current Biology 12(13): 1138-1144.

Carlier, M.-F., V. Laurent, J. Santolini, R. Melki, D. Didry, G.-X. Xia, Y. Hong, N.-H. Chua and D. Pantaloni (1997). "Actin depolymerizing factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility." The Journal of cell biology 136(6): 1307-1322.

Chapman, E. J. and M. Estelle (2009). "Mechanism of auxin-regulated gene expression in plants." Annual review of genetics 43: 265-285.

Chemisquy, M. A., L. M. Giussani, M. A. Scataglini, E. A. Kellogg and O. Morrone (2010). "Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): a combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*." Annals of botany 106(1): 107-130.

Choi, S. C., S. Lee, S.-R. Kim, Y.-S. Lee, C. Liu, X. Cao and G. An (2014). "Trithorax group protein *Oryza sativa* Trithorax1 controls flowering time in rice via interaction with early heading date3." Plant physiology 164(3): 1326-1337.

Conner, J. A., S. Goel, G. Gunawan, M.-M. Cordonnier-Pratt, V. E. Johnson, C. Liang, H. Wang, L. H. Pratt, J. E. Mullet and J. DeBarry (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., M. Mookkan, H. Huo, K. Chae and P. Ozias-Akins (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.

Conner, J. A., M. Podio and P. Ozias-Akins (2017). "Haploid embryo production in rice and maize induced by PsASGR-BBML transgenes." Plant reproduction 30(1): 41-52.

Denninger, P., A. Bleckmann, A. Lausser, F. Vogler, T. Ott, D. W. Ehrhardt, W. B. Frommer, S. Sprunck, T. Dresselhaus and G. Grossmann (2014). "Male–female communication triggers calcium signatures during fertilization in Arabidopsis." Nature communications 5: 4645.

Dharmasiri, N., S. Dharmasiri and M. Estelle (2005). "The F-box protein TIR1 is an auxin receptor." Nature 435(7041): 441.

Digonnet, C., D. Aldon, N. Leduc, C. Dumas and M. Rougier (1997). "First evidence of a calcium transient in flowering plants at fertilization." Development 124(15): 2867-2874.

Donadío, S., L. M. Giussani, E. A. Kellogg, F. O. Zuolaga and O. Morrone (2009). "A preliminary molecular phylogeny of *Pennisetum* and *Cenchrus* (Poaceae-Paniceae) based on the trnL-F, rpl16 chloroplast markers." Taxon 58(2): 392-404.

El Ouakfaoui, S., J. Schnell, A. Abdeen, A. Colville, H. Labbé, S. Han, B. Baum, S. Laberge and B. Miki (2010). "Control of somatic embryogenesis and embryo development by AP2 transcription factors." Plant molecular biology 74(4-5): 313-326.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm and J. Mistry (2013). "Pfam: the protein families database." Nucleic acids research 42(D1): D222-D230.

Friml, J., A. Vieten, M. Sauer, D. Weijers, H. Schwarz, T. Hamann, R. Offringa and G. Jürgens (2003). "Efflux-dependent auxin gradients establish the apical–basal axis of *Arabidopsis*." Nature 426(6963): 147.

Gallois, J.-L., F. R. Nora, Y. Mizukami and R. Sablowski (2004). "WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem." Genes & development 18(4): 375-380.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury and Q. Zeng (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nature biotechnology 29(7): 644.

Grunstein, M. (1997). "Histone acetylation in chromatin structure and transcription." Nature 389(6649): 349.

Gualtieri, G., J. A. Conner, D. T. Morishige, L. D. Moore, J. E. Mullet and P. Ozias-Akins (2006). "A segment of the apospory-specific genomic region is highly microsyntenic not only between the apomicts Pennisetum squamulatum and buffelgrass, but also with a rice chromosome 11 centromeric-proximal genomic region." Plant physiology 140(3): 963-971.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li and M. Lieber (2013). "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." Nature protocols 8(8): 1494.

Hojsgaard, D., S. Klatt, R. Baier, J. G. Carman and E. Hörandl (2014). "Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics." Critical Reviews in Plant Sciences 33(5): 414-427.

Horstman, A., H. Fukuoka, J. M. Muino, L. Nitsch, C. Guo, P. Passarinho, G. Sanchez-Perez, R. Immink, G. Angenent and K. Boutilier (2015). "AIL and HDG proteins act antagonistically to control cell proliferation." Development 142(3): 454-464.

Horstman, A., M. Li, I. Heidmann, M. Weemen, B. Chen, J. M. Muino, G. C. Angenent and K. Boutilier (2017). "The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis." Plant physiology 175(2): 848-857.

Jackson, J. P., A. M. Lindroth, X. Cao and S. E. Jacobsen (2002). "Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase." Nature 416(6880): 556.

Jacob, Y., S. Feng, C. A. LeBlanc, Y. V. Bernatavichute, H. Stroud, S. Cokus, L. M. Johnson, M. Pellegrini, S. E. Jacobsen and S. D. Michaels (2009). "ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing." Nature structural & molecular biology 16(7): 763.

Jiang, D., X. Gu and Y. He (2009). "Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in *Arabidopsis*." The Plant Cell 21(6): 1733-1746.

Jin, J., F. Tian, D.-C. Yang, Y.-Q. Meng, L. Kong, J. Luo and G. Gao (2016). "PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants." Nucleic acids research: gkw982.

Jullien, P. E., D. Susaki, R. Yelagandula, T. Higashiyama and F. Berger (2012). "DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*." Current Biology 22(19): 1825-1830.

Kasahara, R. D., M. F. Portereiko, L. Sandaklie-Nikolova, D. S. Rabiger and G. N. Drews (2005). "MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis." The Plant Cell 17(11): 2981-2992.

Kellogg, E. A. (2017). Evolution of *Setaria*. Genetics and Genomics of *Setaria*, Springer: 3-27.

Kepinski, S. and O. Leyser (2005). "The *Arabidopsis* F-box protein TIR1 is an auxin receptor." Nature 435(7041): 446.

Kerk, N. M., T. Ceserani, S. L. Tausta, I. M. Sussex and T. M. Nelson (2003). "Laser capture microdissection of cells from plant tissues." Plant physiology 132(1): 27-35.

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

Kopylova, E., L. Noé and H. Touzet (2012). "SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data." Bioinformatics 28(24): 3211-3217.

Krueger, F. (2015). Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Bioinformatics, Cambridge, United Kingdom.

Laux, T., K. F. Mayer, J. Berger and G. Jurgens (1996). "The WUSCHEL gene is required for shoot and floral meristem integrity in *Arabidopsis*." Development 122(1): 87-96.

Leibfried, A., J. P. C. To, W. Busch, S. Stehling, A. Kehle, M. Demar, J. J. Kieber and J. U. Lohmann (2005). "WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators." Nature 438(7071): 1172.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome biology 15(12): 550.

Lowe, K., E. Wu, N. Wang, G. Hoerster, C. Hastings, M.-J. Cho, C. Scelonge, B. Lenderts, M. Chamberlin and J. Cushatt (2016). "Morphogenic regulators Baby boom and Wuschel improve monocot transformation." The Plant Cell 28(9): 1998-2015.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nature reviews Molecular cell biology 6(11): 838.

Matzk, F. (1991). "New efforts to overcome apomixis in *Poa pratensis* L." Euphytica 55(1): 65-72.

Means, A. R. and J. R. Dedman (1980). "Calmodulin—an intracellular calcium receptor." Nature 285(5760): 73.

Mu, J., H. Tan, S. Hong, Y. Liang and J. Zuo (2013). "*Arabidopsis* transcription factor genes NF-YA1, 5, 6, and 9 play redundant roles in male gametogenesis, embryogenesis, and seed development." Molecular plant 6(1): 188-201.

Naumann, K., A. Fischer, I. Hofmann, V. Krauss, S. Phalke, K. Irmler, G. Hause, A. C. Aurich, R. Dorn and T. Jenuwein (2005). "Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*." The EMBO journal 24(7): 1418-1429.

Ng, H. H. and A. Bird (2000). "Histone deacetylases: silencers for hire." Trends in biochemical sciences 25(3): 121-126.

Nogler, G. A. (1984). Gametophytic apomixis. Embryology of angiosperms, Springer: 475-518.

Ohnishi, T., H. Takanashi, M. Mogi, H. Takahashi, S. Kikuchi, K. Yano, T. Okamoto, M. Fujita, N. Kurata and N. Tsutsumi (2011). "Distinct gene expression profiles in egg and synergid cells of rice as revealed by cell type-specific microarrays." Plant physiology 155(2): 881-891.

Okada, T., Y. Hu, M. R. Tucker, J. M. Taylor, S. D. Johnson, A. Spriggs, T. Tsuchiya, K. Oelkers, J. C. M. Rodrigues and A. M. G. Koltunow (2013). "Enlarging cells initiating apomixis

in *Hieracium praealtum* transition to an embryo sac program prior to entering mitosis." Plant physiology 163(1): 216-231.

Ozias-Akins, P., D. Roche and W. W. Hanna (1998). "Tight clustering and hemizygosity of apomixis-linked molecular markers in Pennisetum squamulatum implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes." Proceedings of the National Academy of Sciences 95(9): 5127-5132.

Ozias-Akins, P. and P. J. van Dijk (2007). "Mendelian genetics of apomixis in plants." Annu Rev Genet 41: 509-537.

Pagnussat, G. C., M. Alandete-Saez, J. L. Bowman and V. Sundaresan (2009). "Auxin-dependent patterning and gamete specification in the *Arabidopsis* female gametophyte." Science 324(5935): 1684-1689.

Passarinho, P., T. Ketelaar, M. Xing, J. van Arkel, C. Maliepaard, M. W. Hendriks, R. Joosen, M. Lammers, L. Herdies and B. Den Boer (2008). "BABY BOOM target genes provide diverse entry points into cell proliferation and cell growth pathways." Plant molecular biology 68(3): 225-237.

Petersen, T. N., S. Brunak, G. Von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nature methods 8(10): 785.

Pien, S., D. Fleury, J. S. Mylne, P. Crevillen, D. Inzé, Z. Avramova, C. Dean and U. Grossniklaus (2008). "ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation." The Plant Cell 20(3): 580-588.

Qian, W., D. Miki, H. Zhang, Y. Liu, X. Zhang, K. Tang, Y. Kan, H. La, X. Li and S. Li (2012). "A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*." Science 336(6087): 1445-1448.

Rice, J. C. and C. D. Allis (2001). "Histone methylation versus histone acetylation: new insights into epigenetic regulation." Current opinion in cell biology 13(3): 263-273.

Roche, D., P. Cong, Z. Chen, W. W. Hanna, D. L. Gustine, R. T. Sherwood and P. Ozias-Akins (1999). "An apospory-specific genomic region is conserved between Buffelgrass (Cenchrus ciliaris L.) and *Pennisetum squamulatum* Fresen." The Plant Journal 19(2): 203-208.

Ruegger, M., E. Dewey, W. M. Gray, L. Hobbie, J. Turner and M. Estelle (1998). "The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p." Genes & development 12(2): 198-207.

Russell, P. and P. Nurse (1987). "Negative regulation of mitosis by wee1+, a gene encoding a protein kinase homolog." Cell 49(4): 559-567.

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

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

Sherwood, R. T., C. C. Berg and B. A. Young (1994). "Inheritance of apospory in buffelgrass." Crop Science 34(6): 1490-1494.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov (2015). "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." Bioinformatics 31(19): 3210-3212.

Sobel, R. E., R. G. Cook, C. A. Perry, A. T. Annunziato and C. D. Allis (1995). "Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4." Proceedings of the National Academy of Sciences 92(4): 1237-1241.

Sprunck, S., S. Rademacher, F. Vogler, J. Gheyselinck, U. Grossniklaus and T. Dresselhaus (2012). "Egg cell–secreted EC1 triggers sperm cell activation during double fertilization." Science 338(6110): 1093-1097.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature 403(6765): 41.

Struhl, K. (1998). "Histone acetylation and transcriptional regulatory mechanisms." Genes & development 12(5): 599-606.

Tamada, Y., J.-Y. Yun, S. chul Woo and R. M. Amasino (2009). "*ARABIDOPSIS* TRITHORAX-RELATED7 is required for methylation of lysine 4 of histone H3 and for transcriptional activation of FLOWERING LOCUS C." The Plant Cell 21(10): 3257-3269.

Tsuwamoto, R., S. Yokoi and Y. Takahata (2010). "*Arabidopsis* EMBRYOMAKER encoding an AP2 domain transcription factor plays a key role in developmental change from vegetative to embryonic phase." Plant molecular biology 73(4-5): 481-492.

Uranga, J. A., R. A. Pedersen and J. Arechaga (2002). "Parthenogenetic activation of mouse oocytes using calcium ionophores and protein kinase C stimulators." International Journal of Developmental Biology 40(2): 515-519.

Vaillant, I. and J. Paszkowski (2007). "Role of histone and DNA methylation in gene regulation." Current opinion in plant biology 10(5): 528-533.

Yarden, R. I., S. Pardo-Reoyo, M. Sgagias, K. H. Cowan and L. C. Brody (2002). "BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage." Nature genetics 30(3): 285.

Young, M. D., M. J. Wakefield, G. K. Smyth and A. Oshlack (2012). "goseq: Gene Ontology testing for RNA-seq datasets." R Bioconductor.

Yun, J.-Y., Y. Tamada, Y. E. Kang and R. M. Amasino (2012). "*Arabidopsis* trithorax-related3/SET domain GROUP2 is required for the winter-annual habit of *Arabidopsis thaliana*." Plant and Cell Physiology 53(5): 834-846.

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

Appendix 3.A: Supplemental Tables

Table S1: Top 50 most-highly abundant transcripts and corresponding annotations

TRINITY_DN47682_c12_g1_i1	*Zea mays* isolate SM10 mitochondrion sequence
TRINITY_DN42072_c11_g2_i2	*Setaria italica* chloroplast, complete genome
TRINITY_DN43100_c18_g1_i1	no hit
TRINITY_DN36404_c5_g1_i1	PREDICTED: *Zea mays* uncharacterized LOC103646773 (LOC103646773), ncRNA
TRINITY_DN29314_c1_g2_i1	PREDICTED: *Setaria italica* uncharacterized LOC101782662 (LOC101782662), mRNA
TRINITY_DN42072_c11_g2_i1	*Setaria italica* chloroplast, complete genome
TRINITY_DN35402_c5_g1_i1	PREDICTED*: Setaria italica* aspartic proteinase CDR1 (LOC101752612), mRNA
TRINITY_DN42072_c11_g1_i1	PREDICTED: *Setaria italica* 50S ribosomal protein L2, chloroplastic (LOC111256063),
TRINITY_DN33600_c6_g1_i2	no hit
TRINITY_DN33526_c6_g1_i1	PREDICTED: *Setaria italica* germin-like protein 5-1 (LOC101759844), mRNA
TRINITY_DN34548_c1_g1_i1	no hit
TRINITY_DN33637_c14_g2_i2	no hit
TRINITY_DN35104_c2_g1_i1	PREDICTED: *Sorghum bicolor* uncharacterized LOC110434297 (LOC110434297), mRNA
TRINITY_DN49551_c3_g3_i2	*Stenotaphrum secundatum* voucher USDA:PI 410357 chloroplast, complete genome
TRINITY_DN36620_c5_g1_i1	*Saccharum* hybrid cultivar R570_BAC clone Sh128J07
TRINITY_DN33543_c6_g1_i1	no hit
TRINITY_DN48372_c2_g3_i2	*Setaria italica* chloroplast, complete genome
TRINITY_DN39567_c7_g2_i1	*Sorghum bicolor* mitochondrion, complete genome
TRINITY_DN49144_c6_g1_i1	*Cenchrus purpureus* chloroplast, complete genome
TRINITY_DN43040_c5_g1_i1	no hit
TRINITY_DN41690_c3_g1_i2	no hit
TRINITY_DN40881_c3_g1_i1	PREDICTED: *Setaria italica* non-specific lipid-transfer protein 1 (LOC101772575), mRNA
TRINITY_DN47884_c3_g2_i2	PREDICTED: *Setaria italica* uncharacterized protein Os08g0218700/LOC_Os08g12160-like (LOC111256211), mRNA
TRINITY_DN32470_c5_g1_i2	no hit
TRINITY_DN32386_c2_g1_i2	no hit
TRINITY_DN31794_c2_g1_i1	no hit

TRINITY_DN35735_c3_g4_i1	PREDICTED: *Setaria italica* cysteine proteinase inhibitor 8 (LOC101754611), mRNA
TRINITY_DN33850_c2_g1_i1	PREDICTED: *Setaria italica* calreticulin (LOC101774614), mRNA
TRINITY_DN41597_c3_g1_i1	no hit
TRINITY_DN33637_c14_g2_i1	no hit
TRINITY_DN34303_c3_g1_i1	*Panicum virgatum* clone PV_ABa103-K10, complete sequence
TRINITY_DN44613_c2_g2_i3	PREDICTED: *Setaria italica* ATP synthase subunit alpha, mitochondrial (LOC111256054), mRNA
TRINITY_DN39567_c7_g1_i1	*Ferrocalamus rimosivaginus* mitochondrion, partial genome
TRINITY_DN42072_c11_g1_i2	*Paspalidium geminatum* voucher SI:Giussani 313 chloroplast, complete genome
TRINITY_DN41963_c4_g1_i1	no hit
TRINITY_DN37619_c3_g1_i1	no hit
TRINITY_DN37636_c1_g1_i1	PREDICTED: *Setaria italica* uncharacterized LOC105914156 (LOC105914156),ncRNA
TRINITY_DN46651_c3_g3_i1	*Pennisetum glaucum* cultivar 7042S hydroxyproline-rich glycoprotein (HRGP) gene, complete cds
TRINITY_DN40186_c3_g1_i1	PREDICTED: *Setaria italica* skin secretory protein xP2 (LOC101760489), mRNA
TRINITY_DN41288_c2_g1_i4	PREDICTED: *Setaria italica* probable glutathione peroxidase 2 (LOC101786937), transcript variant X1, mRNA
TRINITY_DN44574_c5_g1_i2	PREDICTED: *Setaria italica* glycine-rich RNA-binding protein 2-like (LOC101770940), mRNA
TRINITY_DN32691_c0_g1_i4	no hit
TRINITY_DN34303_c3_g1_i2	*Panicum virgatum* clone PV_ABa103-K10, complete sequence
TRINITY_DN41690_c3_g1_i1	no hit
TRINITY_DN32500_c0_g1_i2	no hit
TRINITY_DN33218_c3_g1_i1	PREDICTED: *Setaria italica* glyceraldehyde-3-phosphate dehydrogenase 3, cytosolic (LOC101785151), mRNA
TRINITY_DN46373_c0_g1_i1	no hit
TRINITY_DN36487_c1_g5_i1	no hit
TRINITY_DN42849_c1_g1_i2	PREDICTED: *Zea mays* embryo-sac basal-endosperm-layer embryo-surrounding-region 2 (ebe2), transcript variant X1, mRNA
TRINITY_DN35176_c0_g3_i1	PREDICTED: *Setaria italica* uncharacterized LOC101781532 (LOC101781532),ncRNA

Table S2: Differentially Expressed Genes List, Log2FC>2, FDR< 0.05 by DESeq2 (partial, top 50 with highest fold-change, sexual up-regulated)

TRINITY_DN47008_c0_g1_i15	PREDICTED: *Setaria italica* protein FAR1-RELATED SEQUENCE 5 (LOC101784300), mRNA
TRINITY_DN46260_c0_g1_i3	PREDICTED: *Setaria italica* exonuclease mut-7 homolog (LOC101785032), mRNA
TRINITY_DN49322_c2_g1_i9	*Cenchrus americanus* genomic sequence
TRINITY_DN44518_c1_g3_i5	no hit
TRINITY_DN44207_c4_g1_i7	PREDICTED: *Setaria italica* CASP-like protein 4C1 (LOC101782111), mRNA
TRINITY_DN42223_c0_g1_i7	PREDICTED: *Setaria italica* putative disease resistance RPP13-like protein 3 (LOC101784866), transcript variant X5, mRNA
TRINITY_DN35079_c2_g1_i3	PREDICTED: *Sorghum bicolor* transcription factor MYBS3 (LOC8058075), mRNA
TRINITY_DN49100_c1_g1_i3	PREDICTED: *Setaria italica* ribosome biogenesis protein bms1 (LOC101764870), mRNA
TRINITY_DN43980_c0_g1_i8	no hit
TRINITY_DN27795_c0_g1_i1	no hit
TRINITY_DN43298_c1_g6_i2	no hit
TRINITY_DN41883_c2_g1_i4	no hit
TRINITY_DN26247_c0_g1_i1	no hit
TRINITY_DN40002_c0_g1_i2	PREDICTED: *Setaria italica* acyl-protein thioesterase 1 homolog 1 (LOC101759623), mRNA
TRINITY_DN43484_c0_g1_i1	no hit
TRINITY_DN38726_c0_g1_i2	no hit
TRINITY_DN29090_c0_g1_i1	no hit
TRINITY_DN48473_c0_g2_i1	PREDICTED: *Setaria italica* NAC domain-containing protein 53 (LOC101766001), mRNA
TRINITY_DN33130_c6_g1_i1	*Pennisetum glaucum* BAC 311G2 complete sequence
TRINITY_DN45085_c2_g2_i15	no hit
TRINITY_DN48585_c1_g1_i6	PREDICTED: *Setaria italica* probable sugar phosphate/phosphate translocator At3g11320 (LOC101762303), mRNA

TRINITY_DN46918_c1_g1_i1	PREDICTED: *Setaria italica* splicing factor U2af large subunit A (LOC101757002), transcript variant X2, misc_RNA
TRINITY_DN49566_c3_g1_i3	PREDICTED: *Setaria italica* serine carboxypeptidase 1-like (LOC101764689), transcript variant X4, mRNA
TRINITY_DN39790_c2_g4_i2	no hit
TRINITY_DN42569_c1_g1_i1	PREDICTED: *Setaria italica* serine protease SPPA, chloroplastic (LOC101782735), mRNA
TRINITY_DN48688_c0_g1_i7	PREDICTED: *Setaria italica* uncharacterized LOC101784111 (LOC101784111), transcript variant X5, mRNA
TRINITY_DN44302_c0_g1_i4	PREDICTED: *Setaria italica* probable methyltransferase PMT26 (LOC101766665), mRNA
TRINITY_DN48955_c2_g1_i12	PREDICTED: *Setaria italica* pentatricopeptide repeat-containing protein At1g11290, chloroplastic (LOC101776357), transcript variant X3, mRNA
TRINITY_DN36411_c0_g1_i11	no hit
TRINITY_DN48782_c1_g1_i19	PREDICTED: *Setaria italica* protein argonaute 4B (LOC101778220), mRNA
TRINITY_DN48209_c3_g2_i3	PREDICTED: *Sorghum bicolor* uncharacterized LOC110435160 (LOC110435160), mRNA
TRINITY_DN35133_c0_g1_i3	no hit
TRINITY_DN35678_c1_g2_i3	PREDICTED: *Panicum hallii* universal stress protein PHOS34-like (LOC112894486), mRNA
TRINITY_DN37796_c1_g3_i1	PREDICTED: *Setaria italica* uncharacterized LOC101785997 (LOC101785997), mRNA
TRINITY_DN45963_c0_g2_i1	PREDICTED: *Setaria italica* plant UBX domain-containing protein 8 (LOC101758300), mRNA
TRINITY_DN48160_c0_g1_i5	PREDICTED: *Setaria italica* phenylalanine--tRNA ligase, chloroplastic/mitochondrial (LOC101773070), mRNA
TRINITY_DN42876_c0_g1_i10	PREDICTED: *Setaria italica* BAG-associated GRAM protein 1 (LOC101765090), mRNA
TRINITY_DN46310_c1_g1_i1	PREDICTED: *Setaria italica* endoplasmin homolog (LOC101762306), mRNA
TRINITY_DN39679_c0_g1_i2	PREDICTED: *Setaria italica* uncharacterized LOC101768145 (LOC101768145), transcript variant X2, mRNA

TRINITY_DN47437_c0_g1_i23	PREDICTED: *Setaria italica* uncharacterized LOC101776789 (LOC101776789), transcript variant X3, mRNA
TRINITY_DN31084_c0_g1_i2	no hit
TRINITY_DN42821_c1_g1_i4	PREDICTED: *Setaria italica* uncharacterized LOC101779560 (LOC101779560), misc_RNA
TRINITY_DN26505_c0_g1_i1	PREDICTED: *Diabrotica virgifera* 60S ribosomal protein L23 (LOC114340694), mRNA
TRINITY_DN29517_c0_g1_i2	no hit
TRINITY_DN47561_c1_g1_i13	PREDICTED: *Panicum hallii* uncharacterized LOC112889872 (LOC112889872), transcript variant X4, mRNA
TRINITY_DN15224_c0_g1_i1	no hit
TRINITY_DN49299_c1_g1_i4	PREDICTED: *Setaria italica* increased DNA methylation 1 (LOC101786810), transcript variant X3, mRNA
TRINITY_DN42879_c4_g4_i4	no hit
TRINITY_DN49280_c1_g1_i7	no hit
TRINITY_DN45955_c0_g1_i4	no hit

Differentially Expressed Genes List, Log2FC>2, FDR< 0.05 by DESeq2 (partial, top 50 with highest fold-change, apomictic up-regulated)

TRINITY_DN38788_c1_g1_i6	PREDICTED: *Setaria italica* kynurenine formamidase (LOC101773399), transcript variant X2, mRNA
TRINITY_DN47016_c1_g1_i5	PREDICTED: *Setaria italica* exonuclease mut-7 homolog (LOC101786161), transcript variant X1, mRNA
TRINITY_DN42828_c1_g1_i10	no hit
TRINITY_DN36785_c0_g2_i3	PREDICTED: *Setaria italica* uncharacterized LOC101776722 (LOC101776722), mRNA
TRINITY_DN35105_c0_g1_i1	PREDICTED: *Setaria italica* adenylate kinase isoenzyme 6 homolog (LOC101782268), transcript variant X4, mRNA
TRINITY_DN32387_c1_g1_i3	PREDICTED: *Setaria italica* selenoprotein F (LOC101758956), mRNA
TRINITY_DN47177_c1_g2_i1	PREDICTED: *Setaria italica* dihydrolipoyl dehydrogenase 1, chloroplastic (LOC101769227), mRNA
TRINITY_DN43721_c1_g1_i1	PREDICTED: *Setaria italica* ribonuclease 2 (LOC101765985), mRNA
TRINITY_DN42753_c2_g1_i3	PREDICTED: *Setaria italica* ethylene-responsive transcription factor WIN1 (LOC101762446), mRNA
TRINITY_DN28993_c0_g1_i1	no hit
TRINITY_DN41497_c1_g1_i2	PREDICTED: *Setaria italica* uncharacterized LOC101761322 (LOC101761322), mRNA
TRINITY_DN46834_c1_g2_i5	PREDICTED: *Setaria italica* ornithine carbamoyltransferase, chloroplastic (LOC101785690), mRNA
TRINITY_DN42822_c2_g2_i1	PREDICTED: *Setaria italica* callose synthase 3 (LOC101759475), mRNA
TRINITY_DN38515_c0_g1_i1	PREDICTED: *Setaria italica* transcription factor GTE10 (LOC101775357), transcript variant X2, mRNA
TRINITY_DN42211_c1_g1_i6	PREDICTED: *Setaria italica* ankyrin-1 (LOC101764016), transcript variant X2, mRNA
TRINITY_DN33412_c2_g2_i3	PREDICTED: *Setaria italica* 30S ribosomal protein S16-2, chloroplastic/mitochondrial (LOC101778176), mRNA
TRINITY_DN36477_c3_g2_i4	*Oryza sativa* Indica Group cultivar Shuhui498 chromosome 2 sequence
TRINITY_DN37502_c0_g2_i2	no hit
TRINITY_DN39675_c1_g2_i1	PREDICTED: *Setaria italica* uncharacterized LOC101758951 (LOC101758951), transcript variant X3, ncRNA

TRINITY_DN41421_c1_g1_i4	PREDICTED: *Setaria italica* protein FATTY ACID EXPORT 6 (LOC101770085), transcript variant X2, mRNA
TRINITY_DN37354_c1_g1_i1	*Zea mays* full-length cDNA clone ZM_BFc0120P02 mRNA, complete cds
TRINITY_DN35853_c1_g3_i1	PREDICTED: *Setaria italica* transmembrane protein 205 (LOC101773892), mRNA
TRINITY_DN34936_c0_g1_i3	PREDICTED: *Setaria italica* synaptotagmin-5 (LOC101759646), transcript variant X1, mRNA
TRINITY_DN46791_c0_g1_i7	PREDICTED: *Setaria italica* cysteine-rich and transmembrane domain-containing protein WIH2 (LOC101785914), mRNA
TRINITY_DN49495_c4_g1_i1	PREDICTED: *Setaria italica* dnaJ homolog subfamily C member 17 (LOC101759327), mRNA
TRINITY_DN39656_c3_g1_i2	PREDICTED: *Setaria italica* RING-box protein 1 (LOC101778950), mRNA
TRINITY_DN40089_c6_g2_i3	PREDICTED: *Setaria italica* small nuclear ribonucleoprotein SmD1a (LOC101769681), mRNA
TRINITY_DN38585_c0_g2_i2	PREDICTED: *Setaria italica* glycine-rich RNA-binding protein 4, mitochondrial (LOC101773770), mRNA
TRINITY_DN39083_c0_g1_i8	*Triticum aestivum* cultivar Chinese Spring mitochondrion, complete genome
TRINITY_DN43049_c0_g1_i3	PREDICTED: *Setaria italica* uncharacterized LOC101781567 (LOC101781567), mRNA
TRINITY_DN48754_c0_g1_i1	PREDICTED: *Setaria italica* pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH10 (LOC101752692), transcript variant X1, mRNA
TRINITY_DN48585_c1_g1_i13	PREDICTED: *Setaria italica* probable sugar phosphate/phosphate translocator At3g11320 (LOC101762303), mRNA
TRINITY_DN47061_c0_g1_i2	PREDICTED: *Setaria italica* ubiquitin carboxyl-terminal hydrolase 17-like protein B (LOC101752481), mRNA
TRINITY_DN45288_c0_g1_i5	PREDICTED: *Setaria italica* calcineurin B-like protein 1 (CBL1), transcript variant X1, mRNA
TRINITY_DN39245_c0_g1_i9	PREDICTED: *Setaria italica* uncharacterized LOC101768714 (LOC101768714), transcript variant X1, mRNA
TRINITY_DN40480_c3_g2_i2	PREDICTED: *Setaria italica* eukaryotic translation initiation factor 5A (LOC101769405), mRNA
TRINITY_DN38633_c0_g1_i1	PREDICTED: *Setaria italica* ADP-ribosylation factor 2 (LOC101774077), mRNA

TRINITY_DN36724_c0_g1_i1	PREDICTED: *Setaria italica* gibberellin 20 oxidase 1-D (LOC101760262), mRNA
TRINITY_DN40360_c1_g1_i4	PREDICTED: *Setaria italica* mitochondrial import inner membrane translocase subunit TIM22-3 (LOC101773495), mRNA
TRINITY_DN37134_c2_g5_i1	no hit
TRINITY_DN40315_c0_g1_i1	PREDICTED: *Setaria italica* vacuolar protein sorting-associated protein 25 (LOC101767133), mRNA
TRINITY_DN49035_c2_g1_i8	PREDICTED: *Setaria italica* pre-mRNA-splicing factor 18 (LOC101781549), mRNA
TRINITY_DN42713_c0_g1_i2	PREDICTED: *Setaria italica* gamma-glutamylcyclotransferase 2-2 (LOC101768570), transcript variant X1, mRNA
TRINITY_DN43164_c0_g1_i7	PREDICTED: *Setaria italica* replication factor C subunit 5 (LOC101753640), mRNA
TRINITY_DN40833_c0_g1_i2	PREDICTED: *Setaria italica* repressor of RNA polymerase III transcription MAF1 homolog (LOC101778163), mRNA
TRINITY_DN28814_c0_g2_i3	no hit
TRINITY_DN43750_c1_g1_i5	PREDICTED: *Setaria italica* arogenate dehydrogenase 2, chloroplastic-like (LOC101769617), mRNA
TRINITY_DN47587_c1_g1_i9	*Saccharum* hybrid cultivar BAC clone 265O22, cultivar R570
TRINITY_DN37683_c1_g1_i3	PREDICTED: *Setaria italica* peptidyl-tRNA hydrolase 2, mitochondrial (LOC101770916), mRNA
TRINITY_DN33436_c1_g2_i2	PREDICTED: *Setaria italica* V-type proton ATPase subunit c"2 (LOC101786609), mRNA
TRINITY_DN36592_c3_g1_i3	no hit

Table S3: Enriched GO terms of apomictic upregulated transcripts (FDR < 0.05).

GO:0008150	biological_process
GO:0009987	cellular process
GO:0008152	metabolic process
GO:0071704	organic substance metabolic process
GO:0044237	cellular metabolic process
GO:0006807	nitrogen compound metabolic process
GO:0044238	primary metabolic process
GO:0006412	translation
GO:0043043	peptide biosynthetic process
GO:0043603	cellular amide metabolic process
GO:0034641	cellular nitrogen compound metabolic process
GO:0043604	amide biosynthetic process
GO:0009058	biosynthetic process
GO:0006518	peptide metabolic process
GO:1901576	organic substance biosynthetic process
GO:0044271	cellular nitrogen compound biosynthetic process
GO:0009059	macromolecule biosynthetic process
GO:0044249	cellular biosynthetic process
GO:0043170	macromolecule metabolic process
GO:0034645	cellular macromolecule biosynthetic process
GO:1901564	organonitrogen compound metabolic process
GO:1901566	organonitrogen compound biosynthetic process

GO:0044260	cellular macromolecule metabolic process
GO:0019538	protein metabolic process
GO:0044267	cellular protein metabolic process
GO:0006725	cellular aromatic compound metabolic process
GO:1901360	organic cyclic compound metabolic process
GO:0046483	heterocycle metabolic process
GO:0006139	nucleobase-containing compound metabolic process
GO:0019438	aromatic compound biosynthetic process
GO:1901362	organic cyclic compound biosynthetic process
GO:0090304	nucleic acid metabolic process
GO:0050896	response to stimulus
GO:0034654	nucleobase-containing compound biosynthetic process
GO:0022414	reproductive process
GO:0016070	RNA metabolic process
GO:0009790	embryo development
GO:0032774	RNA biosynthetic process
GO:0009056	catabolic process
GO:0048856	anatomical structure development
GO:1901575	organic substance catabolic process
GO:0009793	embryo development ending in seed dormancy
GO:0042221	response to chemical

GO:0018130	heterocycle biosynthetic process
GO:0050789	regulation of biological process
GO:0003006	developmental process involved in reproduction
GO:0022607	cellular component assembly
GO:0050794	regulation of cellular process
GO:0009066	aspartate family amino acid metabolic process
GO:0065007	biological regulation
GO:0007338	single fertilization
GO:0010468	regulation of gene expression
GO:0006869	lipid transport
GO:0048608	reproductive structure development
GO:0034622	cellular macromolecular complex assembly
GO:0071826	ribonucleoprotein complex subunit organization
GO:0006351	transcription, DNA-templated
GO:0097659	nucleic acid-templated transcription
GO:0071702	organic substance transport
GO:0007275	multicellular organism development
GO:0010035	response to inorganic substance
GO:0032502	developmental process
GO:0043933	macromolecular complex subunit organization
GO:0006950	response to stress
GO:0071840	cellular component organization or biogenesis

GO:0022618 ribonucleoprotein complex assembly

GO:0009566 fertilization

GO:1901700 response to oxygen-containing compound

GO:0048825 cotyledon development

GO:0051252 regulation of RNA metabolic process

GO:0006355 regulation of transcription, DNA-templated

GO:2000112 regulation of cellular macromolecule biosynthetic process

GO:0006790 sulfur compound metabolic process

GO:1903506 regulation of nucleic acid-templated transcription

GO:2001141 regulation of RNA biosynthetic process

GO:0032790 ribosome disassembly

GO:0010556 regulation of macromolecule biosynthetic process

GO:0016043 cellular component organization

GO:0042631 cellular response to water deprivation

GO:0071462 cellular response to water stimulus

GO:0065003 macromolecular complex assembly

GO:0006413 translational initiation

GO:0032501 multicellular organismal process

GO:0048316 seed development

GO:0009067 aspartate family amino acid biosynthetic process

GO:0009889 regulation of biosynthetic process

GO:0060255 regulation of macromolecule metabolic process

GO:0099402	plant organ development
GO:0031326	regulation of cellular biosynthetic process
GO:0000394	RNA splicing, via endonucleolytic cleavage and ligation
GO:1901605	alpha-amino acid metabolic process
GO:0044281	small molecule metabolic process
GO:0006555	methionine metabolic process
GO:0032988	ribonucleoprotein complex disassembly
GO:0006334	nucleosome assembly
GO:0048609	multicellular organismal reproductive process
GO:0001731	formation of translation preinitiation complex
GO:0006226	dUMP biosynthetic process
GO:0009149	pyrimidine nucleoside triphosphate catabolic process
GO:0009213	pyrimidine deoxyribonucleoside triphosphate catabolic process
GO:0009223	pyrimidine deoxyribonucleotide catabolic process
GO:0046078	dUMP metabolic process
GO:0046080	dUTP metabolic process
GO:0046081	dUTP catabolic process
GO:0032984	macromolecular complex disassembly
GO:0019222	regulation of metabolic process
GO:0048827	phyllome development
GO:1901607	alpha-amino acid biosynthetic process
GO:0048366	leaf development

GO:0051171	regulation of nitrogen compound metabolic process
GO:0031668	cellular response to extracellular stimulus
GO:0005575	cellular_component
GO:0044424	intracellular part
GO:0044464	cell part
GO:0043226	organelle
GO:0032991	macromolecular complex
GO:0043229	intracellular organelle
GO:0044444	cytoplasmic part
GO:0044422	organelle part
GO:0044391	ribosomal subunit
GO:0044446	intracellular organelle part
GO:0043227	membrane-bounded organelle
GO:0030529	intracellular ribonucleoprotein complex
GO:1990904	ribonucleoprotein complex
GO:0043231	intracellular membrane-bounded organelle
GO:0044445	cytosolic part
GO:0005840	ribosome
GO:0043232	intracellular non-membrane-bounded organelle
GO:0043228	non-membrane-bounded organelle
GO:0022627	cytosolic small ribosomal subunit
GO:0000786	nucleosome

GO:0005634	nucleus
GO:0044815	DNA packaging complex
GO:0015935	small ribosomal subunit
GO:0032993	protein-DNA complex
GO:0022626	cytosolic ribosome
GO:0015934	large ribosomal subunit
GO:0090575	RNA polymerase II transcription factor complex
GO:0044798	nuclear transcription factor complex
GO:0005576	extracellular region
GO:0043234	protein complex
GO:0005829	cytosol
GO:0022625	cytosolic large ribosomal subunit
GO:0000788	nuclear nucleosome
GO:0005669	transcription factor TFIID complex
GO:0005618	cell wall
GO:0030312	external encapsulating structure
GO:0016020	membrane
GO:0044428	nuclear part
GO:0005667	transcription factor complex
GO:0005773	vacuole
GO:0044427	chromosomal part
GO:0098805	whole membrane

GO:0031090 organelle membrane

GO:1905368 peptidase complex

GO:0005615 extracellular space

GO:0098588 bounding membrane of organelle

GO:0000502 proteasome complex

GO:1905369 endopeptidase complex

GO:0003674 molecular_function

GO:0003735 structural constituent of ribosome

GO:0005198 structural molecule activity

GO:0005488 binding

GO:0016491 oxidoreductase activity

GO:0000988 transcription factor activity, protein binding

GO:0000989 transcription factor activity, transcription factor binding

GO:0003712 transcription cofactor activity

GO:0008289 lipid binding

GO:0003677 DNA binding

GO:0046872 metal ion binding

GO:0043169 cation binding

GO:0003676 nucleic acid binding

GO:0004097 catechol oxidase activity

GO:0003743 translation initiation factor activity

GO:0004170 dUTP diphosphatase activity

Table S4: Potential transcription factor families identified by blastx search (e-value < 1e -10) against plantTFDB (Jin *et al.,* 2016)

	upregulated in sexual egg	TF family
de novo expressed	TRINITY_DN31511_c0_g1_i2	ARF
de novo expressed	TRINITY_DN24549_c0_g1_i1	B3
de novo expressed	TRINITY_DN44062_c3_g2_i6	B3
de novo expressed	TRINITY_DN44499_c2_g2_i3	bZIP
de novo expressed	TRINITY_DN29890_c0_g1_i2	C3H
de novo expressed	TRINITY_DN38556_c3_g2_i5	C3H
de novo expressed	TRINITY_DN46047_c0_g1_i1	ERF
de novo expressed	TRINITY_DN41784_c0_g1_i6	FAR1
de novo expressed	TRINITY_DN47008_c0_g1_i15	FAR1
de novo expressed	TRINITY_DN49423_c0_g1_i3	FAR1
de novo expressed	TRINITY_DN39601_c1_g1_i2	MYB
de novo expressed	TRINITY_DN35079_c2_g1_i3	MYB_related
de novo expressed	TRINITY_DN36376_c1_g4_i1	MYB_related
de novo expressed	TRINITY_DN37652_c5_g2_i4	MYB_related
de novo expressed	TRINITY_DN42223_c0_g1_i7	NAC
de novo expressed	TRINITY_DN48473_c0_g2_i1	NAC
de novo expressed	TRINITY_DN38430_c0_g1_i1	Trihelix
de novo expressed	TRINITY_DN42161_c0_g1_i5	WRKY
de novo expressed	TRINITY_DN46177_c1_g4_i5	WRKY

de novo expressed	TRINITY_DN46488_c0_g1_i7	WRKY
de novo expressed	TRINITY_DN46615_c0_g1_i2	WRKY
not de novo expressed	TRINITY_DN31511_c0_g1_i4	ARF
not de novo expressed	TRINITY_DN36941_c2_g2_i4	ARF
not de novo expressed	TRINITY_DN40829_c0_g1_i3	bHLH
not de novo expressed	TRINITY_DN40161_c0_g1_i8	C2H2
not de novo expressed	TRINITY_DN36018_c0_g1_i1	C3H
not de novo expressed	TRINITY_DN35350_c0_g2_i3	CO-like
not de novo expressed	TRINITY_DN41080_c0_g1_i10	FAR1
not de novo expressed	TRINITY_DN48797_c0_g2_i1	FAR1
not de novo expressed	TRINITY_DN49271_c2_g1_i12	G2-like
not de novo expressed	TRINITY_DN38697_c3_g1_i1	GATA
not de novo expressed	TRINITY_DN39709_c2_g1_i2	MYB_related
not de novo expressed	TRINITY_DN43055_c1_g1_i3	MYB_related
not de novo expressed	TRINITY_DN44007_c1_g1_i1	MYB_related
not de novo expressed	TRINITY_DN44314_c1_g1_i1	MYB_related
not de novo expressed	TRINITY_DN48470_c4_g1_i3	MYB_related
	upregulated in apomictic egg	TF family
de novo expressed	TRINITY_DN28916_c0_g1_i1	AP2
de novo expressed	TRINITY_DN44016_c2_g3_i3	AP2
de novo expressed	TRINITY_DN35229_c0_g3_i4	B3
de novo expressed	TRINITY_DN40962_c0_g3_i3	BBR-BPC
de novo expressed	TRINITY_DN43063_c0_g2_i2	bHLH

de novo expressed	TRINITY_DN43473_c1_g1_i2	bHLH
de novo expressed	TRINITY_DN49067_c0_g2_i1	bHLH
de novo expressed	TRINITY_DN49357_c2_g1_i1	bHLH
de novo expressed	TRINITY_DN40517_c0_g1_i2	bZIP
de novo expressed	TRINITY_DN45451_c2_g1_i2	bZIP
de novo expressed	TRINITY_DN46606_c3_g1_i3	bZIP
de novo expressed	TRINITY_DN38410_c1_g1_i3	C3H
de novo expressed	TRINITY_DN45219_c0_g1_i1	C3H
de novo expressed	TRINITY_DN45219_c1_g1_i1	C3H
de novo expressed	TRINITY_DN46934_c0_g1_i4	C3H
de novo expressed	TRINITY_DN36576_c3_g1_i2	ERF
de novo expressed	TRINITY_DN39096_c2_g1_i2	ERF
de novo expressed	TRINITY_DN40557_c0_g1_i2	ERF
de novo expressed	TRINITY_DN40560_c0_g1_i1	ERF
de novo expressed	TRINITY_DN40591_c1_g2_i1	ERF
de novo expressed	TRINITY_DN42510_c2_g1_i1	ERF
de novo expressed	TRINITY_DN42753_c2_g1_i3	ERF
de novo expressed	TRINITY_DN46047_c0_g3_i2	ERF
de novo expressed	TRINITY_DN49291_c1_g1_i1	ERF
de novo expressed	TRINITY_DN25458_c0_g1_i2	FAR1
de novo expressed	TRINITY_DN49288_c2_g2_i4	FAR1
de novo expressed	TRINITY_DN42473_c1_g1_i9	G2-like
de novo expressed	TRINITY_DN37172_c1_g1_i12	GATA

de novo expressed	TRINITY_DN36553_c0_g1_i3	GeBP
de novo expressed	TRINITY_DN46808_c0_g3_i6	GRAS
de novo expressed	TRINITY_DN48861_c2_g2_i6	GRAS
de novo expressed	TRINITY_DN41420_c2_g2_i1	HD-ZIP
de novo expressed	TRINITY_DN41007_c2_g3_i2	LBD
de novo expressed	TRINITY_DN36230_c2_g1_i1	MIKC_MADS
de novo expressed	TRINITY_DN42267_c0_g1_i1	MIKC_MADS
de novo expressed	TRINITY_DN44570_c2_g1_i4	MIKC_MADS
de novo expressed	TRINITY_DN44631_c4_g3_i1	MIKC_MADS
de novo expressed	TRINITY_DN47421_c0_g1_i7	MIKC_MADS
de novo expressed	TRINITY_DN49341_c0_g1_i10	MIKC_MADS
de novo expressed	TRINITY_DN40103_c1_g1_i1	MYB
de novo expressed	TRINITY_DN46691_c2_g1_i3	MYB
de novo expressed	TRINITY_DN46691_c2_g1_i8	MYB
de novo expressed	TRINITY_DN46962_c2_g1_i1	MYB
de novo expressed	TRINITY_DN46962_c2_g1_i3	MYB
de novo expressed	TRINITY_DN47820_c0_g1_i11	MYB
de novo expressed	TRINITY_DN35154_c1_g1_i3	MYB_related
de novo expressed	TRINITY_DN36509_c3_g1_i1	MYB_related
de novo expressed	TRINITY_DN37609_c0_g1_i2	MYB_related
de novo expressed	TRINITY_DN46897_c0_g1_i8	MYB_related
de novo expressed	TRINITY_DN47546_c1_g1_i1	MYB_related
de novo expressed	TRINITY_DN47833_c0_g1_i6	MYB_related

de novo expressed	TRINITY_DN48936_c2_g1_i1	MYB_related
de novo expressed	TRINITY_DN49157_c0_g1_i2	MYB_related
de novo expressed	TRINITY_DN35415_c2_g4_i2	NAC
de novo expressed	TRINITY_DN35881_c1_g4_i1	NAC
de novo expressed	TRINITY_DN37929_c6_g5_i1	NAC
de novo expressed	TRINITY_DN39490_c0_g1_i2	NAC
de novo expressed	TRINITY_DN44177_c0_g1_i1	NAC
de novo expressed	TRINITY_DN44895_c0_g5_i1	NAC
de novo expressed	TRINITY_DN30545_c0_g1_i1	NF-YA
de novo expressed	TRINITY_DN31238_c0_g1_i1	NF-YB
de novo expressed	TRINITY_DN35537_c5_g1_i3	NF-YB
de novo expressed	TRINITY_DN44949_c3_g1_i16	NF-YC
de novo expressed	TRINITY_DN35364_c0_g1_i6	Nin-like
de novo expressed	TRINITY_DN48897_c1_g1_i14	Nin-like
de novo expressed	TRINITY_DN49485_c3_g1_i1	SBP
de novo expressed	TRINITY_DN38430_c0_g2_i1	Trihelix
de novo expressed	TRINITY_DN42296_c3_g1_i3	WOX
de novo expressed	TRINITY_DN42509_c1_g1_i1	WRKY
de novo expressed	TRINITY_DN49573_c7_g1_i14	WRKY
de novo expressed	TRINITY_DN49582_c8_g2_i3	WRKY
de novo expressed	TRINITY_DN49035_c2_g1_i8	YABBY
not de novo expressed	TRINITY_DN31511_c0_g1_i1	ARF
not de novo expressed	TRINITY_DN35179_c1_g2_i1	ARF

not de novo expressed	TRINITY_DN40665_c6_g1_i7	bHLH
not de novo expressed	TRINITY_DN40776_c2_g1_i3	bHLH
not de novo expressed	TRINITY_DN44717_c0_g2_i1	bHLH
not de novo expressed	TRINITY_DN35420_c2_g1_i1	bZIP
not de novo expressed	TRINITY_DN35797_c1_g1_i3	bZIP
not de novo expressed	TRINITY_DN40221_c1_g1_i1	bZIP
not de novo expressed	TRINITY_DN40875_c0_g1_i1	bZIP
not de novo expressed	TRINITY_DN44991_c0_g1_i2	bZIP
not de novo expressed	TRINITY_DN37516_c2_g2_i6	C2H2
not de novo expressed	TRINITY_DN41852_c1_g1_i2	C2H2
not de novo expressed	TRINITY_DN41928_c2_g3_i1	C2H2
not de novo expressed	TRINITY_DN49421_c4_g1_i10	C2H2
not de novo expressed	TRINITY_DN33609_c4_g1_i1	C3H
not de novo expressed	TRINITY_DN37675_c3_g1_i1	C3H
not de novo expressed	TRINITY_DN37675_c4_g1_i1	C3H
not de novo expressed	TRINITY_DN39442_c0_g3_i2	DBB
not de novo expressed	TRINITY_DN40352_c2_g3_i1	DBB
not de novo expressed	TRINITY_DN38072_c0_g1_i2	Dof
not de novo expressed	TRINITY_DN37205_c1_g3_i4	ERF
not de novo expressed	TRINITY_DN39728_c1_g3_i2	ERF
not de novo expressed	TRINITY_DN47594_c0_g1_i1	FAR1
not de novo expressed	TRINITY_DN49381_c2_g1_i1	FAR1
not de novo expressed	TRINITY_DN42577_c0_g1_i7	GATA

not de novo expressed	TRINITY_DN47963_c2_g1_i3	GRAS
not de novo expressed	TRINITY_DN48144_c1_g1_i2	GRAS
not de novo expressed	TRINITY_DN48144_c1_g1_i7	GRAS
not de novo expressed	TRINITY_DN35287_c0_g3_i4	HB-other
not de novo expressed	TRINITY_DN48026_c0_g1_i2	HB-other
not de novo expressed	TRINITY_DN38799_c0_g2_i2	HD-ZIP
not de novo expressed	TRINITY_DN46791_c1_g4_i3	HD-ZIP
not de novo expressed	TRINITY_DN39540_c0_g1_i2	MIKC_MADS
not de novo expressed	TRINITY_DN44631_c4_g6_i1	MIKC_MADS
not de novo expressed	TRINITY_DN35951_c4_g3_i1	MYB_related
not de novo expressed	TRINITY_DN48057_c1_g1_i4	MYB_related
not de novo expressed	TRINITY_DN34132_c4_g5_i1	NAC
not de novo expressed	TRINITY_DN41613_c1_g3_i2	NAC
not de novo expressed	TRINITY_DN34773_c0_g1_i2	NF-YA
not de novo expressed	TRINITY_DN37549_c1_g1_i6	NF-YB
not de novo expressed	TRINITY_DN40039_c0_g1_i5	NF-YC
not de novo expressed	TRINITY_DN38292_c0_g3_i1	Nin-like
not de novo expressed	TRINITY_DN38430_c0_g1_i2	Trihelix
not de novo expressed	TRINITY_DN47734_c0_g3_i1	Trihelix
not de novo expressed	TRINITY_DN42296_c3_g1_i2	WOX
not de novo expressed	TRINITY_DN35114_c2_g1_i7	WRKY
not de novo expressed	TRINITY_DN48426_c0_g1_i1	WRKY
not de novo expressed	TRINITY_DN42195_c2_g1_i2	ZF-HD

not de novo expressed	TRINITY_DN42195_c2_g2_i5	ZF-HD

Table S5: BBM target genes (Passarinho *et al.,* 2008, Horstman *et al.,* 2015) and potential corresponding orthologs differentially expressed between apomictic egg and sexual egg.

Candidate BBM target genes	Trinity ID of potential orthologs	BLASTN annotation	upregulated in
BABYBOOM (BBM)	TRINITY_DN28916_c0_g1_i1	ASGR-BBM-like2	apomictic egg, de novo
	TRINITY_DN31800_c0_g1_i1	BBM2	apomictic egg, de novo
AINTEGUMENTA-LIKE 5 (AIL5)	TRINITY_DN44016_c2_g3_i4	AIL5	apomictic egg
	TRINITY_DN44016_c2_g3_i3	AIL5	apomictic egg, de novo
NAC domain protein (ANAC094)	TRINITY_DN35881_c1_g4_i1	NAC domain-containing protein 21/22	apomictic egg, de novo
	TRINITY_DN41613_c1_g3_i2	NAC domain-containing protein 67	apomictic egg
	TRINITY_DN34132_c4_g5_i1	NAC domain-containing protein 83-like	apomictic egg
	TRINITY_DN48473_c0_g2_i1	NAC domain-containing protein 53	sexual egg, de novo
	TRINITY_DN38955_c0_g1_i1	NAC domain-containing protein 73	sexual egg
MADS-box protein (PHERES2)	TRINITY_DN36230_c2_g1_i1	MADS-box protein ZMM17-like	apomictic egg, de novo
	TRINITY_DN44570_c2_g1_i4	MADS-box transcription factor 13	apomictic egg, de novo
	TRINITY_DN47421_c0_g1_i7	MADS-box transcription factor 51	apomictic egg, de novo
	TRINITY_DN44631_c4_g3_i1	MADS-box transcription factor 8	apomictic egg, de novo
	TRINITY_DN39540_c0_g1_i2	MADS-box transcription factor 8	apomictic egg
	TRINITY_DN44631_c4_g6_i1	MADS-box transcription factor 8	apomictic egg
AT-hook protein (AHL8)	TRINITY_DN37016_c0_g1_i1	AT-hook motif nuclear-localized protein 25-like	apomictic egg
	TRINITY_DN36346_c1_g2_i1	AT-hook motif nuclear-localized protein 9	apomictic egg
	TRINITY_DN49391_c2_g1_i5	AT-hook motif nuclear-localized protein 9	apomictic egg
Zinc finger protein C2H2-type	TRINITY_DN38864_c1_g1_i2	zinc finger protein ZAT4	sexual egg, de novo
	TRINITY_DN38864_c1_g1_i4	zinc finger protein ZAT4	sexual egg
Basic Helix-Loop-Helix protein (BHLH122)	TRINITY_DN46468_c1_g1_i3	transcription factor bHLH140	apomictic egg
	TRINITY_DN44717_c0_g2_i1	transcription factor bHLH150-like	apomictic egg
	TRINITY_DN40665_c6_g1_i7	transcription factor bHLH48	apomictic egg
	TRINITY_DN40776_c2_g1_i3	transcription factor bHLH93	apomictic egg
	TRINITY_DN49067_c0_g2_i1	transcription factor bHLH95	apomictic egg, de novo

GDSL-like lipase	TRINITY_DN42627_c0_g1_i2	GDSL esterase/lipase At1g28600	apomictic egg, de novo
	TRINITY_DN42651_c0_g1_i2	GDSL esterase/lipase At2g42990	apomictic egg, de novo
	TRINITY_DN34870_c3_g1_i3	GDSL esterase/lipase At5g33370	apomictic egg, de novo
	TRINITY_DN33973_c6_g1_i4	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN34870_c4_g2_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN33653_c1_g1_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN34870_c3_g1_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN33653_c1_g2_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN48164_c2_g1_i1	GDSL esterase/lipase At5g45910	apomictic egg
	TRINITY_DN40224_c1_g1_i4	GDSL esterase/lipase At5g45910	apomictic egg, de novo
	TRINITY_DN40224_c1_g1_i3	GDSL esterase/lipase At5g45910	apomictic egg
	TRINITY_DN33070_c3_g1_i6	GDSL esterase/lipase LTL1	apomictic egg, de novo
	TRINITY_DN34400_c2_g1_i6	GDSL esterase/lipase LTL1	apomictic egg, de novo
	TRINITY_DN41437_c3_g2_i1	GDSL esterase/lipase LTL1	apomictic egg, de novo
	TRINITY_DN47667_c4_g1_i2	GDSL esterase/lipase LTL1	apomictic egg
CCR4-NOT transcription complex protein	TRINITY_DN45483_c0_g1_i7	CCR4-NOT transcription complex subunit 1	apomictic egg
	TRINITY_DN37483_c0_g1_i4	CCR4-NOT transcription complex subunit 11	apomictic egg, de novo
Leucine-rich receptor-like kinase, LRRII Group	TRINITY_DN49157_c0_g1_i2	probable leucine-rich repeat receptor-like protein kinaseAt5g63930	apomictic egg, de novo
	TRINITY_DN48057_c1_g1_i4	leucine-rich repeat receptor-like tyrosine-protein kinase PXC3	apomictic egg
	TRINITY_DN29208_c0_g1_i1	putative leucine-rich repeat receptor-like serine/threonine-protein kinase At2g14440	sexual egg, de novo
	TRINITY_DN34718_c2_g1_i1	putative leucine-rich repeat receptor-like serine/threonine-protein kinase At2g24130	sexual egg
CRINKLY4	TRINITY_DN48633_c0_g1_i5	putative receptor protein kinase CRINKLY4	apomictic egg
Calcium-binding EF hand protein	TRINITY_DN36179_c1_g1_i1	probable calcium-binding protein CML16	apomictic egg
	TRINITY_DN35507_c0_g1_i1	probable calcium-binding protein CML30	apomictic egg, de novo
	TRINITY_DN47473_c3_g1_i8	probable calcium-binding protein CML7	apomictic egg, de novo
	TRINITY_DN47473_c3_g1_i9	probable calcium-binding protein CML7	apomictic egg, de novo
	TRINITY_DN36886_c0_g2_i1	probable calcium-binding protein CML47	apomictic egg, de novo
	TRINITY_DN39099_c3_g3_i1	probable calcium-binding protein CML20	sexual egg
	TRINITY_DN46420_c2_g2_i4	probable calcium-binding protein CML48	sexual egg, de novo

	TRINITY_DN47473_c3_g1_i13	probable calcium-binding protein CML7	sexual egg, de novo
Cysteine proteinase	TRINITY_DN35494_c0_g2_i2	cysteine proteinase 2	apomictic egg
	TRINITY_DN35494_c0_g2_i1	cysteine proteinase 2	sexual egg
Subtilisin-like serine protease, S8 family	TRINITY_DN48571_c1_g1_i1	subtilisin-like protease SBT1.5	apomictic egg
	TRINITY_DN46307_c1_g1_i3	subtilisin-like protease SBT1.6	apomictic egg
	TRINITY_DN45139_c1_g1_i1	subtilisin-like protease SBT1.7	apomictic egg, de novo
	TRINITY_DN41765_c1_g1_i1	subtilisin-like protease SBT1.8	apomictic egg
	TRINITY_DN43858_c1_g1_i16	subtilisin-like protease SBT3.5	sexual egg
	TRINITY_DN47691_c1_g2_i1	subtilisin-like protease SBT5.3	sexual egg
BTB-POZ domain protein, NPH3 family (NRL27)	TRINITY_DN43249_c2_g1_i1	BTB/POZ domain-containing protein At3g49900	apomictic egg, de novo
	TRINITY_DN47250_c0_g2_i4	BTB/POZ domain-containing protein At3g08570	sexual egg, de novo
RING H2 domain protein	TRINITY_DN42731_c2_g2_i2	RING-H2 finger protein ATL48	apomictic egg, de novo
	TRINITY_DN42731_c2_g2_i6	RING-H2 finger protein ATL48	apomictic egg, de novo
	TRINITY_DN40967_c2_g2_i2	RING-H2 finger protein ATL66	apomictic egg, de novo
	TRINITY_DN41341_c2_g1_i1	RING-H2 finger protein ATL79-like	apomictic egg, de novo
	TRINITY_DN40128_c0_g3_i2	RING-H2 finger protein ATL32	sexual egg
	TRINITY_DN47648_c2_g2_i2	RING-H2 finger protein ATL80-like	sexual egg
U-box/armadillo domain protein	TRINITY_DN46263_c1_g2_i3	U-box domain-containing protein 39	apomictic egg
	TRINITY_DN41195_c0_g1_i1	U-box domain-containing protein 34	sexual egg, de novo
	TRINITY_DN43054_c2_g1_i2	U-box domain-containing protein 4	sexual egg
Zinc finger protein, RanBP2-type	TRINITY_DN42104_c2_g1_i7	ranBP2-type zinc finger protein At1g67325	apomictic egg
PH domain-containing protein (PH16)	TRINITY_DN44818_c3_g1_i2	pleckstrin homology domain-containing protein 1	apomictic egg
Pectate lyase (PLL21)	TRINITY_DN39469_c0_g1_i2	probable pectate lyase 4	sexual egg
Lipid-transfer protein (LTP6-like)	TRINITY_DN35293_c0_g2_i1	non-specific lipid-transfer protein	apomictic egg, de novo
	TRINITY_DN35293_c0_g2_i2	non-specific lipid-transfer protein	apomictic egg, de novo
	TRINITY_DN35736_c0_g1_i7	non-specific lipid-transfer protein 1	apomictic egg
	TRINITY_DN39904_c3_g2_i1	non-specific lipid-transfer protein 1	apomictic egg
	TRINITY_DN40881_c3_g1_i3	non-specific lipid-transfer protein 1	apomictic egg, de novo
	TRINITY_DN49134_c1_g1_i1	non-specific lipid-transfer protein 1	apomictic egg
	TRINITY_DN40881_c3_g1_i1	non-specific lipid-transfer protein 1	apomictic egg
	TRINITY_DN38469_c0_g3_i3	non-specific lipid-transfer protein 2P	apomictic egg, de novo
	TRINITY_DN46671_c3_g3_i1	non-specific lipid-transfer protein 2P-like	apomictic egg

	TRINITY_DN46671_c3_g1_i2	non-specific lipid-transfer protein 2P-like	apomictic egg
	TRINITY_DN42182_c0_g3_i3	non-specific lipid-transfer protein 3-like	apomictic egg, de novo
	TRINITY_DN42182_c0_g2_i6	non-specific lipid-transfer protein 3-like	apomictic egg
	TRINITY_DN42182_c0_g2_i5	non-specific lipid-transfer protein 3-like	apomictic egg
	TRINITY_DN42182_c0_g2_i3	non-specific lipid-transfer protein 3-like	apomictic egg
	TRINITY_DN37410_c1_g1_i1	non-specific lipid transfer protein-like 1	sexual egg
	TRINITY_DN42182_c0_g2_i4	non-specific lipid-transfer protein 3-like	sexual egg
	TRINITY_DN42182_c0_g2_i1	non-specific lipid-transfer protein 3-like	sexual egg
	TRINITY_DN36097_c0_g4_i2	lipid transfer-like protein VAS	sexual egg, de novo
	TRINITY_DN36097_c0_g4_i1	lipid transfer-like protein VAS	sexual egg
	TRINITY_DN35387_c0_g1_i1	non-specific lipid transfer protein GPI-anchored 2	sexual egg
	TRINITY_DN37633_c0_g2_i5	glycolipid transfer protein 1	sexual egg
O-methyltransferase	TRINITY_DN25987_c0_g1_i4	acetylserotonin O-methyltransferase 2	apomictic egg, de novo
	TRINITY_DN42444_c3_g3_i1	caffeic acid 3-O-methyltransferase	apomictic egg, de novo
	TRINITY_DN33193_c0_g1_i16	ubiquinone biosynthesis O-methyltransferase, mitochondrial	sexual egg, de novo
Endo-beta-1,4-glucanase	TRINITY_DN44485_c1_g2_i4	endo-1,3;1,4-beta-D-glucanase	sexual egg, de novo
	TRINITY_DN38609_c0_g1_i2	probable endo-1,3(4)-beta-glucanase ARB_01444	sexual egg, de novo
Polygalacturonase	TRINITY_DN45883_c1_g2_i6	exopolygalacturonase	apomictic egg, de novo
	TRINITY_DN42344_c1_g1_i3	polygalacturonase QRT3	apomictic egg
	TRINITY_DN42284_c0_g1_i4	probable polygalacturonase At1g80170	apomictic egg, de novo
	TRINITY_DN44944_c3_g1_i1	probable polygalacturonase At1g80170	apomictic egg, de novo
	TRINITY_DN38822_c0_g3_i1	probable polygalacturonase	sexual egg
Cytochrome P450 (CYP71B22)	TRINITY_DN31978_c1_g1_i3	cytochrome P450 78A5	apomictic egg, de novo
	TRINITY_DN39023_c0_g1_i1	cytochrome P450 78A5	apomictic egg
	TRINITY_DN47043_c2_g1_i4	cytochrome P450 86A4	apomictic egg, de novo
	TRINITY_DN47918_c2_g2_i3	cytochrome P450 89A2	apomictic egg, de novo
	TRINITY_DN39651_c0_g1_i4	cytochrome P450 90D2	apomictic egg
	TRINITY_DN38686_c0_g1_i5	cytochrome P450 78A5	sexual egg
	TRINITY_DN42905_c6_g1_i2	cytochrome P450 78A5	sexual egg
	TRINITY_DN31978_c1_g1_i4	cytochrome P450 78A5	sexual egg
Actin-depolymerizing factor (ADF9)	TRINITY_DN36541_c0_g1_i1	actin-depolymerizing factor 2	apomictic egg
	TRINITY_DN37217_c0_g3_i4	actin-depolymerizing factor 5	apomictic egg, de novo

EARLY LIGHT INDUCED PROTEIN2	TRINITY_DN40968_c1_g4_i1	low molecular mass early light-inducible protein HV60, chloroplastic	apomictic egg, de novo
	TRINITY_DN40968_c1_g2_i3	low molecular mass early light-inducible protein HV90, chloroplastic	apomictic egg, de novo
Short-chain dehydrogenase/reductase	TRINITY_DN43265_c0_g1_i6	short-chain dehydrogenase TIC 32, chloroplastic	apomictic egg
Copper chaperone (ATX1)	TRINITY_DN37961_c0_g2_i2	copper transport protein ATX1	apomictic egg
phosphatidylinositol N-acetylglucosaminyltransferase subunit P-related	TRINITY_DN39206_c0_g6_i2	phosphatidylinositol N-acetylglucosaminyltransferase subunit P	sexual egg
Glutathione S-transferase TAU 20	TRINITY_DN40959_c2_g1_i1	probable glutathione S-transferase GSTU6	apomictic egg, de novo
	TRINITY_DN42514_c1_g1_i1	glutathione S-transferase	apomictic egg, de novo
	TRINITY_DN34925_c1_g2_i4	glutathione S-transferase 3	apomictic egg, de novo
	TRINITY_DN34925_c1_g2_i5	glutathione S-transferase 3	apomictic egg, de novo
Ankyrin repeat family protein	TRINITY_DN31600_c0_g1_i3	ankyrin repeat, PH and SEC7 domain containing protein secG-like	apomictic egg, de novo
	TRINITY_DN37569_c0_g2_i4	putative ankyrin repeat protein RF_0381	apomictic egg
	TRINITY_DN40207_c1_g2_i1	ankyrin repeat domain-containing protein 2A	apomictic egg, de novo
	TRINITY_DN44660_c1_g2_i1	ankyrin repeat domain-containing protein 2A	apomictic egg
	TRINITY_DN34247_c0_g1_i1	ankyrin repeat-containing protein	apomictic egg

Table S6: WUSCHEL target genes (Leibfried *et al.,* 2005) and potential corresponding orthologs differentially expressed between apomictic egg and sexual egg.

	Trinity ID of potential orthologs	BLASTN annotation	upregulated in
Genes positively responsive to WUS induction			
peroxidase 3 (PER3) (P3) / rare cold-inducible protein (RCI3A) (PRC)	TRINITY_DN32850_c1_g1_i3	peroxidase 1	apomictic egg, de novo
	TRINITY_DN40103_c0_g1_i1	peroxidase 31	apomictic egg, de novo
	TRINITY_DN40103_c0_g1_i2	peroxidase 31	apomictic egg
	TRINITY_DN38275_c2_g2_i2	peroxidase A2	sexual egg, de novo
protein phosphatase 2C, putative / PP2C, putative			
	TRINITY_DN34896_c2_g1_i1	probable protein phosphatase 2C 12	apomictic egg
	TRINITY_DN48826_c0_g1_i1	probable protein phosphatase 2C 47	apomictic egg
	TRINITY_DN47057_c1_g1_i3	probable protein phosphatase 2C 48	apomictic egg, de novo
	TRINITY_DN46143_c0_g1_i8	probable protein phosphatase 2C 6	apomictic egg
	TRINITY_DN44106_c0_g2_i2	probable protein phosphatase 2C 64	apomictic egg, de novo
	TRINITY_DN46098_c1_g2_i5	probable protein phosphatase 2C 8	apomictic egg
	TRINITY_DN40946_c1_g2_i2	probable protein phosphatase 2C 44	apomictic egg
2-oxoglutarate-dependent dioxygenase, putative			
	TRINITY_DN35934_c2_g1_i5	gibberellin 2-beta-dioxygenase	apomictic egg, de novo
myb family transcription factor	TRINITY_DN47754_c0_g2_i2	myb-like protein X	apomictic egg, de novo
	TRINITY_DN46691_c2_g1_i3	myb-related protein 340	apomictic egg, de novo
	TRINITY_DN46691_c2_g1_i8	myb-related protein 340	apomictic egg, de novo
	TRINITY_DN46962_c2_g1_i1	myb-related protein 340	apomictic egg, de novo
	TRINITY_DN46962_c2_g1_i3	myb-related protein 340	apomictic egg, de novo
	TRINITY_DN41632_c2_g1_i2	probable transcription factor MYB58	apomictic egg, de novo
	TRINITY_DN41632_c2_g1_i1	probable transcription factor MYB58	apomictic egg
	TRINITY_DN33498_c3_g4_i1	transcription factor MYB3R-1	sexual egg
	TRINITY_DN35079_c2_g1_i3	transcription factor MYBS3	sexual egg, de novo
	TRINITY_DN49185_c1_g1_i10	myb family transcription factor PHL7	sexual egg

UDP-glucoronosyl/UDP-glucosyl transferase family protein			
	TRINITY_DN46794_c0_g2_i5	DIMBOA UDP-glucosyltransferase BX8	apomictic egg
glucose-6-phosphate 1-dehydrogenase, putative / G6PD, putative			
	TRINITY_DN38648_c0_g1_i4	glucose-6-phosphate 1-dehydrogenase, cytoplasmic isoform	apomictic egg, de novo
	TRINITY_DN48365_c1_g1_i7	glucose-6-phosphate 1-dehydrogenase, cytoplasmic isoform	sexual egg
	TRINITY_DN38648_c0_g1_i5	glucose-6-phosphate 1-dehydrogenase, cytoplasmic isoform	sexual egg, de novo
caffeoyl-CoA 3-O-methyltransferase, putative			
	TRINITY_DN42444_c3_g3_i1	caffeic acid 3-O-methyltransferase	apomictic egg, de novo
phosphoglycerate/bisphosphoglycerate mutase family protein			
	TRINITY_DN42437_c1_g1_i4	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	apomictic egg, de novo
	TRINITY_DN42437_c1_g1_i3	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	apomictic egg
WD-40 repeat family protein	TRINITY_DN39671_c0_g1_i4	WD40 domain-containg protein	apomictic egg, de novo
	TRINITY_DN39671_c0_g1_i1	WD40 domain-containg protein	apomictic egg
	TRINITY_DN44195_c1_g2_i5	uncharacterized WD repeat-containing protein C2A9.03	apomictic egg
	TRINITY_DN45984_c2_g4_i8	WD repeat-containing protein 53	sexual egg
purple acid phosphatase, putative	TRINITY_DN45573_c1_g1_i12	probable purple acid phosphatase 20	apomictic egg, de novo
	TRINITY_DN43548_c0_g1_i2	purple acid phosphatase 2	apomictic egg, de novo
	TRINITY_DN43548_c0_g3_i1	purple acid phosphatase 2	apomictic egg
	TRINITY_DN45573_c1_g1_i8	probable purple acid phosphatase 20	sexual egg, de novo
zinc finger (C3HC4-type RING finger) family protein			
	TRINITY_DN42731_c2_g2_i2	RING-H2 finger protein ATL48	apomictic egg, de novo
	TRINITY_DN42731_c2_g2_i6	RING-H2 finger protein ATL48	apomictic egg, de novo
	TRINITY_DN40967_c2_g2_i2	RING-H2 finger protein ATL66	apomictic egg, de novo

	TRINITY_DN41341_c2_g1_i1	RING-H2 finger protein ATL79-like	apomictic egg, de novo
	TRINITY_DN40128_c0_g3_i2	RING-H2 finger protein ATL32	sexual egg
	TRINITY_DN47648_c2_g2_i2	RING-H2 finger protein ATL80-like	sexual egg
expansin, putative (EXP8)	TRINITY_DN41749_c0_g4_i5	expansin-A4	apomictic egg, de novo
	TRINITY_DN41749_c0_g2_i1	expansin-A4	apomictic egg
	TRINITY_DN47217_c0_g1_i1	expansin-B2	sexual egg
	TRINITY_DN39917_c0_g1_i1	expansin-B9	sexual egg
	TRINITY_DN30843_c0_g1_i2	expansin-A31-like	sexual egg, de novo
cold-responsive protein / cold-regulated protein (cor15a)			
	TRINITY_DN35610_c0_g3_i1	cold-regulated 413 inner membrane protein 1, chloroplastic	apomictic egg
GDSL-motif lipase/hydrolase family protein			
	TRINITY_DN42627_c0_g1_i2	GDSL esterase/lipase At1g28600	apomictic egg, de novo
	TRINITY_DN42651_c0_g1_i2	GDSL esterase/lipase At2g42990	apomictic egg, de novo
	TRINITY_DN34870_c3_g1_i3	GDSL esterase/lipase At5g33370	apomictic egg, de novo
	TRINITY_DN33973_c6_g1_i4	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN34870_c4_g2_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN33653_c1_g1_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN34870_c3_g1_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN33653_c1_g2_i1	GDSL esterase/lipase At5g33370	apomictic egg
	TRINITY_DN48164_c2_g1_i1	GDSL esterase/lipase At5g45910	apomictic egg
	TRINITY_DN40224_c1_g1_i4	GDSL esterase/lipase At5g45910	apomictic egg, de novo
	TRINITY_DN40224_c1_g1_i3	GDSL esterase/lipase At5g45910	apomictic egg
	TRINITY_DN33070_c3_g1_i6	GDSL esterase/lipase LTL1	apomictic egg, de novo
	TRINITY_DN34400_c2_g1_i6	GDSL esterase/lipase LTL1	apomictic egg, de novo
	TRINITY_DN41437_c3_g2_i1	GDSL esterase/lipase LTL1	apomictic egg, de novo
	TRINITY_DN47667_c4_g1_i2	GDSL esterase/lipase LTL1	apomictic egg
jacalin lectin family protein	TRINITY_DN37179_c3_g1_i3	jacalin-related lectin 3	apomictic egg
meprin and TRAF homology domain-containing protein / MATH domain containing protein			
	TRINITY_DN35155_c1_g2_i1	BTB/POZ and MATH domain-containing protein 1-like	apomictic egg
	TRINITY_DN35155_c1_g1_i1	BTB/POZ and MATH domain-containing protein 1-like	

			apomictic egg
	TRINITY_DN34790_c0_g1_i5	BTB/POZ and MATH domain-containing protein 1	apomictic egg
	TRINITY_DN35155_c1_g4_i2	BTB/POZ and MATH domain-containing protein 2-like	apomictic egg, de novo
	TRINITY_DN33292_c1_g1_i5	BTB/POZ and MATH domain-containing protein 2	apomictic egg
	TRINITY_DN31595_c0_g1_i1	BTB/POZ and MATH domain-containing protein 2	apomictic egg
	TRINITY_DN24922_c0_g1_i1	BTB/POZ and MATH domain-containing protein 4	apomictic egg, de novo
	TRINITY_DN38769_c0_g2_i1	BTB/POZ and MATH domain-containing protein 5	apomictic egg
cytochrome P450 family protein	TRINITY_DN31978_c1_g1_i3	cytochrome P450 78A5	apomictic egg, de novo
	TRINITY_DN39023_c0_g1_i1	cytochrome P450 78A5	apomictic egg
	TRINITY_DN47043_c2_g1_i4	cytochrome P450 86A4	apomictic egg, de novo
	TRINITY_DN47918_c2_g2_i3	cytochrome P450 89A2	apomictic egg, de novo
	TRINITY_DN39651_c0_g1_i4	cytochrome P450 90D2	apomictic egg
	TRINITY_DN42905_c6_g1_i2	cytochrome P450 78A5	sexual egg
	TRINITY_DN31978_c1_g1_i4	cytochrome P450 78A5	sexual egg
	TRINITY_DN38686_c0_g1_i5	cytochrome P450 78A5	sexual egg
short-chain dehydrogenase/reductase (SDR) family protein			
	TRINITY_DN43265_c0_g1_i6	short-chain dehydrogenase TIC 32, chloroplastic	apomictic egg
NADH dehydrogenase-related	TRINITY_DN45165_c0_g1_i8	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 2	apomictic egg
	TRINITY_DN37000_c0_g1_i3	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	apomictic egg, de novo
	TRINITY_DN43372_c1_g2_i3	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial	apomictic egg, de novo
	TRINITY_DN39566_c0_g2_i2	NAD(P)H dehydrogenase (quinone) FQR1	sexual egg
	TRINITY_DN38920_c2_g3_i4	NAD(P)H-quinone oxidoreductase subunit O, chloroplastic	sexual egg
	TRINITY_DN40807_c0_g3_i2	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10-A	sexual egg
	TRINITY_DN38691_c1_g1_i4	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial	sexual egg, de novo

	TRINITY_DN49176_c0_g1_i1	probable NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	sexual egg
auxin-responsive protein / indoleacetic acid-induced protein 1 (IAA1)			
	TRINITY_DN32631_c0_g1_i2	putative auxin-responsive protein IAA29	apomictic egg
	TRINITY_DN37013_c0_g1_i2	auxin-responsive protein IAA17	apomictic egg, de novo
	TRINITY_DN44579_c0_g3_i1	auxin-responsive protein IAA2	apomictic egg, de novo
	TRINITY_DN39140_c0_g1_i3	auxin-responsive protein IAA31	apomictic egg, de novo
aspartyl protease family protein	TRINITY_DN47860_c1_g1_i1	aspartyl protease family protein 1	apomictic egg
	TRINITY_DN39463_c1_g1_i5	aspartyl protease family protein 1	sexual egg, de novo
CBS domain-containing protein	TRINITY_DN46010_c0_g1_i2	CBS domain-containing protein CBSCBSPB3	apomictic egg, de novo
	TRINITY_DN36395_c1_g1_i2	CBS domain-containing protein CBSX3, mitochondrial	apomictic egg
	TRINITY_DN33294_c0_g1_i1	CBS domain-containing protein CBSX6	apomictic egg
ADP, ATP carrier protein, mitochondrial, putative / ADP/ATP translocase, putative / adenine nucleotide translocator, putative			
	TRINITY_DN40904_c1_g2_i5	probable envelope ADP,ATP carrier protein, chloroplastic	sexual egg, de novo
zinc finger (CCCH-type) family protein			
	TRINITY_DN49354_c4_g2_i2	zinc finger CCCH domain-containing protein 13	apomictic egg, de novo
	TRINITY_DN45219_c0_g1_i1	zinc finger CCCH domain-containing protein 14	apomictic egg, de novo
	TRINITY_DN39315_c2_g1_i3	zinc finger CCCH domain-containing protein 30	apomictic egg
	TRINITY_DN37675_c3_g1_i1	zinc finger CCCH domain-containing protein 33	apomictic egg
	TRINITY_DN42538_c2_g2_i1	zinc finger CCCH domain-containing protein 40	apomictic egg
	TRINITY_DN45219_c1_g1_i1	zinc finger CCCH domain-containing protein 44	apomictic egg, de novo
	TRINITY_DN37675_c4_g1_i1	zinc finger CCCH domain-containing protein 67-like	apomictic egg
	TRINITY_DN36018_c0_g1_i1	zinc finger CCCH domain-containing protein 12	sexual egg
	TRINITY_DN49354_c4_g2_i1	zinc finger CCCH domain-containing protein 13	sexual egg, de novo
	TRINITY_DN38556_c3_g2_i5	zinc finger CCCH domain-containing protein 13	

			sexual egg, de novo
	TRINITY_DN46469_c0_g1_i3	zinc finger CCCH domain-containing protein 16	sexual egg
	TRINITY_DN49165_c1_g1_i3	zinc finger CCCH domain-containing protein 27	sexual egg
	TRINITY_DN29890_c0_g1_i2	zinc finger CCCH domain-containing protein 8	
			sexual egg, de novo
CBL-interacting protein kinase 14 (CIPK14)			
	TRINITY_DN48944_c2_g4_i1	CBL-interacting protein kinase 17	apomictic egg, de novo
	TRINITY_DN42947_c0_g2_i6	CBL-interacting protein kinase 19	apomictic egg, de novo
	TRINITY_DN42947_c0_g1_i2	CBL-interacting protein kinase 6	apomictic egg, de novo
basic helix-loop-helix (bHLH) family protein			
	TRINITY_DN46468_c1_g1_i3	transcription factor bHLH140	apomictic egg
	TRINITY_DN44717_c0_g2_i1	transcription factor bHLH150-like	apomictic egg
	TRINITY_DN40665_c6_g1_i7	transcription factor bHLH48	apomictic egg
	TRINITY_DN40776_c2_g1_i3	transcription factor bHLH93	apomictic egg
	TRINITY_DN49067_c0_g2_i1	transcription factor bHLH95	apomictic egg, de novo
Genes negatively responsive to WUS induction			
LOB domain protein 3 / lateral organ boundaries domain protein 3 (LBD3)			
	TRINITY_DN41007_c2_g3_i2	LOB domain-containing protein 37	apomictic egg, de novo
two-component responsive regulator / response regulator 7 (ARR7)			
	TRINITY_DN44624_c0_g1_i3	two-component response regulator-like PRR1	
			apomictic egg, de novo
	TRINITY_DN49360_c1_g2_i1	two-component response regulator ORR3-like	sexual egg, de novo
BTB/POZ domain-containing protein	TRINITY_DN43249_c2_g1_i1	BTB/POZ domain-containing protein At3g49900	
			apomictic egg, de novo
	TRINITY_DN47250_c0_g2_i4	BTB/POZ domain-containing protein At3g08570	
			sexual egg, de novo
nodulin family protein	TRINITY_DN48026_c0_g1_i2	nodulin homeobox	apomictic egg
	TRINITY_DN48056_c2_g1_i1	early nodulin-like protein 2	sexual egg
phosphoinositide-specific phospholipase C (PLC1)			
	TRINITY_DN43389_c2_g1_i3	phosphoinositide phosphatase SAC8	sexual egg
oxidoreductase, 2OG-Fe(II) oxygenase family protein			
	TRINITY_DN45687_c0_g1_i2	2-oxoglutarate-Fe(II) type oxidoreductase hxnY-like	sexual egg
	TRINITY_DN45911_c1_g1_i10	2-oxoglutarate-Fe(II) type oxidoreductase	sexual egg, de novo

thioredoxin family protein	TRINITY_DN41456_c0_g1_i4	thioredoxin domain-containing protein PLP3B	apomictic egg
	TRINITY_DN40003_c0_g1_i3	thioredoxin H1	apomictic egg, de novo
	TRINITY_DN43401_c0_g1_i14	thioredoxin H1	apomictic egg
	TRINITY_DN40003_c0_g1_i6	thioredoxin H1	apomictic egg
	TRINITY_DN42364_c2_g1_i2	thioredoxin O, mitochondrial	apomictic egg, de novo
	TRINITY_DN34629_c2_g2_i5	thioredoxin-like 3-2, chloroplastic	apomictic egg
	TRINITY_DN38087_c0_g1_i9	ferredoxin-thioredoxin reductase catalytic chain, chloroplastic	sexual egg
	TRINITY_DN37081_c0_g1_i5	thioredoxin-like protein YLS8	sexual egg
disease resistance protein (TIR-NBS-LRR class), putative			
	TRINITY_DN38942_c0_g1_i1	putative disease resistance protein At1g50180	apomictic egg
	TRINITY_DN44177_c0_g1_i1	putative disease resistance protein RGA1	apomictic egg, de novo
	TRINITY_DN49582_c8_g2_i3	putative disease resistance protein RGA3	apomictic egg, de novo
	TRINITY_DN48426_c0_g1_i1	disease resistance protein RPM1	apomictic egg
	TRINITY_DN47298_c0_g1_i18	probable disease resistance protein At1g61300	apomictic egg, de novo
	TRINITY_DN38227_c3_g1_i2	disease resistance protein RGA2	apomictic egg
	TRINITY_DN42509_c1_g1_i1	disease resistance protein RPP13	apomictic egg, de novo
	TRINITY_DN45379_c0_g1_i9	protein ENHANCED DISEASE RESISTANCE 2-like	apomictic egg, de novo
	TRINITY_DN46615_c0_g1_i2	putative disease resistance RPP13-like protein 2	sexual egg, de novo
	TRINITY_DN42223_c0_g1_i7	putative disease resistance RPP13-like protein 3	sexual egg, de novo
	TRINITY_DN39582_c1_g3_i1	putative disease resistance protein RGA4	sexual egg, de novo
	TRINITY_DN46488_c0_g1_i7	putative disease resistance RPP13-like protein 3	sexual egg, de novo
	TRINITY_DN40583_c0_g2_i3	disease resistance protein RPM1	sexual egg

Appendix 3.B: Supplemental Experimental Procedures

Library prep

Egg apparatus sections were collected and stored in the lid of 0.2ml PCR tubes at -80 centigrade:

Sample	# of ovaries embedded	# of egg apparatus sections (8 μm thick)
B-2S-1	500	772
B-2S-2	500	972
B-2S-3	500	1000
B-2S-4	500	1046
B-12-9-1	500	752
B-12-9-2	500	1356
B-12-9-3	500	1074
B-12-9-4	500	1155

RNA extraction using QIAGEN RNeasy Micro Plus Kit

1. Add 10 μl β-mercaptoethanol to 1ml Buffer RLT Plus before use.
2. Add 25 μl of Buffer RLT Plus on the lid of the PCR tube, spin down for 5 seconds in a QuickSpin centrifuge.
3. Add another 25 μl of Buffer RLT Plus on the lid of the PCR tube, spin down for 5 seconds in a QuickSpin centrifuge again.
4. Transfer 50 μl of Buffer RLT Plus and egg apparatus sample mixtures to a 1.5 ml Eppendorf tube.

5. Add 150 µl of Buffer RLT Plus to the PCR tube, spin down for 5 seconds in a QuickSpin centrifuge.
6. Transfer 150 µl Buffer RLT Plus and egg apparatus sample mixtures to the same 1.5ml Eppendorf tube.
7. Add another 150 µl of Buffer RLT Plus to the PCR tube, spin down for 5 seconds in a QuickSpin centrifuge to thoroughly recover egg apparatus residues.
8. Transfer 150 µl of Buffer RLT Plus and egg apparatus sample mixtures to the same 1.5ml Eppendorf tube.
9. With 350 µl of egg apparatus and Buffer RLT Plus mixtures in the 1.5 ml Eppendorf tube, vortex for 30s.
10. Transfer the lysate to a gDNA Eliminator spin column placed in a 2 ml collection tube. Centrifuge for 30s at 12,000 rpm. Discard the column, and save the flow-through.
11. Add 1 volume of 70% ethanol to the flow-through, and mix well and slowly by pipetting. Do not centrifuge. Proceed immediately to step 12.
12. Transfer the sample, including any precipitate that may have formed, to an RNeasy MinElute spin column placed in a 2 ml collection tube. Close the lid, and centrifuge for 30s at 12,000 rpm. Discard the flow-through.
13. Add 700 µl Buffer RW1 to the RNeasy MinElute spin column. Close the lid, and centrifuge for 30s at 12,000 rpm. Discard the flow-through.
14. Add 500 µl Buffer RPE to the RNeasy MinElute spin column. Close the lid, and centrifuge for 30s at 12,000 rpm. Discard the flow-through.

15. Add 500 µl of 80% ethanol to the RNeasy MinElute spin column. Close the lid, and centrifuge for 2 min at 12,000 rpm to wash the spin column membrane. Discard the collection tube with the flow-through.
16. Place the RNeasy MinElute spin column in a new 2 ml collection tube. Open the lid of the spin column, and centrifuge at full speed (14,000 rpm) for 5 min to dry the membrane. Discard the collection tube with the flow-through.
17. Place the RNeasy MinElute spin column in a new 1.5 ml collection tube. Add 14 µl RNase-free water directly to the center of the spin column membrane, close the lid gently and incubate at room temperature for 2 min. Centrifuge for 1 min at full speed (14,000 rpm) to elute the RNA.
18. Check the quality and quantity with Bioanalyzer.

Result:

Sample	# of egg apparatus sections used for RNA extraction	RNA concentration (pg/ µl)	RIN (RNA Integrity Number)
B-2S-1	370	259	7
B-2S-2	972	596	2.4
B-2S-3	1000	302	1.2
B-2S-4	1046	346	1.6
B-12-9-1	752	889	2.2
B-12-9-2	1356	1129	2.2
B-12-9-3	1074	738	2.1
B-12-9-4	1155	1273	2.3

cDNA synthesis using the SMART-Seq v4 Ultra Low Input RNA Kit (Revised)

A. First-Strand cDNA Synthesis

1. Thaw the 5X Ultra Low First-Strand Buffer at room temperature and 10X Lysis Buffer, RNase Inhibitor, 3' SMART-Seq CDS Primer II A (12 uM), SMART-Seq v4 Oligonucleoide (48 uM) on ice. Vortex each reagent gently and spin down briefly.
2. Prepare the 10X Reaction Buffer by mixing the 10X Lysis Buffer with the RNase Inhibitor as follows (scale-up as needed):

   19 ul 10X Lysis Buffer

   1 ul RNase Inhibitor

   20 ul Total Volume

   Mix briefly, then spin down.
3. Transfer 2 ng of total RNA in less than 9.5 ul of total volume to a 0.2-ml RNase-free PCR tube. Bring the volume to 9.5 ul with nuclease-free water. Add 1 ul of 10X Reaction Buffer. Mix the sample gently by pipetting and incubate at room temperature for 5 minutes.
4. After the incubation, place the samples on ice and add 2 ul of the 3' SMART-Seq CDS Primer II A (12 uM). Vortex gently and then spin the tube briefly

   10.5 ul Total RNA in Reaction Buffer

   2 ul 3' SMART-Seq CDS Primer II A (12 uM)

   20 ul Total Volume

5. Incubate the tubes at 72 ^{0}C in a lid-pre-heated thermal cycler for 3 minutes.

6. Prepare the Master Mix, plus 10% of the total reaction mix volume by adding the reagents in order as follows at room temperature:

4 ul	5X Ultra Low First-Strand Buffer
1 ul	SMART-Seq v4 Oligonucleotide (48 uM)
0.5 ul	RNase Inhibitor (40 U/ul)
5.5 ul	Total volume per reaction

7. Place the tubes on ice for 2 minutes right after the 3 minutes of incubation at 72 ^{0}C.
8. Preheat the thermal cycler to 42 ^{0}C.
9. Add 2 ul per reaction, plus 10%, of the SMARTScribe Reverse Transcriptase to the Master Mix just before use.
10. Add 7.5 ul of the Master Mix to each sample. Mix by gently pipetting, and spin briefly.
11. Place the tubes in a thermal cycler with a heated lid (42 ^{0}C). Run the following program:

42 ^{0}C	90 minutes
70 ^{0}C	10 minutes
4 ^{0}C	hold

B. cDNA Amplification by LD PCR

1. Thaw 2X SeqAmp PCR Buffer, PCR Primer II A (12 uM) on ice. Gently vortex and spin down briefly.
2. Prepare PCR Master Mix, plus 10% of the total reaction mix volume by combining the following reagents in the order as shown below:

(Remove the SeqAmp DNA polymerase from the freezer, gently mix the tube by pipetting up and down several times, and add to the Master Mix just before use)

25 ul	2X SeqAmp PCR Buffer
1 ul	PCR Primer II A (12 uM)
1 ul	SeqAmp DNA Polymerase
3 ul	Nuclease-Free water
30 ul	Total volume per reaction

3. Add 30 ul of the PCR Master Mix to each tube containing 20 ul of first-strand cDNA products. Gently vortex and spin down briefly.
4. Place the tubes in the preheated thermal cycler and run the following protocol:

95 ^{0}C	1 minute	
98 ^{0}C	20 seconds	14 cycles
65 ^{0}C	30 seconds	
68 ^{0}C	3 minutes	
72 ^{0}C	10 minutes	
4 ^{0}C	hold	

C. Purification of Amplified cDNA using the Agencourt RNAClean XP

1. Bring the Agencourt RNAClean XP beads to room temperature and vortex well until evenly mixed.
2. Add 1 ul of 10X Lysis Buffer to each PCR product generated from above.
3. Add 50 ul of well-mixed Agencourt RNAClean XP beads to each sample.
4. Mix by pipetting the entire volume up and down 20 times to mix thoroughly.
5. Incubate at room temperature for 8 minutes to let the cDNA bind to the beads.

6. After the incubation, place the samples on a magnetic stand for ~5 minutes or longer, until the liquid looks clear.
7. Remove the supernatant and discard.
8. Keep the samples on the magnetic stand. Add 200 ul of freshly made 80% ethanol to each sample. Wait for 30 seconds and discard the supernatant without disturbing the beads.
9. Repeat the ethanol wash (Step 8) once.
10. Try to remove all the remaining ethanol with a pipette carefully.
11. Let the samples air-dry at room temperature for 2.5-3.5 minutes until the pellet is no longer shiny, but before a crack appears.
12. Once the beads are dry, add 25 ul of Elution Buffer to cover the bead pellet. Remove the samples from the magnetic stand and mix by pipetting for 15 minutes to thoroughly resuspend the beads.
13. Incubate at room temperature for 5 minutes to rehydrate.
14. Place the samples back on the magnetic stand for 1 minute or longer, until the solution is clear.

(Precautions: In the step 11 of Purification of Amplified cDNA using the Agencourt RNAClean XP, make sure to start re-suspending the pellet after 2.5-3.5 min of air-dry since the pellet looks shiny even beyond 4 min air-dry and seems to never crack given longer time. In the step 12 when re-suspending the pellet with elution buffer, the pellet looks abnormally sticky and will be washed off as a whole piece by the elution buffer instead of getting re-suspended. Increasing the elution buffer volume from 17 ul to 25 ul, suspension time from 2 min to 15 min plus constantly mixing the beads with elution buffer using a pipet is critical to improve cDNA recovery and yield).

D. Second-time cDNA Amplification by LD PCR

1. Aliquot 7 ul of each sample from above, bring to the total volume of 20 ul by adding elution buffer.
2. Thaw 2X SeqAmp PCR Buffer, PCR Primer II A (12 uM) on ice. Gently vortex and spin down briefly.
3. Prepare PCR Master Mix, plus 10% of the total reaction mix volume by combining the following reagents in the order as shown below:

(Remove the SeqAmp DNA polymerase from the freezer, gently mix the tube by pipetting up and down several times, and add to the Master Mix just before use)

25 ul	2X SeqAmp PCR Buffer
1 ul	PCR Primer II A (12 uM)
1 ul	SeqAmp DNA Polymerase
3 ul	Nuclease-Free water
30 ul	Total volume per reaction

4. Add 30 ul of the PCR Master Mix to each tube containing 20 ul of cDNA products. Gently vortex and spin down briefly.
5. Place the tubes in the preheated thermal cycler and run the following protocol:

95 ^{0}C	1 minute	
98 ^{0}C	20 seconds	14 cycles
65 ^{0}C	30 seconds	
68 ^{0}C	3 minutes	
72 ^{0}C	10 minutes	

4 ^{0}C hold

E. Purification of Amplified cDNA using the Agencourt RNAClean XP

1. Bring the Agencourt RNAClean XP beads to room temperature and vortex well until evenly mixed.
2. Add 1 ul of 10X Lysis Buffer to each PCR product generated from above.
3. Add 50 ul of well-mixed Agencourt RNAClean XP beads to each sample.
4. Mix by pipetting the entire volume up and down 20 times to mix thoroughly.
5. Incubate at room temperature for 8 minutes to let the cDNA bind to the beads.
6. After the incubation, place the samples on a magnetic stand for ~5 minutes or longer, until the liquid looks clear.
7. Remove the supernatant and discard.
8. Keep the samples on the magnetic stand. Add 200 ul of freshly made 80% ethanol to each sample. Wait for 30 seconds and discard the supernatant without disturbing the beads.
9. Repeat the ethanol wash (Step 8) once.
10. Try to remove all the remaining ethanol with a pipette carefully.
11. Let the samples air-dry at room temperature for 2.5-3.5 minutes until the pellet is no longer shiny, but before a crack appears.
12. Once the beads are dry, add 25 ul of Elution Buffer to cover the bead pellet. Remove the samples from the magnetic stand and mix by pipetting for 15 minutes to thoroughly resuspend the beads.
13. Incubate at room temperature for 5 minutes to rehydrate.
14. Place the samples back on the magnetic stand for 1 minute or longer, until the solution is clear.

15. Transfer clear supernatant containing purified cDNA from each tube into PCR tubes. Store at – 80 ^{0}C.

16. cDNA QC.

(Precautions: In the step 11 of Purification of Amplified cDNA using the Agencourt RNAClean XP, make sure to re-suspend the pellet before it cracks. If you over-dry the pellet, cracks will appear in the pellet and you may lose some of the cDNA yield. In the step 12 when re-suspending the pellet with elution buffer, the pellet looks normal but still hard to be washed off and re-suspended. Increasing the elution buffer volume from 17 ul to 25 ul, suspension time from 2 min to 15 min plus constantly mixing the beads with elution buffer using a pipet can help to improve cDNA recovery and yield).

Sample	cDNA yield
B-2S-1	11.5 ng/µl
B-2S-2	13.6 ng/µl
B-2S-3	12.5 ng/µl
B-2S-4	11.1 ng/µl
B-12-9-1	10.2 ng/µl
B-12-9-2	13.2 ng/µl
B-12-9-3	11.6 ng/µl
B-12-9-4	10.6 ng/ul

TruSeq Nano DNA Library Prep (Low Sample)

Fragment cDNA

1. Aliquot 100 ng of cDNA and add elution buffer in a final volume of 50 µl in Covaris tube.
2. Shear the cDNA using the Covaris E220 Evolution with 350 bq insert setting.

Clean Up Fragmented cDNA

1. Transfer each of the 50 µl sheared cDNA sample to a 96-well PCR plate. Take out Sample Purification Beads and bring it to room temperature, vortex at least 1 minute.
2. Add 80 µl well-mixed Sample Purification Beads to each sample. Gently mix the entire volume up and down at least 10 times.
3. Incubate the PCR plate at room temperature for 5 minutes.
4. Place the plate on a magnetic stand for 8 minutes or until the liquid is clear.
5. Use a pipette to remove and discard 125 µl of clear supernatant from each well.
6. With the plate on the magnetic stand, add 200 µl freshly made 80% ethanol to each well.
7. Incubate the plate at room temperature for 30 seconds, and use a pipette to remove and discard the supernatant without disturbing the beads.
8. Repeat steps 6 and 7.
9. Remove and discard any ethanol residues from each well.
10. With the plate on the magnetic stand, let the cDNA-beads mixture air-dry at room temperature for ~ 2 minutes but before a crack appears.
11. With the plate on the magnetic stand, add 62.5 µl Resuspension Buffer to each well.
12. Remove the plate from the magnetic stand.

13. Re-suspend the beads in each well by repeatedly dispensing the Resuspension Buffer over the beads until they are immersed in the solution. Do not touch the beads with the pipette tip. Gently pipette the entire volume up and down 10 times to mix thoroughly.
14. Incubate the plate at room temperature for 2 minutes.
15. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
16. Transfer 60 µl of the clear supernatant from each well to a new 96-well plate. Do not disturb the beads.
17. Proceed immediately to Perform End Repair and Size Selection.

Perform End Repair and Size Selection

This process coverts the overhangs resulting from fragmentation into blunt ends using End Repair Mix 2. The 3' to 5' exonuclease activity of this mix removes the 3' overhangs and the 5' to 3' polymerase activity fills in the 5' overhangs. Following end repair, the appropriate library size is selected using different ratios of the Sample Purification Beads.

Preparation:

1. Prepare an ice bucket.
2. Remove the End Repair Mix 2 from – 20 ^{0}C storage. Thaw it at room temperature and then place it on ice.
3. Remove the Sample Purification Beads and Resuspension Buffer from 4 ^{0}C storage and let stand for at least 30 minutes to bring them to room temperature.
4. Preprogram the thermal cycler with the following program:
    - Choose the thermal cycler preheat lid option and set to 100 ^{0}C.
    - 30 ^{0}C for 30 minutes
    - Hold at 4 ^{0}C

End Repair Reaction

1. Centrifuge the thawed End Repair Mix 2 at 600 x g for 5 seconds.
2. Add 40 µl End Repair Mix 2 to each well. Gently pipette the entire volume up and down 10 times to mix thoroughly.
3. Seal the plate with adhesive seal.
4. Return the End Repair Mix 2 tube to -20 ^{0}C storage.
5. Place the plate on the preprogrammed thermal cycler and run the following program:
    - Choose the thermal cycler preheat lid option and set to 100 ^{0}C.
    - 30 ^{0}C for 30 minutes
    - Hold at 4 ^{0}C
6. Remove the plate from the thermal cycler when the program reaches 4 ^{0}C.

Remove Large DNA Fragments

1. Vortex the Sample Purification Beads for at least 1 minute or until they are well dispersed.
2. Combine Sample Purification Beads and PCR grade water in a tube to create a diluted bead mixture of 160 µl per 100 µl of end-repaired sample. Determine the volumes using the following formulas, which include 15% excess for multiple samples:

Diluted Bead Mixture for a 350 bp Insert Size

	Formula	8 samples
Sample Purification Beads	# of samples X 109.25 µl	874 µl
PCR grade water	# of samples X 74.75 µl	598 µl

3. Vortex the diluted bead mixture for 5 seconds to make sure that the beads are evenly dispersed.
4. Add 160 µl of the diluted bead mixture to each well of the plate. Set a 200 µl pipette to 200 µl, and then gently pipette the entire volume up and down 10 times to mix thoroughly.
5. Incubate the plate at room temperature for 5 minutes.
6. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
7. Set a 200 µl pipette to 125 µl. Transfer 125 µl of the supernatant, which contains the DNA of interest, from each well of the plate to a new 96-well plate. Take care not to disturb the beads.

Note: Do not discard the supernatant. It contains the DNA of interest.

8. Repeat step 7. A total of 250 µl of DNA of interest is transferred to a new plate.

Remove Small DNA Fragments

Note: In the following steps, use undiluted Sample Purification Beads.

1. Vortex the Sample Purification Beads for at least 1 minute or until they are well dispersed.
2. Add 30 µl undiluted Sample Purification Beads to each well. Set a 200 µl pipette to 200 µl , and then gently pipette the entire volume up and down 10 times to mix thoroughly.
3. Incubate the plate at room temperature for 5 minutes.
4. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
5. Using a 200 µl pipette set to 138 µl, remove and discard 138 µl of the supernatant from each well. Take care not to disturb the beads.

6. Repeat step 5 to remove and discard a total of 276 µl of the supernatant from each well. Note: Leave the plate on the magnetic stand while performing the following steps 7-12.
7. With the plate on the magnetic stand, add 200 µl freshly prepared 80% ethanol to each well. Do not disturb the beads.
8. Incubate the plate at room temperature for 30 seconds, and then remove and discard all of the supernatant from each well. Take care not to disturb the beads.
9. Repeat steps 7 and 8 to perform an 80% ethanol wash 2 times.
10. Remove and discard any ethanol residues with a 10 µl pipette.
11. With the plate on the magnetic stand, let the samples air-dry at room temperature for 3 minutes. Do not allow the beads to crack.
12. With the plate on the magnetic stand, add 20 µl Resuspension Buffer to each well. Make sure the Resuspension Buffer runs over the beads. Do not touch the beads with the pipette tip.
13. Remove the plate from the magnetic stand.
14. Resuspend the beads in each well by repeatedly dispensing the Resuspension Buffer over the bead pellet until it is immersed in the solution. Gently pipette the entire volume up and down 10 times to mix thoroughly.
15. Incubate the plate at room temperature for 2 minutes.
16. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
17. Transfer 17.5 µl of the clear supernatant from each well to a new 96-well plate.
18. Proceed immediately to Adenylate 3' Ends or safely stop the protocol here and store at -20 ^{0}C for up to 7 days.

Adenylate 3' Ends

A single ‘A’ nucleotide is added to the 3’ ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. A corresponding single ‘T’ nucleotide on the 3’ end of the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera formation.

Preparation:

1. Remove the A-Tailing Mix from -20 ^{0}C storage and thaw it at room temperature.
2. Remove the Resuspension Buffer from 4 ^{0}C storage and bring it to room temperature.
3. Preprogram the thermal cycler with the following program:
    - Choose the preheat lid option and set to 100 ^{0}C
    - 37 ^{0}C for 30 minutes
    - 70 ^{0}C for 5 minutes
    - 4 ^{0}C for 5 minutes
    - Hold at 4 ^{0}C

Adenylate 3’ Ends Reaction

1. Centrifuge the thawed A-Tailing Mix tube at 600 x g for 5 seconds.
2. Add 12.5 µl thawed A-Tailing Mix to each well. Set a 20 µl pipette to 20 µl, then gently pipette the entire volume up and down 10 times to mix thoroughly.
3. Seal the plate with adhesive seal.
4. Return the A-Tailing Mix tube to – 20 ^{0}C storage.
5. Place the plate on the preprogrammed thermal cycler. Close the lid, then select and run the program:
    - Choose the preheat lid option and set to 100 ^{0}C
    - 37 ^{0}C for 30 minutes

- 70 ^{0}C for 5 minutes
- 4 ^{0}C for 5 minutes
- Hold at 4 ^{0}C

6. When the thermal cycler temperature has been at 4 ^{0}C for 5 minutes, remove the plate from the thermal cycler.
7. Proceed immediately to Adapter Ligations.

Ligate Adapters

This process ligates multiple indexing adapters to the ends of the DNA fragments, preparing them for hybridization onto a flow cell.

Preparation:

1. Remove the following from -20 ^{0}C storage and thaw them at room temperature:
    - Appropriate DNA Adapter tubes.
    - Stop Ligation Buffer.

Note: Do not remove the Ligation Mix 2 tube from -20 ^{0}C storage until instructed to do so in the procedure.

2. Remove the Resuspension Buffer from 4 ^{0}C storage and bring it to room temperature.
3. Remove the Sample Purification Beads from 4 ^{0}C storage and let stand for at least 30 minutes to bring them to room temperature.
4. Preprogram the thermal cycler with the following program:
    - Choose the thermal cycler preheat lid option and set to 100 ^{0}C
    - 30 ^{0}C for 10 minutes
    - Hold at 4 ^{0}C

Ligate Adapters Reactions

1. Centrifuge the thawed tubes at 600 x g for 5 seconds.
2. Centrifuge the Stop Ligation Buffer tube at 600 x g for 5 seconds.
3. Immediately before use, remove the Ligation Mix 2 tube from -20 ^{0}C storage.
4. Remove the adhesive seal from the plate.
5. Add 2.5 µl Resuspension Buffer to each well.
6. Add 2.5 µl Ligation Mix 2 to each well.
7. Return the Ligation Mix 2 tube to -20 ^{0}C storage immediately after use.
8. Seal the plate with an adhesive seal.
9. Add 2.5 µl thawed DNA Adapter Index to each sample according to the table:

Sample	DNA Adapter Index	Index sequence
B-2S-1	AD005	ACAGTG
B-2S-2	AD006	GCCAAT
B-2S-3	AD002	CGATGT
B-2S-4	AD004	TGACCA
B-12-9-1	AD012	CTTGTA
B-12-9-2	AD019	GTGAAA
B-12-9-3	AD007	CAGATC
B-12-9-4	AD016	CCGTCC

10. Seal the plate with an adhesive seal.
11. Place the sealed plate on the preprogrammed thermal cycler. Close the lid and run the program:
    - Set the lid temperature to 40 ^{0}C
    - 30 ^{0}C for 10 minutes
    - Hold at 4 ^{0}C

Note: Set the lid temperature to 40 ^{0}C instead of 100 ^{0}C.

12. Remove the adhesive seal from the plate.
13. Add 5 µl Stop Ligation Buffer to each well to inactivate the ligation. Set a 200 µl pipette to 30 µl, then gently pipette the entire volume up and down 10 times to mix thoroughly.

Clean Up the Ligation Products

1. Vortex the Sample Purification Beads for at least 1 minute or until they are well dispersed.
2. Add 42.5 µl mixed Sample Purification Beads to each well. Set a 200 µl pipette to 60 µl and then gently pipette the entire volume up and down 10 times to mix thoroughly.
3. Incubate the plate at room temperature for 5 minutes.
4. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
5. Remove and discard 80 µl of the supernatant from each well. Take care not to disturb the beads.

   Note: Leave the plate on the magnetic stand while performing the following steps 6-11.
6. With the plate on the magnetic stand, add 200 µl freshly prepared 80% ethanol to each well. Do not disturb the beads.

7. Incubate the plate at room temperature for 30 seconds, and then remove and discard all of the supernatant from each well. Take care not to disturb the beads.
8. Repeat steps 6 and 7 to perform an 80% ethanol wash 2 times.
9. Remove and discard any ethanol residues from each well with a 10 µl pipette.
10. With the plate on the magnetic stand, let the samples air-dry at room temperature for 2 minutes. Do not allow the beads to crack.
11. With the plate on the magnetic stand, add 52.5 µl Resuspension Buffer to each well. Make sure the Resuspension Buffer runs over the beads. Do not touch the beads with pipette tip.
12. Remove the plate from the magnetic stand.
13. Resuspend the beads in each well by repeatedly dispensing the Resuspension Buffer over the bead pellet until it is immersed in the solution. Gently pipette the entire volume up and down 10 times to mix thoroughly.
14. Incubate the plate at room temperature for 2 minutes.
15. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
16. Transfer 50 µl of the clear supernatant from each well to a new 96-well plate. Take care not to disturb the beads.
17. Vortex the Sample Purification Beads for at least 1 minute or until they are well dispersed.
18. Add 50 µl mixed Sample Purification Beads to each well for a second cleanup. Set a 200 µl pipette to 75 µl, and then gently pipette the entire volume up and down 10 times to mix thoroughly.
19. Incubate the plate at room temperature for 5 minutes.

20. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
21. Remove and discard 95 µl of the supernatant from each well. Take care not to disturb the beads.

    Note: Leave the plate on the magnetic stand while performing the following steps 22-27.
22. With the plate on the magnetic stand, add 200 µl freshly prepared 80% ethanol to each well. Do not disturb the beads.
23. Incubate the plate at room temperature for 30 seconds, and then remove and discard all of the supernatant from each well. Take care not to disturb the beads.
24. Repeat steps 22 and 23 to perform an 80% ethanol wash 2 times.
25. Remove and discard any ethanol residues from each well with a 10 µl pipette.
26. With the plate on the magnetic stand, let the samples air-dry at room temperature for 5 minutes. Do not allow the beads to crack.
27. With the plate on the magnetic stand, add 27.5 µl Resuspension Buffer to each well. Make sure the Resuspension Buffer runs over the beads. Do not touch the beads with the pipette tip.
28. Remove the plate from the magnetic stand.
29. Resuspend the beads in each well by repeatedly dispensing the Resuspension Buffer over the bead pellet until it is immersed in the solution. Gently pipette the entire volume up and down 10 times to mix thoroughly.
30. Incubate the plate at room temperature for 2 minutes.
31. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
32. Transfer 25 µl of the clear supernatant from each well of the plate to a new 96-well plate. Take care not to disturb the beads.

33. Proceed to Enrich DNA Fragments or safely stop the protocol at – 20 ^{0}C for up to 7 days.

Enrich DNA Fragments

This process uses PCR to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. The PCR is performed with a PCR Primer Cocktail that anneals to the ends of the adapters. Minimize the number of PCR cycles to avoid skewing the representation of the library.

Preparation:

1. Remove the Enhanced PCR Mix and PCR Primer Cocktail from - 20^{0}C storage and thaw them at room temperature.
    - Invert the thawed tubes to mix.
    - Centrifuge the thawed tubes to 600 x g for 1 minute.
    - Do not vortex the tubes.
2. Remove the Resuspension Buffer from 4 ^{0}C and bring it to room temperature.
3. Remove the Sample Purification Beads from 4 ^{0}C and let stand for at least 30 minutes to bring them to room temperature.
4. Preprogram the thermal cycler with the following program:

    Choose the preheat lid option and set to 100 ^{0}C.

95 ^{0}C	3 minutes	
98 ^{0}C	20 seconds	8 cycles
60 ^{0}C	15 seconds	
72 ^{0}C	30 seconds	

72 ^{0}C 5 minutes

4 ^{0}C hold

Enrich DNA Fragments Reactions:

1. Add 5 µl thawed PCR Primer Cocktail to each well.
2. Add 20 µl thawed Enhanced PCR Mix to each well. Set a 200 µl pipette to 35 µl, then gently pipette the entire volume up and down 10 times to mix thoroughly.
3. Seal the plate with adhesive seal.
4. Place the sealed PCR plate on the preprogrammed thermal cycler. Close the lid, then run the program:

   Choose the preheat lid option and set to 100 ^{0}C.

   95 ^{0}C 3 minutes

   98 ^{0}C 20 seconds ┐

   60 ^{0}C 15 seconds ├ 8 cycles

   72 ^{0}C 30 seconds ┘

   72 ^{0}C 5 minutes

   4 ^{0}C hold

Clean up PCR products:

1. Remove the adhesive seal from the plate.
2. Vortex the Sample Purification Beads for at least 1 minute or until they are well dispersed.

3. Add 50 µl mixed Sample Purification Beads to each well. Set a 200 µl pipette to 75 µl , then gently pipette the entire volume up and down 10 times to mix thoroughly.
4. Incubate the plate at room temperature for 5 minutes.
5. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
6. Remove and discard 95 µl of the supernatant from each well.

   Note: Leave the plate on the magnetic stand while performing the following 80% ethanol wash steps.
7. With the plate on the magnetic stand, add 200 µl freshly prepared 80% ethanol to each well. Do not disturb the beads.
8. Incubate the plate at room temperature for 30 seconds, then remove and discard all of the supernatant from each well.
9. Repeat steps 7 and 8 to perform an 80 % ethanol wash 2 times.
10. Remove and discard ethanol residue from each well with a 10 µl pipette.
11. With the plate on the magnetic stand, let the samples air-dry at room temperature for 5 minutes. Do not allow the beads to crack.
12. With the plate on the magnetic stand, add 32.5 µl Resuspension Buffer to each well. Make sure the Resuspension Buffer runs over the beads. Do not touch the beads with the pipette tip.
13. Remove the plate from the magnetic stand.
14. Resuspend the beads in each well by repeatedly dispensing the Resuspension Buffer over the bead pellet until it is immersed in the solution. Gently pipette the entire volume up and down 10 times to mix thoroughly.
15. Incubate the plate at room temperature for 2 minutes.

16. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
17. Transfer 30 µl of the clear supernatant from each well of the PCR plate to new PCR tubes and label them.

Second-time Enrich DNA Fragments Reactions:

1. Aliquot 25 µl of the PCR products generated from above into a new 96-well plate and label the sample name.
2. Add 5 µl thawed PCR Primer Cocktail to each well.
3. Add 20 µl thawed Enhanced PCR Mix to each well. Set a 200 µl pipette to 35 µl, then gently pipette the entire volume up and down 10 times to mix thoroughly.
4. Seal the plate with adhesive seal.
5. Place the sealed PCR plate on the preprogrammed thermal cycler. Close the lid, then run the program:

   Choose the preheat lid option and set to 100 ^{0}C.

95 ^{0}C	3 minutes	
98 ^{0}C	20 seconds	8 cycles
60 ^{0}C	15 seconds	
72 ^{0}C	30 seconds	
72 ^{0}C	5 minutes	
4 ^{0}C	hold	

Clean up PCR products:

1. Remove the adhesive seal from the plate.
2. Vortex the Sample Purification Beads for at least 1 minute or until they are well dispersed.
3. Add 50 µl mixed Sample Purification Beads to each well. Set a 200 µl pipette to 75 µl, then gently pipette the entire volume up and down 10 times to mix thoroughly.

4. Incubate the plate at room temperature for 5 minutes.
5. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.
6. Remove and discard 95 µl of the supernatant from each well.

   Note: Leave the plate on the magnetic stand while performing the following 80% ethanol wash steps.
7. With the plate on the magnetic stand, add 200 µl freshly prepared 80% ethanol to each well. Do not disturb the beads.
8. Incubate the plate at room temperature for 30 seconds, then remove and discard all of the supernatant from each well.
9. Repeat steps 7 and 8 to perform an 80 % ethanol wash 2 times.
10. Remove and discard ethanol residue from each well with a 10 µl pipette.
11. With the plate on the magnetic stand, let the samples air-dry at room temperature for 5 minutes. Do not allow the beads to crack.
12. With the plate on the magnetic stand, add 32.5 µl Resuspension Buffer to each well.

    Make sure the Resuspension Buffer runs over the beads. Do not touch the beads with the pipette tip.
13. Remove the plate from the magnetic stand.
14. Resuspend the beads in each well by repeatedly dispensing the Resuspension Buffer over the bead pellet until it is immersed in the solution. Gently pipette the entire volume up and down 10 times to mix thoroughly.
15. Incubate the plate at room temperature for 2 minutes.
16. Place the plate on the magnetic stand for 5 minutes or until the liquid is clear.

17. Transfer 30 µl of the clear supernatant from each well of the PCR plate to new PCR tubes and label them.
18. Library QC with Qubit and Fragment Analyzer.

For the library concentration:

Library	Concentration (ng/µl)
B-2S-1	79.2
B-2S-2	71.6
B-2S-3	61.6
B-2S-4	61.4
B-12-9-1	81.4
B-12-9-2	72.2
B-12-9-3	60.2
B-12-9-4	71.6

Data Preprocessing

Raw reads were generated from two separate runs on NextSeq (300 Cycles) PE150 Mid Output flow cell (B-2S-1, B-2S-2, B-12-9-1 and B-12-9-2 on the same run, and B-2S-3, B-2S-4, B-12-9-3, B-12-9-4 on the same run):

Step 1: trimmomatic cleaning

Sample	Total reads
B-2S-1-R1	40,198,224
B-2S-1-R2	40,198,224
B-2S-2-R1	37,259,526
B-2S-2-R2	37,259,526
B-2S-3-R1	36,653,170
B-2S-3-R2	36,653,170
B-2S-4-R1	23,683,476
B-2S-4-R2	23,683,476
B-12-9-1-R1	54,663,861
B-12-9-1-R2	54,663,861
B-12-9-2-R1	40,162,002
B-12-9-2-R2	40,162,002
B-12-9-3-R1	27,254,678
B-12-9-3-R2	27,254,678
B-12-9-4-R1	28,127,046
B-12-9-4-R2	28,127,046

#PBS -S /bin/bash

#PBS -N j_trim

#PBS -q highmem_q

#PBS -l nodes=1:ppn=16

#PBS -l walltime=96:00:00

#PBS -l mem=100gb

#PBS -M yk20528@uga.edu

#PBS -m abe

cd $PBS_O_WORKDIR

module load Trimmomatic/0.36-Java-1.8.0_144

time java -jar /usr/local/apps/eb/Trimmomatic/0.36-Java-1.8.0_144/trimmomatic-0.36.jar PE -threads 4 R1.fastq R2.fastq sample-R1-paired.fastq sample-R1-unpaired.fastq sample-R2-paired.fastq sample-R2-unpaired.fastq ILLUMINACLIP:trimmomatic.fa:2:30:10 LEADING:3 SLIDINGWINDOW:4:15 MINLEN:36

The trimmomatic.fa attached is built based on overrepresented sequences detected by FASTQC, Illumina TruSeq universal adapter, TruSeq Index adapters (1-27, 17, 24 and 26 are reserved) and SMARTer adapters

Results:

Sample	Total reads	Number of reads passed trimming	Number of trimmed reads in pairs	Number of trimmed reads unpaired
B-2S-1-R1	40,198,224	31,502,905	7,267,216	24,235,689
B-2S-1-R2	40,198,224	8,117,544	7,267,216	850,328
B-2S-2-R1	37,259,526	31,094,294	6,197,566	24,896,728
B-2S-2-R2	37,259,526	6,519,653	6,197,566	322,087
B-2S-3-R1	36,653,170	29,576,294	4,949,274	24,627,020
B-2S-3-R2	36,653,170	5,787,938	4,949,274	838,664
B-2S-4-R1	23,683,476	19,531,626	3,046,296	16,485,330
B-2S-4-R2	23,683,476	3,649,862	3,046,296	603,566
B-12-9-1-R1	54,663,861	46,568,803	11,120,280	35,463,523
B-12-9-1-R2	54,663,861	11,950,956	11,120,280	830,676
B-12-9-2-R1	40,162,002	35,996,410	7,525,014	28,480,614
B-12-9-2-R2	40,162,002	7,865,427	7,525,014	340,415
B-12-9-3-R1	27,254,678	22,746,362	4,143,660	18,602,702
B-12-9-3-R2	27,254,678	4,719,572	4,143,660	575,912
B-12-9-4-R1	28,127,046	23,285,974	4,057,581	19,228,393

B-12-9-4-R2	28,127,046	4,826,876	4,057,581	769,295

Step 2:

rRNA removal using SortMeRNA

Index a few Pearl millet rRNA gene (see rRNA_Pearl_Millet.txt attached) and comprehensive rRNA databases downloaded from https://www.arb-silva.de/, run the scripts:

```
#!/bin/bash
#PBS -q batch
#PBS -l nodes=1:ppn=1
#PBS -l walltime=10:00:00
#PBS -l mem=20gb
#PBS -M yk20528@uga.edu
#PBS -m abe
cd $PBS_O_WORKDIR
module load SortMeRNA/2.1-foss-2016b
time indexdb_rna --ref rRNA_Pearl_Millet.fa,rRNA_Pearl_Millet_db:\
SILVA_132_LSURef_tax_silva.fasta,SILVA_132_LSURef_tax_silva_db:\
SILVA_132_SSURef_tax_silva.fasta,SILVA_132_SSURef_tax_silva_db
```

Separate each file into rRNA and non-rRNA groupings using the scripts:

```
#!/bin/bash
#PBS -q highmem_q
#PBS -l nodes=1:ppn=16
#PBS -l walltime=480:00:00
#PBS -l mem=150gb
#PBS -M yk20528@uga.edu
#PBS -m abe
cd $PBS_O_WORKDIR
module load SortMeRNA/2.1-foss-2016b
time  sortmerna --ref
rRNA_Pearl_Millet.fa,rRNA_Pearl_Millet_db:SILVA_132_LSURef_tax_silva.fasta,SILVA_13
2_LSURef_tax_silva_db:SILVA_132_SSURef_tax_silva.fasta,SILVA_132_SSURef_tax_silva_
db --reads file.fq --aligned rRNA.fq --fastx --blast 1 --log --other nonrRNA.fq -v
```

SILVA_132_LSURef_tax_silva.fasta (This SSU dataset containing all high quality, aligned 16S/18S ribosomal RNA sequences with a minimum length of 1200 bases for Bacteria and Eukarya and 900 bases for Archaea.)

SILVA_132_SSURef_tax_silva.fasta (This is the LSU reference database containing only high quality, aligned 23S/28S ribosomal RNA sequences with a minimum length of 1900 bases.)

Results:

Sample	Number of reads passed trimmomatic trimming	Total reads failing E-value threshold (non-rRNA)	Total non-rRNA reads
B-2S-1-R1	31,502,905	10,765,826	14,454,229
B-2S-1-R2	8,117,544	3,688,403	
B-2S-2-R1	31,094,294	16,531,372	20,534,545
B-2S-2-R2	6,519,653	4,003,173	
B-2S-3-R1	29,576,294	7,144,027	9,431,974
B-2S-3-R2	5,787,938	2,287,947	
B-2S-4-R1	19,531,626	4,260,134	5,265,418
B-2S-4-R2	3,649,862	1,005,284	
B-12-9-1-R1	46,568,803	21,632,173	27,114,879
B-12-9-1-R2	11,950,956	5,482,706	
B-12-9-2-R1	35,996,410	22,524,073	27,658,011
B-12-9-2-R2	7,865,427	5,133,938	
B-12-9-3-R1	22,746,362	8,673,324	10,265,448
B-12-9-3-R2	4,719,572	1,592,124	

B-12-9-4-R1	23,285,974	9,968,960	11,936,830
B-12-9-4-R2	4,826,876	1,967,870	

Step3: Trim_galore processing

Based on FASTQC per base sequence content plot, remove the first X bases from both ends to avoid poor qualities or biases.

```
#PBS -S /bin/bash
#PBS -N j_TrimGalore
#PBS -q highmem_q
#PBS -l nodes=1:ppn=2
#PBS -l walltime=48:00:00
#PBS -l mem=10gb
#PBS -M yk20528@uga.edu
#PBS -m abe
cd $PBS_O_WORKDIR
module load Trim_Galore/0.4.5-foss-2016b
trim_galore file.fastq --clip_R1 X1 --three_prime_clip_R1 X2
```

Results:

Sample	Total reads failing E-value threshold (non-rRNA)	Number of bases to be removed from their 5' end to avoid poor qualities or biases	Number of bp to be removed from their 3' end to avoid poor qualities or biases	Reads survived with a minimum length of 20 bases after trimming
B-2S-1-R1	10,765,826	9	5	10,746,240
B-2S-1-R2	3,688,403	35	5	3,119,600
B-2S-2-R1	16,531,372	9	5	16,525,688
B-2S-2-R2	4,003,173	35	5	3,694,022
B-2S-3-R1	7,144,027	9	5	7,111,026
B-2S-3-R2	2,287,947	35	5	1,897,395
B-2S-4-R1	4,260,134	9	5	4,214,318
B-2S-4-R2	1,005,284	35	5	738,964
B-12-9-1-R1	21,632,173	9	5	21,615,726
B-12-9-1-R2	5,482,706	35	5	4,981,145
B-12-9-2-R1	22,524,073	9	5	22,513,702

B-12-9-2-R2	5,133,938	35	5	4,820,005
B-12-9-3-R1	8,673,324	8	5	8,657,360
B-12-9-3-R2	1,592,124	35	5	1,359,411
B-12-9-4-R1	9,968,960	8	5	9,957,078
B-12-9-4-R2	1,967,870	35	5	1,619,338

Step 4: Biological contamination removal:

The de novo assembly was constructed using Trinity 2.5.1 by firstly concatenating all cleaned reads from B-2S and B-12-9 respectively and run trinity with parameters as follows: --seqType fq --single --no_normalize_reads. Transcripts assembled by trinity were annotated by BLASTN search against NCBI nt database with e-value set as 1e-10 for saving hits. Then construct a custom non-plant contamination database by extracting aligned subject sequences from the assembly that hit (< 1e-10) bacteria, fungi and animals sequences, index the non-plant contamination database to clean each library individually till no significant amount of non-plant hits is seen in the final assembly.

De Novo assembly construction using the script:

```
#!/bin/bash
#PBS -N j_s_trinity
#PBS -q highmem_q
#PBS -l nodes=1:ppn=16
#PBS -l walltime=480:00:00
#PBS -l mem=150gb
#PBS -M yk20528@uga.edu
#PBS -m abe
cd $PBS_O_WORKDIR
singularity exec /usr/local/singularity-images/trinity-2.5.1--0.simg Trinity --seqType fq --single file.fastq   --no_normalize_reads –output --max_memory 150G
```

Download the nt database from NCBI using the command:

wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/nt.tar.gz

Decompress the file:

tar -xzf nt.tar.gz

Index the decompressed NCBI nt database for BLASTN search:

#PBS -S /bin/bash

#PBS -N j-blast_trn

#PBS -l nodes=1:ppn=4:AMD

#PBS -q batch

#PBS -l walltime=200:00:00

#PBS -l mem=20gb

#PBS -M yk20528@uga.edu

#PBS -m abe

cd $PBS_O_WORKDIR

module load BLAST+/2.6.0-foss-2016b-Python-2.7.14

makeblastdb -in nt -dbtype nucl

Perform BLASTN search and report the top hit only:

```
 #PBS -S /bin/bash
#PBS -N j-blast_trn
#PBS -l nodes=1:ppn=16
#PBS -q highmem_q
#PBS -l walltime=200:00:00
#PBS -l mem=200gb
#PBS -M yk20528@uga.edu
#PBS -m abe
cd $PBS_O_WORKDIR
module load BLAST+/2.6.0-foss-2016b-Python-2.7.14
blastn -num_threads 16 -query Trinity.fasta -db nt -out blastn.outfmt6 -evalue 1e-10 -
max_target_seqs 1 -outfmt "6 qseqid sseqid pident qlen length slen qcovs mismatch gapopen
qstart qend sstart send qseq sseq evalue bitscore db_hit_len pct_hit_len_aligned hit_descr"
```

Examine the detailed alignment statistics using the scripts:

```
#PBS -S /bin/bash
#PBS -N j-blast_trn
#PBS -l nodes=1:ppn=16
#PBS -q highmem_q
#PBS -l walltime=200:00:00
#PBS -l mem=200gb
#PBS -M yk20528@uga.edu
```

```
#PBS -m abe

cd $PBS_O_WORKDIR

singularity exec /usr/local/singularity-images/trinity-2.5.1--0.simg

analyze_blastPlus_topHit_coverage.pl blastn.outfmt6 Trinity.fasta nt
```

Examine all trinity contigs that have non-plant hit descriptions, and extract the aligned portion of the subject sequences as a biological contamination database. And remove the biological contamination using the SortMeRNA.

Results:

Sample	Total reads prior to biological contamination removal	Total reads failing E-value threshold (non-contaminants)
B-2s-1	13,865,840	11,801,994
B-2s-2	20,219,710	19,288,065
B-2s-3	9,008,421	7,362,519
B-2s-4	4,953,282	4,359,291
B-12-9-1	26,596,871	25,809,903
B-12-9-2	27,333,707	26,707,906
B-12-9-3	10,016,771	9,479,492
B-12-9-4	11,576,416	10,753,322

Assembly before biological contamination removal	# of plant hit descriptions	# of non-plant hit descriptions	# of unique plant hit descriptions
B-2S	39,770	7,836	16,730
B-12-9	53,698	3,482	20,913

Assembly after biological contamination removal	# of plant hit descriptions	# of non-plant hit descriptions	# of unique plant hit descriptions
B-2S	37,856	84	16,651
B-12-9	51,693	29	20,787

Transcript Abundance Estimation and Differential Expression Analysis

Kallisto quantification

Build kallisto index using the script:

```
#PBS -S /bin/bash
#PBS -q batch
#PBS -N jobname
#PBS -l nodes=1:ppn=1:AMD
#PBS -l walltime=4:00:00
#PBS -l mem=2gb
#PBS -M yk20528@uga.edu
#PBS -m abe
cd $PBS_O_WORKDIR
module load kallisto/0.43.1-foss-2016b
/usr/local/apps/eb/kallisto/0.43.1-foss-2016b/bin/kallisto index Trinity.fasta -i kallisto_index
```

To perform kallisto quantification, it requires the information of estimated average fragment length and estimated standard deviation of fragment length of each input library. Due to several trimming process done, estimation using library prep length (350 bp) and sequencing length (150 bp) is no longer reliable.

The average and standard deviation of the fragment length of each input library are computed using the scripts:

```
awk 'BEGIN { t=0.0;sq=0.0; n=0;} ;NR%4==2
{n++;L=length($0);t+=L;sq+=L*L;}END{m=t/n;printf("total %d avg=%f
stddev=%f\n",n,m,sq/n-m*m);}'  file.fq
```

Library	Total reads	Average length	Standard deviation
B2-S-1	11,801,994	106.684078	1385.641220
B2-S-2	19,288,065	113.878378	1053.144795
B2-S-3	7,362,519	108.283124	1366.066499
B2-S-4	4,359,291	105.361971	1552.199437
B-12-9-1	25,809,903	110.949123	1153.564779
B-12-9-2	26,707,906	115.669719	952.001423
B-12-9-3	9,479,492	113.533710	1133.045747
B-12-9-4	10,753,322	112.171880	1234.771453

```
#PBS -S /bin/bash

#PBS -q batch

#PBS -N jobname

#PBS -l nodes=1:ppn=1:AMD

#PBS -l walltime=4:00:00

#PBS -l mem=2gb

#PBS -M yk20528@uga.edu

#PBS -m abe


cd $PBS_O_WORKDIR


module load kallisto/0.43.1-foss-2016b
```

/usr/local/apps/eb/kallisto/0.43.1-foss-2016b/bin/kallisto quant -i kallisto_index -o output --single -l -s file.fastq

(l stands for average length, s stands for standard deviation)

Build gene and transcript-level abundance estimates for each of the libraries, construct a matrix of raw counts and a matrix of TMM-normalized expression values using the script:

```
#!/bin/bash
#PBS -N j_s_kallisto
#PBS -q batch
#PBS -l nodes=1:ppn=1
#PBS -l walltime=48:00:00
#PBS -l mem=20gb
#PBS -M yk20528@uga.edu
#PBS -m abe

module load Trinity/2.6.6-foss-2016b

/usr/local/apps/eb/Trinity/2.6.6-foss-2016b/trinityrnaseq-Trinity-v2.6.6/util/abundance_estimates_to_matrix.pl --est_method kallisto \
--gene_trans_map Trinity.fasta.gene_trans_map \
--out_prefix  kallisto \
--cross_sample_norm TMM \
--name_sample_by_basedir \
```

B2-S-1_quant/abundance.tsv \

B2-S-2_quant/abundance.tsv \

B2-S-3_quant/abundance.tsv \

B2-S-4_quant/abundance.tsv \

B-12-9-1_quant/abundance.tsv \

B-12-9-2_quant/abundance.tsv \

B-12-9-3_quant/abundance.tsv \

B-12-9-4_quant/abundance.tsv

Results:

6 files are generated:

yk_kallisto.gene.counts.matrix

yk_kallisto.gene.TMM.EXPR.matrix

yk_kallisto.gene.TPM.not_cross_norm

yk_kallisto.isoform.counts.matrix

yk_kallisto.isoform.TMM.EXPR.matrix

yk_kallisto.isoform.TPM.not_cross_norm

DESeq2 differential expression analysis in R studio

```
library(edgeR)
library(DESeq2)
data = read.table("yk_kallisto.isoform.counts.matrix", header=T, row.names=1, com='')
col_ordering = c(5,6,7,8,1,2,3,4)
rnaseqMatrix = data[,col_ordering]
rnaseqMatrix = round(rnaseqMatrix)
rnaseqMatrix = rnaseqMatrix[rowSums(cpm(rnaseqMatrix) > 1) >= 2,]
conditions = data.frame(conditions=factor(c(rep("B-12-9", 4), rep("B2-S", 4))))
rownames(conditions) = colnames(rnaseqMatrix)
ddsFullCountTable <- DESeqDataSetFromMatrix(
    countData = rnaseqMatrix,
    colData = conditions,
    design = ~ conditions)
dds = DESeq(ddsFullCountTable)
contrast=c("conditions","B-12-9","B2-S")
res = results(dds, contrast)
baseMeanA <- rowMeans(counts(dds, normalized=TRUE)[,colData(dds)$conditions == "B-12-
9"])
baseMeanB <- rowMeans(counts(dds, normalized=TRUE)[,colData(dds)$conditions == "B2-
S"])
res = cbind(baseMeanA, baseMeanB, as.data.frame(res))
res = cbind(sampleA="B-12-9", sampleB="B2-S", as.data.frame(res))
```

```
res$padj[is.na(res$padj)]  <- 1
res = as.data.frame(res[order(res$pvalue),])
write.table(res, file='kallisto.isoform.counts.matrix.B-12-9_vs_B2-S.DESeq2.DE_results', sep='
', quote=FALSE)
write.table(rnaseqMatrix, file='kallisto.isoform.counts.matrix.B-12-9_vs_B2-
S.DESeq2.count_matrix', sep='  ', quote=FALSE)
```

CHAPTER 4

A CASE STUDY OF EGG CELL RNA-SEQ DATA VALIDATION AND UTILIZATION: APOMICTIC AND SEXUAL EGG CELL LOCALIZATION AND FUNCTIONAL VARIATIONS IN BUFFELGRASS (*CENCHRUS CILIARIS*) AS REVEARLED BY *EC1.3* (*EGG CELL 1.3*) IN SITU HYBRIDIZATION [5]

[5] Y. Ke, N. Thangthong, J. Conner and P. Ozias-Akins. To be submitted to *Plant Reproduction*

Abstract

Apomictic development involves a series of events defining the trajectory of embryo formation from a maternal cell in the ovule without fusion with sperm cell, which is intrinsically different than sexual reproduction in multiple facets. Examining these events in apomictic ovules in parallel with the sexual ovules will help us understand the molecular mechanisms of apomixis and its application in plant breeding as well as the eventual goal of synthesizing apomixis in major crops. Here we designed *EC1.3* (*egg cell 1.3*) probe based on egg cell RNA-seq data and analyses generated in the prior chapter of this dissertation, and used it to investigate sexual and apomictic egg localization and functional variations. This work partially validated our RNA-seq data and analyses as the *EC1.3* signal is shown to be egg cell-specific, and its utilization in quantitatively examining sexual and apomictic egg distribution revealed several properties related to apomictic eggs as well as a potential mechanism of how apomixis circumvents fertilization allowing parthenogenesis to proceed. In addition, in situ hybridization work was done on a subset of differentially expressed genes and validated our egg cell RNA-seq data and analyses. The present study serves as a validation of previous RNA-seq data and analyses and its utilization in examining egg cell distribution further deepened our understanding of the mechanisms of apomixis.

Introduction

The developmental process of apomixis encompasses some of the sexual reproduction events, resulting in viable seeds. The key difference is that the embryo in apomixis results from a maternal cell within the ovule instead of syngamy (fusion of egg and sperm). The apomictically-derived seed is genetically identical to the maternal plant.

In angiosperms, female sexual reproduction takes place in the ovule through a series of events, giving rise to a genetically unique seed. These events include: the formation and differentiation of a megaspore mother cell from nucellus, megaspore formation from megaspore mother cell through meiosis, embryo sac development from functional megaspore through mitosis, embryo and endosperm development from egg and polar nuclei through double fertilization. While in apomixis, some of these events are either omitted or adaptively altered but eventually result in a viable seed. Differences in these events and origins of cells that give rise to the embryo lead to three categories of apomixis: apospory, diplospory and adventitious embryony. This manuscript will focus on apospory, as this is the reproduction mode of apomictic buffelgrass.

Apospory is the second most commonly seen mechanism in apomictic species, employed by 110 genera of known apomictic species (Hojsgaard *et al.,* 2014). In obligate apospory, multiple random cells called aposporous initials in nucellus develop into aposporous embryo sacs through mitosis, giving rise to unreduced gametes. The number of aposporous embryo sacs varies with the species but usually ranges from 2 to 8 (Nogler 1984; Chapman *et al.,* 1990) and some of them are immature and may not be accurately counted. In the sexually reproducing species, generally there is only one cell called the megaspore mother cell that gives rise to a tetrad of megaspores with one functional megaspore developing into an embryo sac. The contrast in the number of cells that eventually give rise to embryo sacs between sexual and apospory may suggest that aposporous nucellus has less of a restrictive control than that of sexuals allowing multiple embryo sac differentiation events.

The number of nuclei and their arrangement observed in the aposporous embryo sacs also varies with the species. In *Hieracium*, the nuclei number and their distribution in the aposporous

embryo sacs resemble those seen in sexual embryo sacs (Koltunow 1993). In *panicoideae*, the aposporous embryo sacs were reported to contain four nuclei including one polar nucleus, one egg cell and two synergids (Brown *et al.,* 1958). In *Pennisetum*, the sexual embryo sac is of the *Polygonum* type comprised of proliferated antipodals, two polar nuclei, one egg cell, and two synergids, while the apomictic embryo sac is of a four-nucleate type with two polar nuclei (the central cell), one egg cell, and one synergid or one polar nucleus, one egg cell and two synergids (Naumova *et al.,* 2001). The multiple aposporous embryo sacs observed in apomictic buffelgrass usually have one that is similar to sexual embryo sac in terms of position (close to the micropylar end) and orientation (micropylar-chalazal polarity) while others are apparently randomly distributed, and can be in the nucellus, integuments and even ovary wall. Here we present a cytological examination of egg cell localization in sexual and apomictic buffelgrass ovaries by *EC1.3* (*Egg cell 1.3*) in situ hybridization, investigating the physical localization and potential functional variations between sexual and apomictic embryo sac development. *EC1*, *egg cell 1*, was discovered as the largest cluster of ESTs (expressed sequence tags) isolated from wheat egg (Sprunck *et al.,* 2005). Further identification and characterization of *EC1* orthologs in *Arabidopsis* confirmed its specificity to egg cells, and its critical role in triggering sperm cell activation during double fertilization (Sprunck *et al.,* 2012).

Here we designed an egg cell specific ssRNA probe based on *EC1.3* orthologs discovered in apomictic buffelgrass egg cell library, and applied it to both unpollinated sexual and apomictic ovary sections collected on the day of anthesis, investigating its egg cell-specificity and localization variations. In addition, other differentially expressed genes were also tested through in situ hybridization.

Materials and Methods

Plant Material and Florets Collection

The buffelgrass (*Cenchrus ciliaris*) plants used as a source of materials were vegetatively propagated tillers of the sexual genotype B-2s and natural apomictic genotype B-12-9 (Sherwood *et al.,* 1994). The plants were grown in the field for florets collection from July to August in 2019. The heads were bagged prior to stigma exsertion and stigmas were manually removed with tweezers upon appearance. The heads remained bagged until the day of anther exsertion (anthesis) at which time florets were collected. From 8:00 – 11:00 am on the day of anthesis, florets with fresh anthers half or fully exserted were collected and immediately fixed in ethanol (ethyl alcohol 200 proof, catalog number:111000200, PHARMCO) : acetic acid (acetic acid, glacial, catalog number: A38S-212, Fisher Chemical), 3:1 (Kerk *et al.,* 2003) and stored at 4 °C for overnight. Fixed florets were then transferred and stored in 70% ethanol (DEPC-treated water) at 4 °C.

Ovary Section Preparation

Intact ovaries were carefully dissected from the florets on ice using fine tweezers to avoid physical damage, and were dehydrated in a graded ethanol series on ice with gentle shaking for 20 minutes in each step as follows: 80% ethanol, 90% ethanol, and three changes of 100% ethanol. Ovaries were stored overnight in 100% ethanol at 4 °C, and transferred the next day to a xylene series with gentle shaking for 20 minutes in each step as follows: ethanol: xylene (3:1), ethanol: xylene (1:1), ethanol: xylene (1:3), and three changes of 100% xylene. Xylene-cleared ovaries were then transferred to a Paraplast series in an incubator (54 °C) with shaking for 9 hours in each step as follows: xylene and Paraplast (1:1) mixture, and seven changes of 100 % paraplast. Paraplast-infiltrated ovaries were embedded in Paraplast blocks using a tissue

embedding center, and 8-µm sections were cut on a rotary microtome (Leica RM2145, Germany). The sections were floated on methanol and mounted on microscope slides (Fisherbrand Probe On™ Plus). Slides were heated on a slide warmer at 42 °C for overnight to stretch the Paraplast ribbons and air-dry.

Slide Pretreatment

Slides with Paralast ribbons were de-paraffinized in xylene for 10 minutes twice, followed by a hydration series: 2 times of 100% ethanol, 95% ethanol, 70% ethanol, 50% ethanol, 30% ethanol, and 2 times of DEPC-treated water (2 minutes for each). Rehydrated slides were incubated in 0.2 M HCl for 10 minutes, and then washed in DEPC-treated water, 2 times of 2x SSC (saline sodium citrate), then DEPC-treated water, 5 minutes each. Slides were then treated with a mixture of 100 mM Tris pH 8.0, 50 mM EDTA pH 8.0 and 1 µg/mL proteinase K at 37 °C for 20 minutes, followed by a wash in PBS (phosphate buffered saline) for 2 minutes. Slides were incubated in 2 mg/mL glycine in PBS for 2 minutes to block proteinase K, and washed in PBS for 30 seconds twice. Slides were fixed in 4 % formaldehyde in PBS for 10 minutes and washed in PBS for 5 minutes twice, then dehydrated as follows: 2 times of DEPC-treated water, 30% ethanol, 50% ethanol, 70% ethanol, 95% ethanol and 2 times of 100% ethanol (2 minutes each). Slides were dried for 15 minutes in the fume hood.

EC1.3 Probe Synthesis

The forward primer 5'-GGCGAGATCATCCTGTACCT-3' and reverse primer 5'-GTCGCAGTATCCCTTGAGCAT-3' were designed based on the *EC1.3* (*egg cell-specific 1.3*) sequences derived from the egg cell RNA-seq assembly, and used for PCR to obtain *EC1.3* amplicon from the apomictic egg cell sequencing library. The *EC1.3* amplicon of expected size was cloned into pGEM®-T Easy Vector Systems (Promega) and sequenced through Macrogen,

and reverse transcribed into 162bp ssRNA probe using MAXIscript™ SP6/T7 Kit (Invitrogen).

Hybridization

Hybridization buffer was made with each of the following components at a final concentration of: 50% HiDi™ formamide (applied biosystems),100 µg/ml Yeast tRNA (Invitrogen), 0.75% blocking reagents (Roche), 2.5% dextran sulfate (Millipore), 2.5mM EDTA and DEPC-treated water to make up for the total volume. Probe (0.1 µl) was added to 10 µl of hybridization buffer and heated at 80 °C for 3 minutes. 190 µl hybridization buffer was added to 10 µl pre-denatured probe-hybridization buffer mixture and kept at 55 °C until use. For each slide with pretreated ovary sections, 200 µl pre-denatured probe-hybridization mixture were added, and sealed with another microscope slide on top. The slide pairs were incubated in a humidified environment at 50 °C for overnight.

Antibody Reaction and Detection

After incubation for hybridization, slides were first washed in 2x SSC (Sigma-Aldrich) and 50% formamide (Sigma-Aldrich ACS reagent) for 1 hour at 50 °C each, followed by another wash in TBS buffer (100 mM Tris-HCl pH 7.5, 400 mM NaCl) for 5 minutes. Slides were then incubated in 0.5% blocking reagents (Roche) in TBS for 1 hour followed by a wash in 1% BSA (bovine serum albumin, Sigma-Aldrich), 0.3% (v/v) Triton™ X-100 (Sigma-Aldrich) in TBS for 30 minutes. Slides were further incubated in Anti-Digoxigenin-AP Fab fragments (Roche) : 1% BSA, 0.3% (v/v) Triton™ X-100 in TBS (1:3000) for 90 minutes, followed by two washes in 1% BSA, 0.3% (v/v) TritonTM X-100 in TBS for 30 minutes each. For detection, slides were first washed in detection buffer (100mM Tris-HCl pH 9.5, 100mM NaCl, 50 $MgCl_2$) for 5 minutes, and then incubated in NBT/BCIP (Roche) detection buffer mixture (200 µl NBT/BCIP per 10 ml detection buffer) in dark for 12 hours.

Results and Discussions

EC1.3 is both sexual and apomictic egg cell specific

The positive signals are exclusively seen in the egg cell region with SP6 antisense probe (Figure 4.1A, Figure 4.2A) and T7 probe as the negative control (Figure 4.1B, Figure 4.2B), confirming that the *EC1.3* signals are egg cell specific for both sexual and apomictic eggs. We noticed that the *EC1.3* signals were also observed in multiple aposporous eggs in multiple apomictic embryos sacs (Figure 4.3), suggesting multiple eggs could be functional in terms of containing *EC1.3* transcripts. These results support the validity of our egg cell RNA-seq data that *EC1.3* sequences obtained from LCM-captured eggs also are detected in egg cells through in situ hybridization on ovary sections.

From a perspective of quantitative and positional *EC1.3* signals, investigations into the sexual and apomictic egg cell variations

As the *EC1.3* antisense probe was confirmed to be specific to both sexual and apomictic egg cells, we next quantitatively examined the specification and distribution of egg cells in sexual and apomictic ovary sections by doing *EC1.3* in situ hybridization. We processed 51 sexual ovaries and 94 apomictic ovaries, and collected the localization information of each egg cell structure observed and the presence or absence of *EC1.3* signal. We found that for 51 sexual ovaries examined, 51 egg cell structures were observed and 92.2% of them showed good *EC1.3* signal. For the 94 apomictic ovaries examined, 169 egg cell structures were observed and only 20.7% of them showed good *EC1.3* signals. These results suggest 92.2% of the unpollinated sexual egg cells observed on the day of anthesis are mature and likely to be functional for fertilization. Only 20.7% of the unpollinated apomictic egg cells displayed clear *EC1.3* signal, suggesting the majority of apomictic egg cells may not be receptive to fertilization. This is

consistent with low frequencies (1.3%) of BIII hybrids (hybrids from unreduced 2n egg cells fertilized with meiotically reduced n sperm cells) formation observed in apomictic buffelgrass (Bashaw *et al.,* 1990). Low frequencies of BIII hybrids also occur in other apomictic *Cenchrus* (Bashaw *et al.,* 1992). Our results suggest that the low frequency of BIII hybrid formation may be due in part to the inferior capacity for apomictic egg cells to get fertilized. Interestingly, we see a higher computed expression level of *EC1.3* transcripts in apomictic eggs than the sexual eggs based on the prior egg cell RNA-seq study, which seems to be contradictory to the current observations. The possible explanations could be that the RNA-seq data representing *EC1.3* expression is a result of the computed number of RNA molecules extracted from about 8000 egg apparatus sections collected at the micropylar end of nearly 4000 ovary sections from 8 individual plants matching *EC1.3* contigs, which is intrinsically representing the average expression level of *EC1.3* of egg cells at the micropylar end. While *EC1.3* in situ hybridization is only qualitative rather than quantitative, it is possible that the 20% of apomictic egg cells containing *EC1.3* as revealed by in situ hybridization work may have a much higher abundance than sexual egg at per single cell level to compensate the less number of egg expressing *EC1.3*. Reduced seed set was only observed in *Arabidopsis* when *EC1.2/3* were downregulated in egg cells of the triple knockout *ec1.1/4/5*, indicating that *EC1.3* plays a role in seed set but in a redundant manner (Sprunck *et al.,* 2012). We only observed abundant expression of *EC1.3* and *EC1.4* orthologs and low expression of *EC1.2-like* in buffelgrass eggs according to the RNA-seq study, suggesting *EC1.3* may play a major role with *EC1.4* in seed set in buffelgrass. Another reason that fertilization doesn't happen in apomictic egg could be the formation of a cell wall acting as a physical barrier preventing fertilization (Vielle *et al.,* 1995). The speculated high expression of *EC1.3* per apomictic egg is possibly correlated with the relatively high seed set and

high fertility of the BIII hybrids observed in *Cenchrus* (Bashaw *et al.,* 1990; Bashaw *et al.,* 1992). To test this hypothesis, we may need further *EC1.3* expression data of the egg cells of the *Cenchrus* BIII hybrids.

We next examined the localization of the sexual and apomictic egg cells by modeling an ovary section coordinate system where we set the micropyle (Busri *et al.,* 1993) as the origin and micropylar-chalazal polarity as the y axis, then plotted the relative coordinates of the center point of the egg cell. The apomictic egg cells were more widely distributed along the y axis, while the sexual eggs are strictly restricted to micropylar ends (Figure 4.4 and Figure 4.5), consistent with previous studies of sexual and apomictic embryo sac and egg cell arrangement (Willemse *et al.,* 1984; Chapman *et al.,* 1990; Koltunow 1993). As for *EC1.3* signals, there was no obvious correlation between the presence of the signal with its localization, suggesting the specification of apomictic egg cells was independent of localization.

In addition, preliminary characterizations of DEGs *ASGR-BBML*, *KRYPTONITE*, *NF-YA10* generated in the egg cell RNA-seq study were also achieved through in situ hybridization. With a previously designed *ASGR-BBML* probe and newly synthesized *KRYPTONITE*, *NF-YA10* probes. The *ASGR-BBML* signals were only observed in the apomictic eggs with SP6 probes, while absent from apomictic eggs with T7 probes and sexual eggs with either SP6 or T7 probes (Figure 4.6), which validated our RNA-seq data analyses that *ASGR-BBML* gene is de novo expressed in apomictic eggs. As for *KRYPTONITE, NF-YA10*, positive signals were observed exclusively in egg cell region for both sexual and apomictic ovary sections (Figure 4.7 and Figure 4.8), which validated their presence in both sexual and apomictic egg cell RNA-seq data. Their specificity or predominant expression in the egg cell compared to other regions of ovule suggest they may play a specific role in the egg cell development. To exhaustively test the

function of candidate parthenogenesis genes identified in this study, transformation experiments are needed.

Through *EC1.3* in situ hybridization, we validated our RNA-seq data and bioinformatic analyses by observation that *EC1.3* probe was egg cell-specific in both sexual and apomictic genotypes. Also, we utilize this probe to investigate sexual and apomictic egg cell developmental questions and corroborate previously identified wide-spread distribution of the egg cells in apomictic genotypes and strictly micropylar localization of sexual eggs. Moreover, the low frequency of positive *EC1.3* signals seen in the apomictic ovary sections may provide a molecular explanation of low successful fertilization events in natural apomictic eggs, representing a potential mechanisms for how apomixis circumvents fertilization allowing parthenogenesis to proceed. In addition, the *ASGR-BBML*, *KRYPTONITE*, and *NF-YA10* in situ hybridization work also validated the RNA-seq data and computational analyses. In all, the present study provide a case study of the validation and utilization of our egg cell RNA-seq data.

Figures and Tables

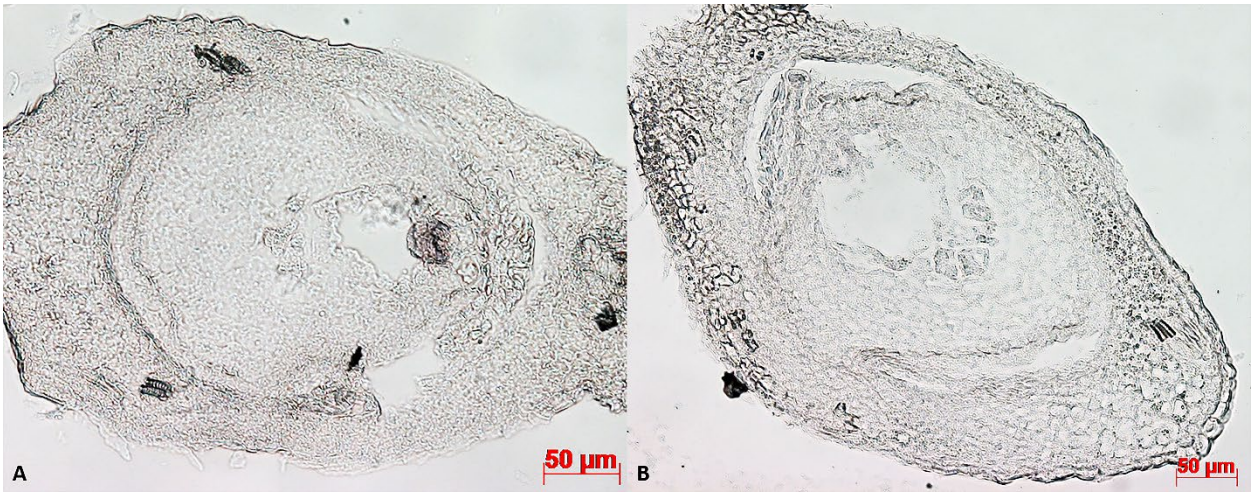

Figure 4.1: (A) SP6 antisense probe positive signal in sexual ovary section; (B) T7 sense probe negative control in sexual ovary section.

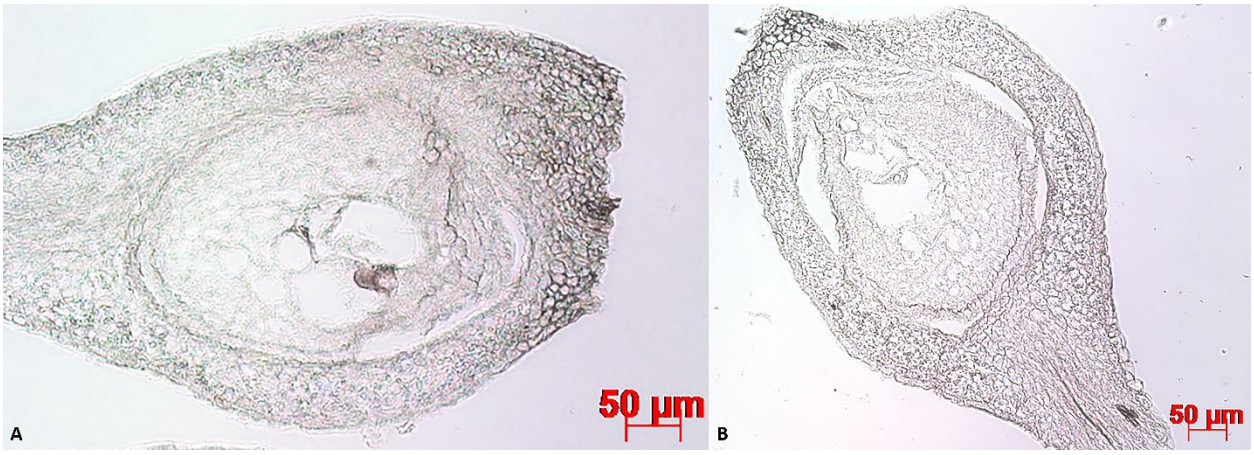

Figure 4.2: (A) SP6 antisense probe positive signal in apomictic ovary section; (B) T7 sense probe negative control in apomictic ovary section.

Figure 4.3: Two egg cells indicated by black arrows showed *EC1.3* signals in two separate aposporous embryo sacs.

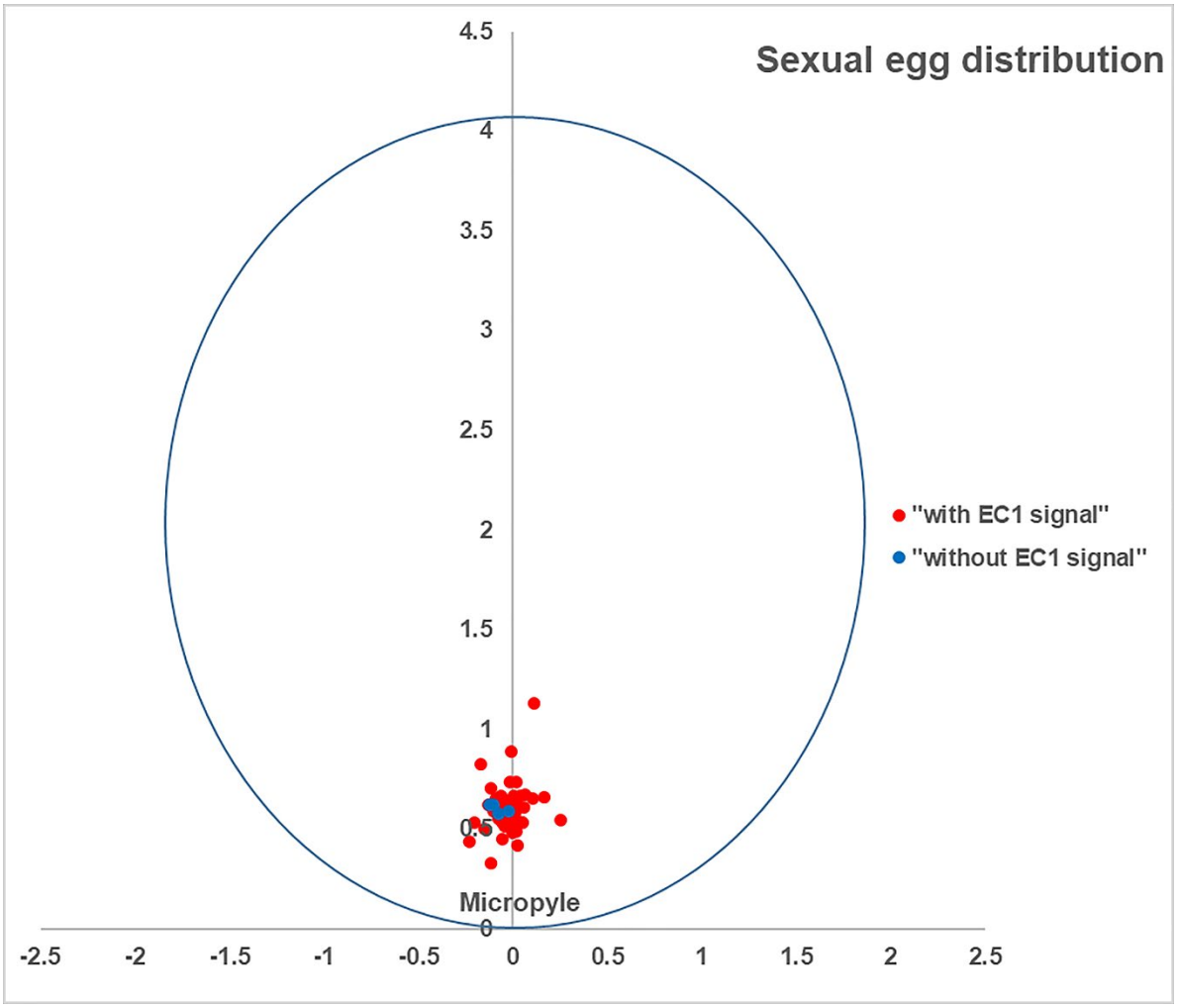

Figure 4.4: Sexual egg cell and *EC1.3* signals distribution in an ovary section model.

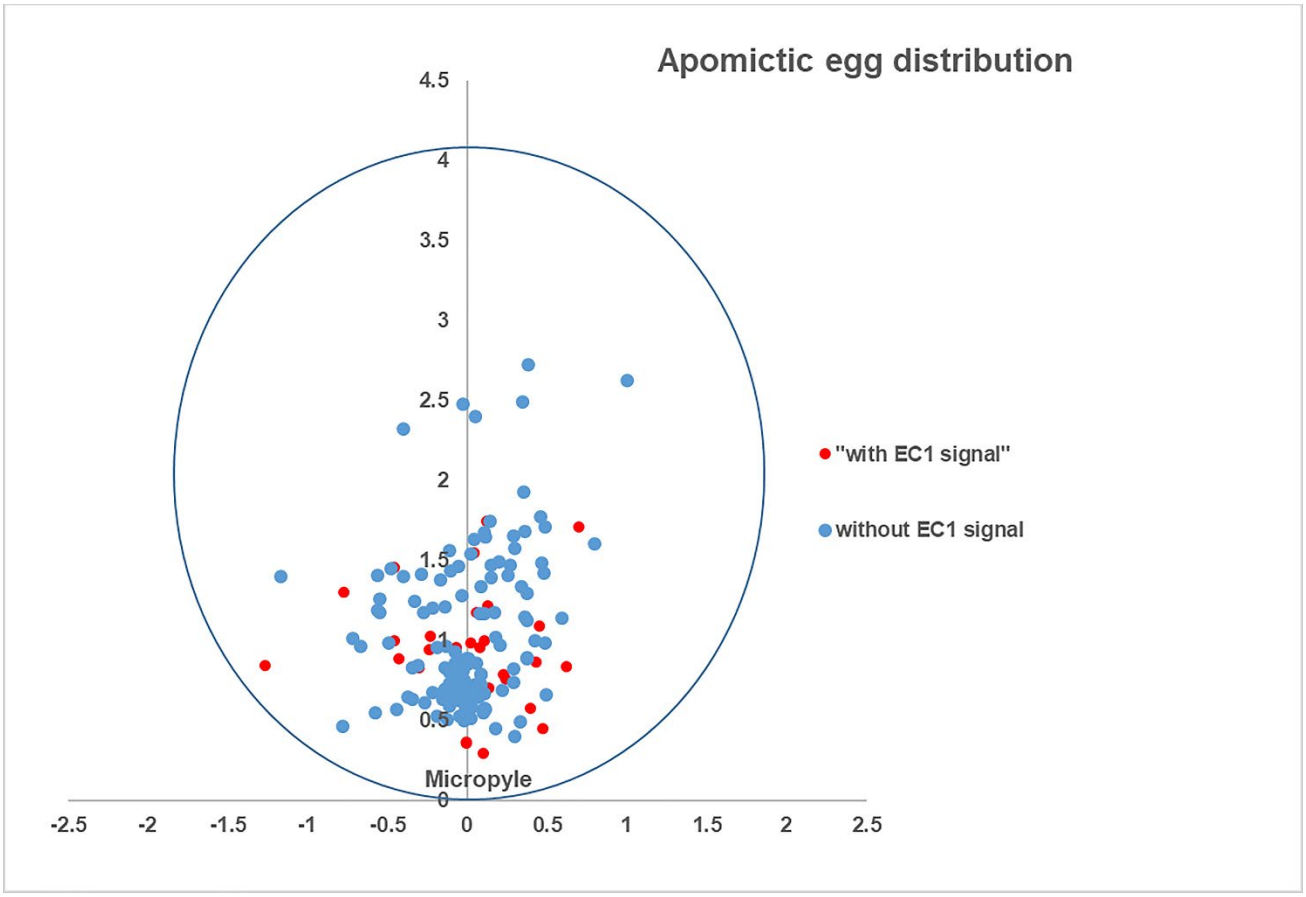

Figure 4.5: Apomictic egg cell and *EC1.3* signals distribution in an ovary section model.

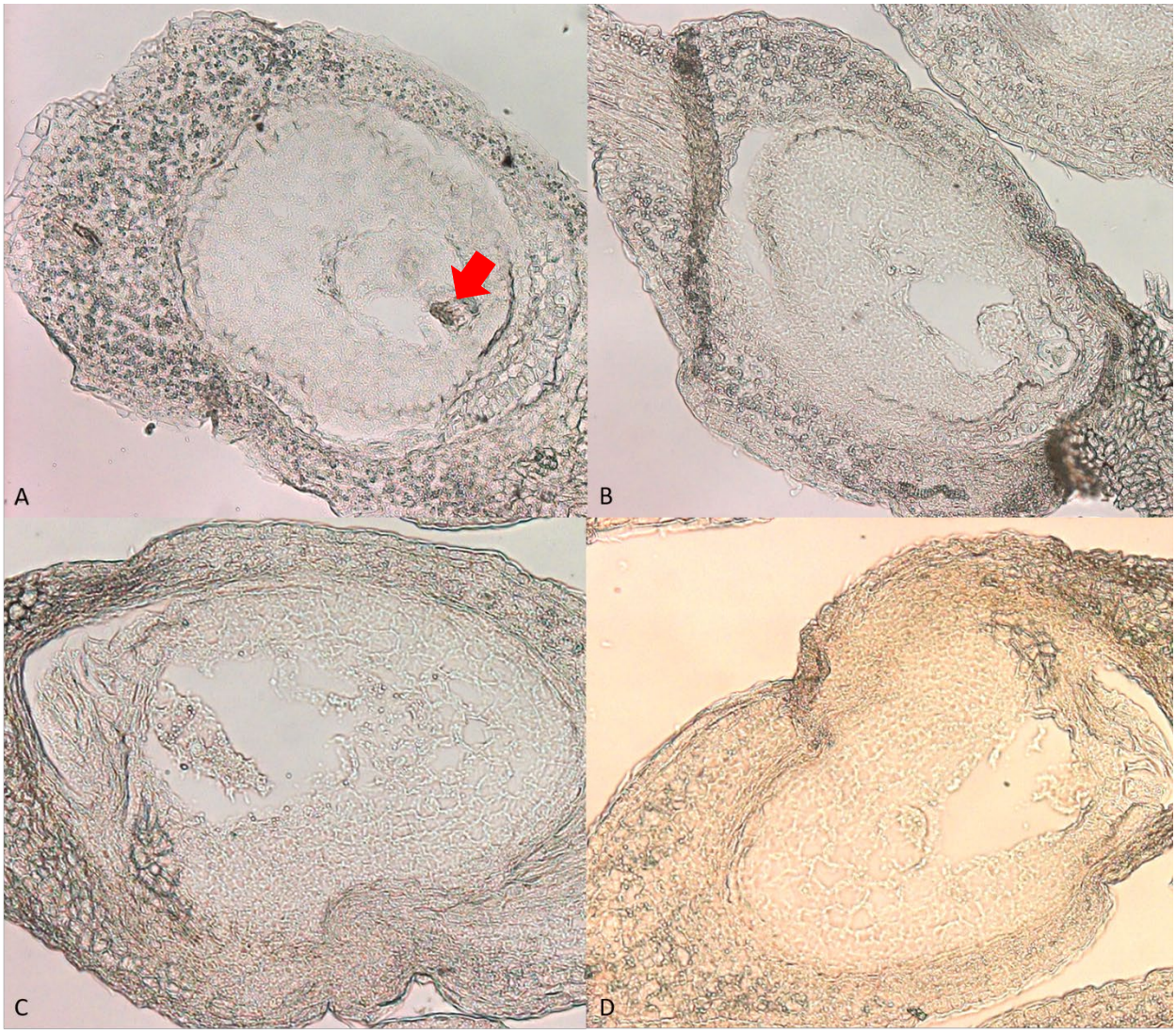

Figure 4.6: (A) *ASGR-BBML* SP6 antisense probe positive signal (red arrow) in an apomictic ovary section; (B) *ASGR-BBML* T7 sense probe negative control in an apomictic ovary section; (C) *ASGR-BBML* SP6 antisense probe negative signal in a sexual ovary section; (D) *ASGR-BBML* T7 sense probe negative control in a sexual ovary section.

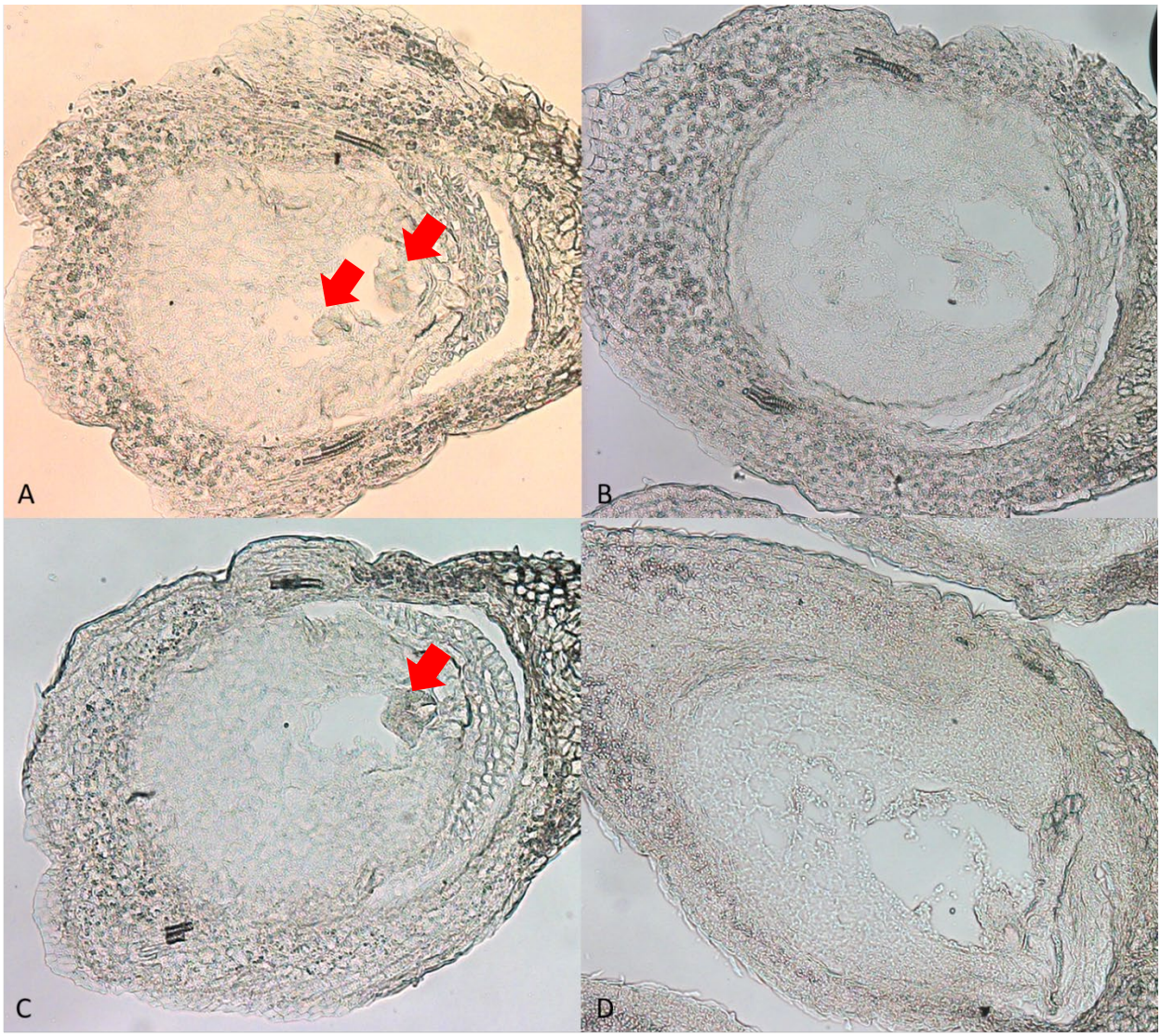

Figure 4.7: (A) *KRYPTONITE* SP6 antisense probe positive signal (red arrows) in an apomictic ovary section; (B) *KRYPTONITE* T7 sense probe negative control in an apomictic ovary section; (C) *KRYPTONITE* SP6 antisense probe positive signal (red arrow) in a sexual ovary section; (D) *KRYPTONITE* T7 sense probe negative control in a sexual ovary section.

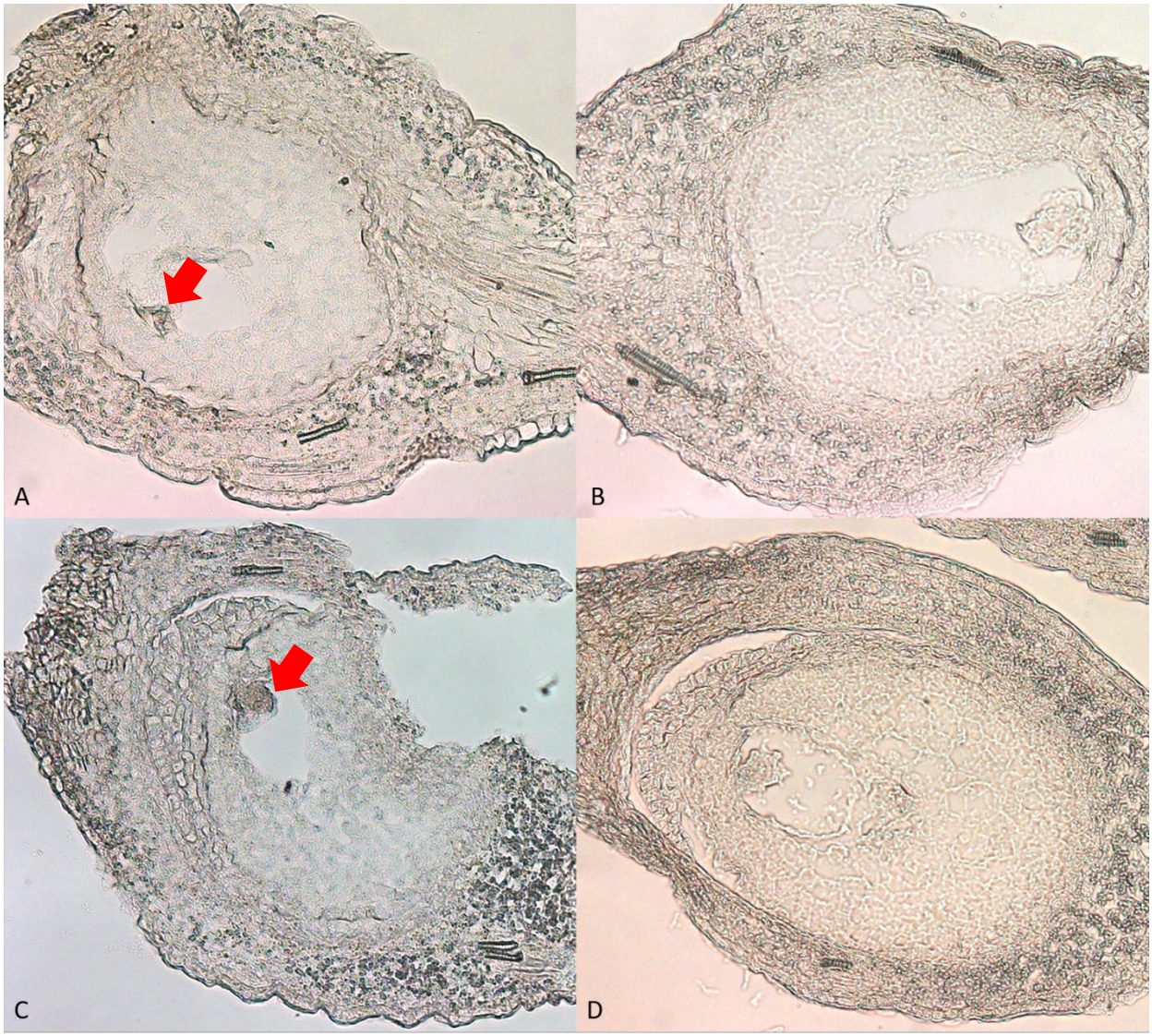

Figure 4.8: (A) *NF-YA10* SP6 antisense probe positive signal (red arrows) in an apomictic ovary section; (B) *NF-YA10* T7 sense probe negative control in an apomictic ovary section; (C) *NF-YA10* SP6 antisense probe positive signal (red arrow) in a sexual ovary section; (D) *NF-YA10* T7 sense probe negative control in a sexual ovary section.

Table 4.1:

	# of ovaries	# of egg cell structures	% egg cell per ovaries	# of *EC1.3* signals	% of *EC1.3* signals
B-2s (sexual)	51	51	1	47	92.2%
B-12-9 (apomictic)	94	169	1.8	35	20.7%

References

Bashaw, E. C. and K. W. Hignight (1990). "Gene transfer in apomictic buffelgrass through fertilization of an unreduced egg." Crop Science 30(3): 571-575.

Bashaw, E. C., M. A. Hussey and K. W. Hignight (1992). "Hybridization (n+ n and 2n+ n) of facultative apomictic species in the *Pennisetum* agamic complex." International journal of plant sciences 153(3, Part 1): 466-470.

Brown, W. V. and W. H. P. Emery (1958). "Apomixis in the Gramineae: panicoideae." American Journal of Botany: 253-263.

Busri, N., G. P. Chapman and J. Greenham (1993). "The embellum: a newly defined structure in the grass ovule." Sexual Plant Reproduction 6(3): 191-198.

Hojsgaard, D., S. Klatt, R. Baier, J. G. Carman and E. Hörandl (2014). "Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics." Critical Reviews in Plant Sciences 33(5): 414-427.

Kerk, N. M., T. Ceserani, S. L. Tausta, I. M. Sussex and T. M. Nelson (2003). "Laser capture microdissection of cells from plant tissues." Plant physiology 132(1): 27-35.

Koltunow, A. M. (1993). "Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules." The Plant Cell 5(10): 1425.

Naumova, T. N. and J.-P. Vielle-Calzada (2001). "Ultrastructural analysis of apomictic development." Flowering of Apomixis: From Mechanisms to Genetic Engineering, Y. Savidan, JG Carman, and T. Dresselhaus, eds (Mexico: CIMMYT, IRD, European Commission DG VI): 44-63.

Nogler, G. A. (1984). Gametophytic apomixis. Embryology of angiosperms, Springer: 475-518.

Sherwood, R. T., C. C. Berg and B. A. Young (1994). "Inheritance of apospory in buffelgrass." Crop Science 34(6): 1490-1494.

Sprunck, S., U. Baumann, K. Edwards, P. Langridge and T. Dresselhaus (2005). "The transcript composition of egg cells changes significantly following fertilization in wheat (Triticum aestivum L.)." The Plant Journal 41(5): 660-672.

Sprunck, S., S. Rademacher, F. Vogler, J. Gheyselinck, U. Grossniklaus and T. Dresselhaus (2012). "Egg cell–secreted EC1 triggers sperm cell activation during double fertilization." Science 338(6110): 1093-1097.

Vielle, J. P., B. L. Burson, E. C. Bashaw and M. A. Hussey (1995). "Early fertilization events in the sexual and aposporous egg apparatus of *Pennisetum ciliare* (L.) Link." The Plant Journal 8(2): 309-316.

Willemse, M. T. M. and J. L. Van Went (1984). The female gametophyte. In 'Embryology of angiosperms'.(Ed. BM Johri) pp. 159–191, Springer-Verlag: Berlin Heidelberg.

CHAPTER 5

SUMMARY

Engineering apomixis into major crops is the ultimate goal of apomixis research. Very recently, synthetic apomictic rice has been made through editing four sexually-derived genes, *PAIR1*, *OsREC8*, *OsOSD1*, and *OsMATL*, where a rate of 6% of clonal offspring from hybrid rice was induced. Another recent work in rice induced apomictic progeny at a rate up to 29% through editing *PAIR1*, *OsREC8*, *OsOSD1* and ectopically expressing *BBM1* under an egg cell-specific promoter, suggesting there could be multiple genes to manipulate to increase the apomixis induction rate. In an effort to identify more genes that can be used to pursue simple and efficient apomixis engineering, learning from nature is a must.

The main goal of this research is to identify natural apomixis genes, specifically, the parthenogenesis gene. The goal is achieved in three parts.

In the first study, RNA-seq was conducted on high-quality, abundant ovule RNA collected from the unpollinated sexual and apomictic ovules on the day of anthesis when parthenogenesis is initiating, and a de novo reference transcriptome assembly with near-complete coding sequences information for *Cenchrus ciliaris* was created. Through differential expression analyses, putative transcriptional pathways enriched in apomictic ovules were identified, providing us the hints to parthenogenesis development. As a reference, these data were further used to guide the assembly and annotation of the egg cell sequencing data.

In the second study, we overcame the technical challenges in capturing the egg cell that is deeply embedded in the ovaries with few contamination from other tissues and severe physical

damage, revising protocols to isolate RNA of sufficient quality and quantity, conversion to cDNA, and preparation of sequencing library, and eventually generated laser capture microdissection-based RNA-seq data on sexual and apomictic egg apparatus on the day of anthesis. With the guidance of ovule RNA-seq data generated from the first study, a de novo transcriptome with sequence information for the *Cenchrus ciliaris* egg apparatus was created. Through computational analysis, transcriptional profiles that distinguish apomictic egg from its sexual counterpart were identified, functional roles for a few transcription factors in promoting natural parthenogenesis were suggested. In addition, a novel *BBM2* gene was discovered and its full-length coding sequence was obtained. Further functional validation of *BBM2* awaits to be tested.

In the third study, in situ hybridization was conducted on *EC1.3*, *ASGR-BBML*, *KRYPTONITE*, and *NF-YA10* that were shown to be differentially expressed between sexual and apomictic eggs. The de novo expression of *ASGR-BBML* in apomictic eggs and egg cell specificity for all four genes were validated. In addition, we utilized *EC1.3* to investigate sexual and apomictic egg cell developmental questions, corroborated previously identified sexual and apomictic egg cell localization patterns and came up with a potential mechanism for how apomixis circumvents fertilization allowing parthenogenesis to proceed.

In all, the results of this work not only fulfilled the main goal of identifying natural parthenogenesis gene, but also provided two sets of valid tissue-specific transcriptomic resources for the global community of apomixis researchers. In conclusion, laser capture microdissection-based RNA-seq on unpollinated sexual and apomictic egg cells on the day of anthesis is able to identify natural parthenogenesis genes and provide novel insights into parthenogenesis from a high resolution perspective.