IDENTIFICATION AND CHARACTERIZATION OF APOSPORY CANDIDATE GENES IN *PENNISETUM* AND *CENCHRUS*

by

YAJUAN ZENG

(Under the Direction of Peggy Ozias-Akins)

ABSTRACT

Apomixis, asexual seed reproduction by avoidance of meiosis and fertilization of the egg cell, holds great potential for agriculture as a means to fix hybrid vigor since offspring generated through apomixis are genetic clones of the maternal plant. Apospory is a form of apomixis where the embryo develops from an unreduced egg that is derived from the aposporous embryo sac. In both *Pennisetum squamulatum* and *Cenchrus ciliaris*, apospory inherits as a dominant trait under genetic control of a single locus and transmits with an apospory-specific genomic region (ASGR). In order to understand the molecular mechanism regulating apospory, two major studies have been conducted.

Aposporous initial specification is a critical event for the occurrence of apospory. To elucidate the mechanism controlling this process, two transcriptomes, derived from ovules of an apomictic donor parent and its apomictic backcross derivative at the stage of apospory initiation, were sequenced using 454-FLX technology. Analysis of the two transcriptomes allowed identification of the ASGR-carrier chromosome linked transcripts. Moreover, one of these alien expressed genes was assigned to the ASGR by screening a limited number of apomictic and sexual F_1s. Identification of potential ASGR-linked candidate genes will provide significant

insight into the regulation of apospory initiation. The results suggest that the strategy of comparative sequencing of donor parent and backcross transcriptomes is an efficient approach to identify alien expressed genes in a recurrent parent background.

Previous functional analysis of *BABY BOOM* (*BBM*) genes from other species suggests that *BBM* plays a role in embryo development. To investigate the function of *BABY BOOM-like* genes identified from the ASGR compared with related genes not associated with the ASGR, *ASGR-BBM-like* and *non-ASGR-BBM-like* (*N-ASGR-BBM-like*) genes isolated from *Cenchrus ciliaris* were subjected to sequence analysis and expression characterization. Semi-quantitative RT-PCR with aposporous ovaries from different stages indicated that *ASGR-BBM-like* genes started transcription before pollination and the transcription was up-regulated upon pollination. In contrast, there was no expression of *N-ASGR-BBM-like* genes until one day after pollination. Moreover, the transcription level of *N-ASGR-BBM-like* was much lower than that of *ASGR-BBM-like* at this time point. Temporal and quantitative differences in transcript level between the two suggest that *ASGR-BBM-like* plays a role in parthenogenesis in buffelgrass.

INDEX WORDS: *Pennisetum squamulatum*, *Cenchrus ciliaris*, apospory initiation, 454 sequencing, parthenogenesis, semi-quantitative RT-PCR, *BBM-like* gene

IDENTIFICATION AND CHARACTERIZATION OF APOSPORY CANDIDATE GENES IN *PENNISETUM* AND *CENCHRUS*

by

YAJUAN ZENG

BS, Fujian Normal University, P.R.China, 2000

MS, Fujian Normal University, P.R.China, 2003

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

2009

© 2009

Yajuan Zeng

All Rights Reserved

IDENTIFICATION AND CHARACTERIZATION OF APOSPORY CANDIDATE GENES IN *PENNISETUM* AND *CENCHRUS*

by

YAJUAN ZENG

Major Professor: Peggy Ozias-Akins

Committee: Wayne Hanna
Kelly Dawe
Joann Conner

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2009

DEDICATION

I dedicate this dissertation to my family for their unconditional and endless love and support.

ACKNOWLEDGEMENTS

I would like to give my first gratitude to my major professor and chair of my committee, Dr. Ozias-Akins, for her guidance, encouragement, and support from the initial to the final level in the last five and half years. I also truly appreciate her patience with my dissertation writing and her careful review and editing of my dissertation.

I would like to thank committee members: Dr. Joann Conner, who has shared her excellent techniques and research ideas, Dr. Wayne Hanna, who has shared his expertise and precious materials and equipments, Dr. Kelly Dawe, who has shared his insightful comments, suggestions, encouragement and support and also taught me a wonderful course about plant genetics, and Dr. Emidio Albertini, who has shared his great techniques and ideas, kind friendship and encouragement.

It has been a great opportunity to work with a group of great people in Dr. Ozias-Akins' lab in the last five and half years. I thank all former and present lab members who have assisted me in any form. My thanks go to Drs. Ye Chu, Heqiang Huo, Yukio Akiyama, Shailendra Goel, Maria Laura Ramos, Manjit Singh, Anjanabha Bhattacharya and Joe Knoll, for sharing their valuable techniques and discussions. I also thank Ms. Evelyn Morgan, Ms. Judith Anne Bell, Ms. Gunawati Gunawan, Ms. Congling Wu, Ms. Jackie Merriman and Mr. Greg Thomas for their great technical support. I greatly appreciate everyone I've known in the group, including Dr. Song Zhang, Dr. Vishvas Kulkarni, Ms. Paola Faustinelli and Manju Chandran for their friendship, encouragement and being fun to be with.

I want to give my thanks to the faculties and staff members of Department of Horticulture and NESPAL for their assistance through the entire process, especially Ms. Evelyn Folds for her patience in providing any assistance requested.

I want to thank all of my friends for their priceless friendship. Many of them have been my best cheerleaders. I want to give my special thanks to Ryan Contreras, Lixing Yang and Limei Liu for their help during my research and dissertation writing.

I sincerely appreciate my family's unconditional love, support and confidence in me when I was pursuing my dream, even when the dream went beyond boundaries of language and geography.

My final and deepest appreciation must go to my husband, who is always there for me. It is impossible for me to complete the study and research for my Ph.D. degree without his love, support, encouragement, understanding and companionship.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Apomixis is a term used for asexual reproductive processes which occur during seed development (Nogler, 1984). Unlike zygotic embryogenesis, embryos derived through apomixis form from unreduced somatic cells in the ovule via mitosis and without the fusion of a haploid egg with a sperm cell. The apomixis pathway involves three key components (Grimanelli et al., 2001; Koltunow and Grossniklaus, 2003; Ozias-Akins, 2006). The first is apomeiosis, in which meiosis is avoided prior to embryo sac formation. The second is parthenogenesis, autonomous development of the embryo independent of fertilization of the egg cell by the sperm. The third is endosperm development, either requiring fertilization of the central cell or its autonomous development. Avoidance of meiosis and fertilization of the egg cell in apomixis results in progeny genetically identical to their mother plant. This feature has practical significance in agricultural breeding since it can easily perpetuate hybrid vigor and maintain specific genotypes through successive seed generations (Hanna and Bashaw, 1987; Hanna, 1995; Koltunow et al., 1995; Savidan, 2000). However, apomixis is rarely found among major crops (Bicknell and Koltunow, 2004). Moreover, apomictic species of significant relatedness to sexual crops to be used for introgression of apomixis to its close sexual species is also limited, which complicates conventional breeding of apomictic crops (Dujardin and Hanna, 1989). Therefore, the better solution for introduction of apomixis into crops may be by genetic engineering of desired sexual crops. Toward this goal, the genetic control of apomixis from naturally occurring apomictic species needs to be determined.

We are studying a form of apomixis called apospory in *Pennisetum squamulatum* and *Cenchrus ciliaris*. *P. squamulatum* is a wild relative of pearl millet and is cross-compatible with pearl millet when used as a pollen donor in the interspecific cross (Dujardin and Hanna, 1989). A restriction fragment length polymorphism (RFLP) study with an obligate apomictic backcross line resulted in identification of two apomixis-linked molecular makers (Ozias-Akins et al., 1993), while twelve sequence characterized amplified region (SCAR) markers strictly cosegregated with apospory in an interspecific hybrid F_1 (*P. glaucum* x *P. squamulatum*) population of 397 individuals that segregated for apomixis and sexuality and therefore defined a contiguous apospory-specific genomic region (ASGR) in *P. squamulatum* (Ozias-Akins et al., 1998). When tested for linkage to apospory in two small segregating populations of *C. ciliaris*, most of the SCAR markers showed no recombination with the ASGR (Roche et al., 1999). In order to characterize the ASGR, two bacterial artificial chromosome (BAC) libraries were constructed with an apomictic polyhaploid derived from a cross between the obligate apomict, *P. squamulatum*, and sexual pearl millet and apomictic *C. ciliaris* line, B-12-9 (Roche et al., 2002). ASGR-linked BACs were identified by screening of the two BAC libraries with SCAR markers. By using the ASGR-linked BACs as probes, fluorescence in situ hybridization (FISH) showed that the ASGR was a physically large, heterochromatic, hemizygous, retrotransposon-rich region in both *P. squamulatum* and *C. ciliaris* (Goel et al., 2003; Akiyama et al., 2004; Akiyama et al., 2005). Genetic mapping of long terminal repeat (LTR)-based sequence specific amplified polymorphism (SSAP) markers showed that most of the markers were highly clustered and associated with apomixis (Huo et al., 2009), which was consistent with the observation of low recombination in ASGR. Putative protein-coding genes unrelated to transposable elements from ASGR were identified by partial sequencing of ASGR-linked BACs (Conner et al., 2008).

However, the characteristics of the ASGR impede high-resolution genetic mapping and map-based cloning of apomixis genes.

To extend our attempts to understand the molecular mechanism regulating apospory, two major studies focusing on isolation and characterization of apospory candidate genes were conducted. The first study centered on expression profiling of ovule tissues at the stage of aposporous initial specification by analyzing two 454-FLX technology generated transcriptomes, one from an apomictic donor parent and the second from its apomictic backcross derivative. Identical sequences between the two genotypes are likely to represent transcribed genes from the ASGR since the ASGR-carrier chromosome is the only shared genetic component. The objective of the second study was to characterize the expression pattern of *BABYBOOM-like (BBM-like)* genes in *C. ciliaris*, one group specific to the ASGR and the second group not present in the ASGR and either unlinked or linked in repulsion. *BBM-like* genes are known to be involved in embryogenesis and the *ASGR-BBM-like* gene could play a role in parthenogenetic development of the unreduced egg during apomictic reproduction.

References

Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in Pennisetum squamulatum reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. Plant Physiol **134:** 1733-1741

Akiyama Y, Hanna WW, Ozias-Akins P (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. Theor Appl Genet **111:** 1042-1051

Bicknell RA, Koltunow AM (2004) Understanding apomixis: recent advances and remaining conundrums. Plant Cell **16 Suppl:** S228-245

Conner JA, Goel S, Gunawan G, Cordonnier-Pratt MM, Johnson VE, Liang C, Wang H, Pratt LH, Mullet JE, Debarry J, Yang L, Bennetzen JL, Klein PE, Ozias-Akins P (2008) Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of Pennisetum and Cenchrus. Plant Physiol **147:** 1396-1411

Dujardin M, Hanna WW (1989) Developing apomictic pearl millet-characterization of a BC_3 plant. J Genet Breed **43:** 145-151

Goel S, Chen Z, Conner JA, Akiyama Y, Hanna WW, Ozias-Akins P (2003) Delineation by fluorescence in situ hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in Pennisetum squamulatum and Cenchrus ciliaris. Genetics **163:** 1069-1082

Grimanelli D, Leblanc O, Perotti E, Grossniklaus U (2001) Developmental genetics of gametophytic apomixis. Trends Genet **17:** 597-604

Huo H, Conner J, Ozias-Akins P (2009) Genetic mapping of the apospory-specific genomic region in Pennisetum squamulatum using retrotransposon-based molecular markers. Theor Appl Genet **119:** 199-212

Koltunow AM, Grossniklaus U (2003) Apomixis: a developmental perspective. Annu Rev Plant Biol **54:** 547-574

Nogler GA (1984) Gametophytic apomixis. *In* BM Johri, ed, Embryology of Angisoperms. Springer-Verlag, Berlin, pp 475–518

Ozias-Akins P (2006) Apomixis: Developmental Characteristics and Genetics. Critical Reviews in Plant Sciences **25:** 199-214

Ozias-Akins P, Lubbers EL, Hanna WW, McNay JW (1993) Transmission of the apomictic mode of reproduction in Pennisetum: co-inheritance of the trait and molecular markers. Theor Appl Genet **85:** 632-638

Ozias-Akins P, Roche D, Hanna WW (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. Proc Natl Acad Sci U S A **95:** 5127-5132

Roche D, Cong P, Chen Z, Hanna WW, Gustine DL, Sherwood RT, Ozias-Akins P (1999) Short Communication: An apospory-specific genomic region is conserved between Buffelgrass (Cenchrus ciliaris L.) and Pennisetum squamulatum Fresen. Plant J **19:** 203-208

Roche D, Conner A, Budiman A, Frisch D, Wing R, Hanna W, Ozias-Akins P (2002) Construction of BAC libraries from two apomictic grasses to study the microcolinearity of their apospory-specific genomic regions. Theor Appl Genet **104:** 804-812

CHAPTER 2

LITERATURE REVIEW

Overview: Modes of Reproduction in Flowering Plants

Seed, the end product of reproduction in flowering plants and from which new plants grow, can be reproduced by two pathways: sexual reproduction and apomixis. The sexual pathway involves the union of male and female gametes, resulting in genetically varied offspring. In contrast, apomictic offspring are genetic clones of the maternal plant, a significant feature with great potential for plant breeding and agricultural production.

Sexual Reproduction

The life cycle of plants that reproduce sexually alternates between a diploid sporophytic phase and a haploid gametophytic phase (Reiser and Fischer, 1993; Grossniklaus and Schneitz, 1998; Grimanelli et al., 2001; Crawford and Yanofsky, 2008). The major reproductive function of the diploid sporophyte is to produce haploid spores through meiosis, the processes of which are known as megasporogenesis (female) and microsporogenesis (male). Ovule is the site of processes essential for plant reproduction, including formation of megagametophyte, fertilization (except for the apomicts which avoid the fertilization during the reproduction), embryogenesis, and the formation of the final product of reproduction, the seed. In *Arabidopsis* and many other plants, ovules derive from specialized meristematic regions as a primordium within the carpels referred to as the placentas (Schneitz, 1999). Therefore, the carpel must maintain existing or produce new meristematic region, the placenta, to allow the generation of ovule primordium. Outgrowth of the ovule primordium leads to the formation of the inner and outer integuments

from the surface of the ovule primordium. The two integuments grow to cover and enclose the nucellus of the ovule and leaving a small opening, the micropyle, for fertilization. The nucellus is usually several cells thick and provides the cellular initial for the differentiation of the megaspore mother cell (Robinson-Beers et al., 1992; Skinner et al., 2004). Megasporogenesis starts with the differentiation of the megaspore mother cell. A tetrad of megaspores results from meiosis of the megaspore mother cell and usually only one of the megaspores survives and becomes the functional megaspore while the other three degenerate. Microsporogenesis occurs within the anthers from multiple sporogenous cells. Unlike the megaspores, all four microspores resulting from meiosis of a microspore mother cell survive and are functional. Once the functional spores differentiate, the gametophytic phase of the plant life cycle begins. The spores undergo mitotic cell proliferation and differentiation to develop into gametophytes. Angiosperms have both female and male gametophytes. The male gametophyte, also referred to as the pollen grain or microgametophyte, develops within the anther and usually consists of two sperm cells enclosed within a large vegetative cell. The female gametophyte, also referred to as the embryo sac or megagametophyte, develops within the ovule. The Polygonum-type embryo sac, the most common type of megagametophyte in flowering plants, comprises seven cells, which are one egg cell, two synergid cells, one binucleate central cell and three antipodal cells (Willemse, 1984). Pollination in flowering plants begins when pollen lands on a stigma. Then the pollen tube grows through the stigma and style to deliver the sperm cells to the ovule. Pollen tube growth is critical for directing the sperm cell precisely to the egg to achieve the ultimate goal of fertilization (Sogo et al., 2004). Several independent studies have established that a high apical concentration of Ca^{2+} occurs in growing pollen tubes, and inhibition of the apical Ca^{2+} gradients results in the reversible inhibition of pollen tube growth (Pierson et al., 1994; Malho

and Trewavas, 1996; Franklin-Tong, 1999). The egg cell fuses with one of the sperm cells to produce the zygote which develops into the embryo and the other sperm cell fertilizes the central cell in the embryo sac to form the endosperm. In plants reproduced sexually, the events of MMC differentiation, megasporogenesis, megagametogenesis and double fertilization are in strictly ordered and defined sequence to assure the production of a fertile seed.

Apomixis

One hypothesis about why sexual reproduction is so widespread is that sexual reproduction can eliminate deleterious mutations and create variations for adaptation through genetic recombination (Otto and Lenormand, 2002). However, some theoretical work has led to the finding of the costs associated with sexual reproduction, such as time and energy consumption and the risk of introducing negative genetic element from another individual (Otto and Lenormand, 2002). Plants have evolved with considerable flexibility in the pathway of reproduction. In some flowering plants, the paternal contribution is not necessary for the generation of progeny because meiosis and fertilization of the egg by the male gamete can be avoided during reproduction. The trait is called apomixis or asexual reproduction through seeds (Nogler, 1984). The most important difference between apomixis and sexual reproduction is that the apomictic embryo derives from unreduced diploid cells in the maternal ovule rather than from a diploid zygote produced by fusion of haploid egg and sperm cells. During sexual reproduction, the gametes have half the number of chromosomes of the sporophyte because of meiosis; the diploid chromosome number in the embryo is restored by the fusion of egg and sperm. Different from sexual reproduction, the absence of chromosome reduction and restoration in apomixis results in progeny genetically identical to their mother plant.

Significance of Apomixis

The practical significance of apomixis is that it can easily perpetuate hybrid vigor and maintain specific genotypes through successive seed generations in agricultural breeding (Hanna and Bashaw, 1987; Hanna, 1995; Koltunow et al., 1995; Savidan, 2000). However, apomixis is rarely found among major crops with the exception of subtropical fruit trees like mango, apple and *Citrus* (Bicknell and Koltunow, 2004). The possible reason for its low representation among crops is human selection of superior plants for cultivation (Bicknell and Koltunow, 2004). Transfer of apomixis to its close sexual species by conventional breeding methods has been unsuccessful except in pearl millet (Dujardin and Hanna, 1989). One difficulty contributing to the unsuccessful introgression of apomixis is the limited number of apomictic species of significant relatedness to sexual crops, which results in complex breeding strategies to accommodate the cross-incompatibility during apomictic crop breeding (Dujardin, 1984; Dujardin and Hanna, 1989). Therefore, the better solution for introduction of apomixis into crops may be the genetic engineering of the crops of interest. To achieve the goal, it is useful to obtain information about the genetic control of apomixis from naturally occurring apomictic species.

Developmental Mechanisms and Genetic Control of Apomixis

Apomixis has been observed in more than 400 species across 40 families and is most common in the *Poaceae*, *Rosaceae*, *Rutacea*e and *Asteraceae* (Hanna and Bashaw, 1987; Koltunow, 1993; Carman, 1997; Bicknell and Koltunow, 2004; Ozias-Akins, 2006). Studies of apomixis in different species have greatly expanded our knowledge of apomixis even though the genetic mechanisms underlying apomixis are still undetermined.

Three Essential Components during Apomixis Development

The hypothesis that gametophytic apomixis is a deregulation of the sexual reproduction process has been widely accepted since it was proposed in 1984 based on the study of apospory in *Ranunculus auricomus* (Nogler, 1984). It has been supported by the findings that apomixis is controlled by relative few loci and mutagenesis can induce components of apomixis in sexual plants (Grimanelli et al., 2001; Grossniklaus et al., 2001; Ozias-Akins, 2006; Ozias-Akins and van Dijk, 2007; Ravi et al., 2008). Based on this hypothesis, apomixis can be viewed as the product of putative cell fate changes and omission of critical steps in the sexual process regulated by the same signal pathways as sexual reproduction. By comparison with the essential steps of sexual reproduction, the apomixis pathway involves three key components (Grimanelli et al., 2001; Koltunow and Grossniklaus, 2003; Ozias-Akins, 2006). The first step is apomeiosis, in which meiosis is avoided prior to embryo sac formation. The second step is parthenogenesis, autonomous development of the embryo independent of fertilization of the egg cell by the sperm. Third is endosperm development, either requiring the fertilization of the central cell or its autonomous development. These three important events of apomixis occur by different mechanisms in different types of apomicts.

Different Types of Apomixis

Based on the time at which apomixis initiates during ovule development relative to the sexual pathway and the origin of the cell that initiates apomixis, apomixis is often classified into three types, termed apospory, diplospory and adventitious embryony (Nogler, 1984; Koltunow, 1993; Savidan, 2000; Koltunow and Grossniklaus, 2003; Ozias-Akins and van Dijk, 2007). Diplospory is initiated at the time of megaspore mother cell differentiation by altering cell fate of the MMC. Apospory is initiated after differentiation of the MMC and aposporous initials

differentiate from the nucellar cells surrounding the MMC. Both apospory and diplospory result in formation of a megagametophyte in the ovule without meiosis and the embryo develops from the unreduced egg cell inside the unreduced megagametophyte. Therefore, apospory and diplospory are commonly referred to as gametophytic apomixis. In adventitious embryony, embryos are initiated directly from somatic cells in ovular tissue with no intervening gametophytic phase so it has been referred to as sporophytic apomixis. In sporophytic apomixis, sexual reproduction occurs in parallel and may result in competition between reduced and unreduced embryos.

Sporophytic apomixis

Sporophytic apomixis occurs late in ovule development. Cells from either the inner integument or the nucellus initiate embryo development without forming an embryo sac structure (Koltunow, 1993; Koltunow and Grossniklaus, 2003). This is the major difference between sporophytic apomixis and gametophytic apomixis. Typically adventitious embryos form from nucellar cells and integumentary originated embryos are much less common (Koltunow, 1993; Ozias-Akins, 2006). Sporophytic apomixis and zygotic reproduction can coexist within the same ovule. Endosperm resulting from fertilization of the central cell in the reduced zygotic embryo sac can provide nutrients for either the survival of the sporophytic embryo or both unreduced and reduced embryos (Ozias-Akins, 2006; de Meeus et al., 2007). In the event that fertilization of the central cell is not successful, adventitious embryos can obtain nutrients for their development from the degrading nucellar and integument cells (Koltunow, 1993).

Sporophytic apomixis has been reported previously to be inherited as a quantitative trait in the mapping study of a cross of *Citrus volkameriana* × *Poncirus trifoliata* with identification of six quantitative trait loci (García et al., 1999). In a more recent study a possible adventitious

embryony controlling mechanism was proposed in which two genes A1 and A2 controlled sporophytic embryony in genus *Citrus* and *Poncirus* (Hong et al., 2001). Further studies in species reproducing by sporophytic apomixis will be necessary to confirm the genetic control of the trait.

Aposporous apomixis

Aposporous apomixis initiates from nucellar cells following megaspore mother cell differentiation (Koltunow, 1993). These nucellar cells are referred to as aposporous initials and usually possess large nuclei and dense cytoplasm. The aposporous initials develop into unreduced aposporous embryo sacs by mitosis instead of meiosis. Meiosis of the megaspore mother cell can either be completed resulting in functional chromosomally reduced megaspores or be arrested during the process. In both cases the products of sexual reproduction usually degenerate (Nogler, 1984). Different aposporous species or even genotypes within a species can have different numbers of aposporous embryo sacs formed in an ovule (Nogler, 1984; Bashaw, 1990). Endosperm development usually requires pollination and fertilization of the central cell, termed pseudogamy. The aposporous egg cell is rarely fertilized in most of the aposporous species (Nogler, 1984). However, exceptions have been reported. In *Hypericum perforatum*, fertilization of the unreduced egg cells occurred at high frequency in several accessions as indicated by a high 6C embryo peak of flow cytometric analyses (Matzk et al., 2001). And the fertilization of unreduced egg cells with unreduced pollen was observed at a low frequency in the Greenland populations of *Arabis holboellii* (Naumova et al., 2004). The fertilization of unreduced egg cell results in increased ploidy level of apomicts.

Apospory behaves as a dominant trait under genetic control of a single locus in *Pennisetum squamulatum*, *Cenchrus ciliaris* and some other species studied (Ozias-Akins et al.,

1998; Bicknell et al., 2000; Martinez et al., 2001; Jessup et al., 2002). In *P. squamulatum* and *C. ciliaris* the locus transmitted with apospory is a heterochromatic, retrotransposon-rich region of one chromosome arm, where the recombination is highly suppressed (Ozias-Akins et al., 1998; Goel et al., 2003; Akiyama et al., 2004; Akiyama et al., 2005). However, in aposporous *Poa pratensis*, genetic control of apomeiosis and parthenogenesis seemed to be separated and controlled by different loci (Albertini et al., 2001). Supported by the observation of discrete classes of expressivity that could best be explained by a five-gene complex genetic model, it was postulated that in *Poa pratensis*, apospory initiation and parthenogenesis were unlinked and they were controlled by single genes separately (Matzk et al., 2005).

Diplosporous apomixis

In diplospory, a megaspore mother cell differentiates from the nucellus and develops into an unreduced female gametophyte by variable developmental processes prior to diplosporous embryo sac initiation. Thus there are different types of diplospory (Koltunow, 1993; Crane, 2001). The most common type of diplospory is mitotic diplospory, also known as *Antennaria* type, in which the megaspore mother cell is inhibited from starting meiosis and directly undergoes mitosis to form an unreduced gametophyte. A second type is meiotic diplospory, where meiosis of the megaspore mother cell is initiated but arrested at meiosis I and the chromosome number in the cell is restored to the somatic cell chromosome number. Cell proliferation via mitosis subsequently leads to embryo sac differentiation. As in mitotic diplospory, the aberrant meiotic process occurring in megaspore mother cells in meiotic diplospory results in chromosomally unreduced embryo sacs.

One cell in diplosporous embryo sac is specified to perform the function of an egg cell, although it is chromosomally unreduced compared with meiotically derived egg cells.

Parthenogenesis of the unreduced egg cell results in diplosporous embryo formation without fertilization. Endosperm development is derived from unreduced polar nuclei either autonomously or pseudogamously. However, the processes of embryo and endosperm development seem to proceed in ordered sequence like in sexual reproduction except for the omission or aberrant nature of meiosis and fertilization. Therefore, the plants must have alternative signals or triggers for regulation of diplosporous embryo sac formation and embryo and endosperm development (Koltunow, 1993; Koltunow and Grossniklaus, 2003).

It is still unclear when and how the megaspore mother cell is initiated for the diplospory pathway. The mechanisms of inhibition of meiosis or arrest of meiosis are crucial for understanding the initiation of diplospory. In crosses of sexual and apomictic *Taraxacum*, it was reported that the genetic control of diplosporous embryo sac development and parthenogenesis was independent (Tas and Van Dijk, 1999; Van Dijk et al., 1999; Noyes and Rieseberg, 2000).

Apomixis Candidate Genes Identified from Apomicts

Toward the final goal to induce apomixis in crops by genetic manipulation, efforts have been made to identify genes that confer naturally evolved apomixis. Differential gene expression analysis between closely related apomictic and sexual genotypes with application of differential display technique is an efficient strategy for discovery of candidate apomixis genes. Apomixis- and sex-specific transcripts have been detected from different species. In *Brachiaria*, two apomixis-specific fragments were isolated from ovaries at anthesis (Leblanc et al., 1997). In a later study with *Brachiaria* that focused on earlier developmental stages, including sporogenesis and gametogenesis stages, eleven differentially expressed fragments were identified (Rodrigues et al., 2003). Three expressed sequence tags (ESTs) from *Paspalum notatum*, all highly similar in sequence, showed differential expression in flowers between apomictic and sexual F_1

individuals after apospory initiation (Pessino et al., 2001). Comparative analysis of gene expression patterns of three members of the *FER TILIZATION-INDEPENDENT SEED* (*FIS*)–class genes *MEDEA* (*MEA/FIS1*), *FIS2*, and *FERTILIZATION-INDEPENDENT ENDOSPERM* (*FIE/FIS3*) in *Hieracium* showed that the spatial and temporal expression patterns in apomictic *Hieracium* were comparable to those observed in embryo sacs, embryos, and endosperm in sexual *Hieracium* (Tucker et al., 2003). This finding indicates that sexual and apomictic reproduction share molecular pathways. A similar strategy, cDNA-AFLP (amplified fragment length polymorphism) has also been applied to isolate apomixis candidate genes. In *Poa pratensis*, 179 transcript fragments from spikelets showed quantitatively different expression between apomictic and sexual genotypes by cDNA-AFLP (Albertini et al., 2004). However, after further analysis of two candidate genes, *PpSERK* and *APOSTART*, by reverse transcription polymerase chain reaction (RT-PCR) and in situ hybridization, neither one was specifically expressed in the apomictic genotype (Albertini et al., 2005). Screening of cDNA libraries is another strategy used for identification of candidate apomixis genes. Putative egg-cell-specific and parthenogenesis-related genes were isolated from two egg-cell-specific cDNA libraries constructed with sexual and parthenogenetic egg cells dissected from the 'Salmon' system of wheat lines (Kumlehn et al., 2001). Also one gene showing apomict specific expression was identified from a *Panicum maximum* ovule cDNA library and shown to be expressed in both aposporous initials and aposporous embryos at four days after anthesis (Chen et al., 1999; Chen et al., 2005). By suppression subtractive hybridization, two transcripts expressed in embryo sacs were isolated from *Pennisetum ciliare* ovary libraries showing either higher expression in an apomictic than sexual genotype or specific expression in apomixis (Singh et al., 2007). In addition, two principal loci that collectively control apomixis in *Hieracium caespitosum* have

been identified by deletion mapping (Catanach et al., 2006). Loss of apomeiosis (*LOA*) locus regulates events associated with apomeiosis and the other unlinked locus, loss of parthenogenesis (*LOP*), controls events associated with the avoidance of fertilization (parthenogenesis).

Although differentially expressed transcripts have been identified from apomictic species, none of these candidate genes has been shown to be linked with apomixis. Genetic control of apomixis is still unclear given the differences in spatial and temporal expression patterns of candidate genes; moreover, it may be even more complicated with the contribution of epigenetic regulation (Lohe and Chaudhury, 2002; Koltunow and Grossniklaus, 2003). Therefore, proteomic analysis may provide complementary insight into understanding apomixis mechanisms. Recent proteomic analysis with apomictic sugar beet by matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has led to identification of twenty-four unique proteins, which included five protein spots present only in sugar beet apomictic monosomic addition line M14 and two unique to sexual sugar beet genotype *B. vulgaris* L.; the rest of the identified proteins were either up-regulated or down-regulated in apomictic line M14 (Zhu et al., 2008). Most of these proteins were shown to be involved in several important biological processes, such as cell division and energy metabolism. In a more recent study on apomictic sugar beet, a comparative proteomic and transcriptomic analysis of the sexual and apomictic flowers of sugar beet using high-resolution two-dimensional gel electrophoresis (2-DE) and liquid chromatography-mass spectrometry analysis (LC-MS) was completed (Li et al., 2009). A total of seventy-one differentially expressed protein spots, which could be potential protein markers for apomictic development were identified, although the level of proteomic data corresponding to transcriptomic data was low. These data provided new

insights into proteins expressed during apomictic development and the value of proteomic analysis to the study of apomixis.

Candidate Apomixis Gene Isolation Aided by Studies in Sexual Reproduction

Since apomixis is hypothesized to be due to deregulation of the sexual process, genes involved in megasporogenesis, female gametophyte development, embryogenesis and endosperm formation in sexual reproduction may also provide important insights into the process of apomixis.

Studies of ovule developmental mutants have revealed some important genes from sexually reproducing plants. The MADS box genes in plants have been shown to function in meristem identity and play essential roles in cell fate control and cell differentiation during floral organ development (Davies and Schwarz-Sommer, 1994; Theissen and Saedler, 1995). Other genes that function in the regulation of reproductive organ identity have been identified. *BELL* (*BEL1*), directs normal integument development, in part by suppressing AG expression in this structure (Ray et al., 1994). *AGAMOUS* (*AG*) is required for development of the stamens and carpels and inhibits function of *APETALA2* (*AP2*). The loss of function mutant *ag* converts stamens and carpels to petals and sepals, which is opposite to the mutant of *AP2* (Mizukami and Ma, 1992; Yu et al., 2004; Wang et al., 2008). *AlNTEGUMENTA* (*ANT*) shares functions with *AP2* in initiating integument development and regulating megasporogenesis (Elliott et al., 1996; Klucher et al., 1996; Morcillo et al., 2007); *LEUNIG* (*LUG*), encoding a glutamine-rich protein, is a key regulator for floral organ identity and development by interacting with *AG* (Conner and Liu, 2000). Genes required for female sporogenesis and gametogenesis during sexual reproduction are those involved in megaspore mother cell differentiation and embryo sac development. *ABSENCE OF FIRST DIVISION* (*AFD*) (Grossniklaus et al., 1998) and *DYAD* are

required for progression through female meiosis; in the mutant *dyad*, the megaspore mother cell enters but fails to complete meiosis, arresting at the end of meiosis I in the majority of ovules (Siddiqi et al., 2000). *AMEIOTIC1* (*AM1*) has been demonstrated to be required for all meiotic processes (Golubovskaya et al., 1993; Golubovskaya et al., 1997; Pawlowski et al., 2009). *SPOROCYTELESS* (*SPL*) regulates sporocyte development including microsporogenesis and megasporogenesis (Yang et al., 1999). *SWITCH1* (*SWI1*) is required for early meiotic events that are at the crossroad of sister chromatid cohesion, recombination and axial element formation (Mercier et al., 2003). *OSD1* (*omission of second division*) was isolated from *Arabidopsis* and shown to be directly involved in controlling entry into the second meiotic division (d'Erfurth et al., 2009). *BLH1* (*BEL1-like homeodomain 1*) regulates the cell-fate switch of synergid to egg cell in *Arabidopsis eostre* mutant by misexpression (Pagnussat et al., 2007). *DIANA*/*AGAMOUS-LIKE61* (*AGL61*) is required to maintain normal central cell morphology (Bemer et al., 2008).

Genetic analysis of embryogenesis and somatic embryogenesis in *Arabidopsis* and other plants have revealed several genes involved in regulation of embryogenesis in plants. The *SOMATIC EMBRYOGENESIS RECEPTOR KINASE* (*SERK*), isolated from carrot during somatic embryogenesis, is only expressed in the zygotic embryo up to the early globular stage but neither in unpollinated flowers nor in any other plant tissue (Schmidt et al., 1997). Overexpression of *Arabidopsis SERK 1* (*AtSERK1*) was sufficient to increase the embryogenic potential of the cultured cells and its expression was detected in the nucellus during megasporogenesis, in the functional megaspore and in the egg cell within the embryo sac (Hecht et al., 2001). *LEAFY COTYLEDON* (*LEC*) / *FUSCA* (*FUS*) genes were found essential for in vitro somatic embryogenesis induction with the studies of capacity for somatic embryogenesis in

lec1, *lec2* and *fus3* mutants of *Arabidopsis* (Gaj et al., 2005). Embryogenic potential was totally repressed in double (*lec1 lec2*, *lec1 fus3*, *lec2 fus3*) and triple (*fus3 lec1 lec2*) mutants. Furthermore, the ability to form somatic embryos in response to 2,4-dichlorophenoxyacetic acid was greatly decreased in the mutants (Gaj et al., 2005). The *BABY BOOM* (*BBM*) gene, a member of the AP2/ERF multigene family, was identified as a marker for embryo development in *Brassica napus* microspore-derived embryo cultures (Boutilier et al., 2002). Ectopic expression of *BnBBM* in *Arabidopsis* primarily induced spontaneous somatic embryo formation from seedlings with a low frequency of spontaneous shoots and callus development (Boutilier et al., 2002). It has been shown that *BBM* preferentially expressed in the basal region of the *Arabidopsis* embryo (Casson et al., 2005). Furthermore, expression of *BnBBM* was detected in microspore-derived embryos of *Brassica napus* during the induction and development stages as well as in developing zygotic seeds (Malik et al., 2007). A *BBM-like* gene, *EgAP2-1*, isolated from oil palm, was also shown to be an important regulator in both zygotic and somatic embryo development (Morcillo et al., 2007). However, ectopic expression of *BBM* in tobacco did not induce spontaneous somatic embryogenesis, which was different from observations in *Arabidopsis* and *Brassica*, although activated cell proliferation occurred (Srinivasan et al., 2007). Identification of *BBM* target genes in *Arabidopsis* seedlings with the application of DNA microarray analysis showed that these genes were predominantly associated with cell proliferation and growth (Passarinho et al., 2008). All the functional analysis data suggests an important role of *BBM* gene during embryo development in flowering plants. Another important gene known to be involved in embryogenesis is *WUSCHEL* (*WUS*). In *Arabidopsis* shoot and floral meristems, the *WUS* gene was required for stem cell identity and *WUS* expression was sufficient to induce meristem cell identity (Laux et al., 1996; Mayer et al., 1998; Schoof et al.,

2000). Overexpression of *WUS*/*PGA6* caused high-frequency somatic embryo formation in *Arabidopsis* without any exogenous plant hormones (Zuo et al., 2002). The result suggested that *WUS*/*PGA6* played a key role during embryogenesis, presumably by promoting the vegetative-to-embryogenic transition and/or maintaining the identity of the embryonic stem cells (Zuo et al., 2002). However, it is still unclear in which specific cell(s) these related genes are expressed during zygotic embryogenesis in plants.

Another essential component of apomixis is endosperm formation, which is the nutritional supply for a viable embryo and seed. Endosperm develops in response to fertilization in sexually reproducing plants. In sporophytic apomicts, development of adventitious embryos usually depends on the endosperm formed in the coexisting sexual embryo sac within the same ovule (Koltunow and Grossniklaus, 2003). Thus, endosperm development in sporophytic apomixis is regulated by the same signal pathway as in sexual reproduction. In pseudogamous (fertilization-dependent) gametophytic apomicts, fertilization of the central cell is required for apomictic endosperm development, similar to that in sexual plants. However, autonomous endosperm development occurs in some gametophytic apomicts (Koltunow and Grossniklaus, 2003), which obviously requires a modified signal pathway compared to that in sexually reproducing plants. Functional screenings for mutants that allow autonomous endosperm development in *Arabidopsis* have led to isolation of genes that repress endosperm development in the absence of fertilization in sexually reproducing plants, which might also control autonomous endosperm development in apomicts (Chaudhury et al., 1997; Grossniklaus et al., 2001). The *fis* mutants of *Arabidopsis* showed some degree of seed development without pollination, which suggested that the products of the *Fertilization-Independent Seed* (*FIS*) genes were likely to play repressive regulatory roles in the development of seed before pollination

during normal sexual reproduction (Chaudhury et al., 1997). The maternal loss-of-function mutant *medea* of *Arabidopsis* shows aberrant growth during embryogenesis and led to isolation of the *MEDEA* (*MEA*) gene (Grossniklaus et al., 1998). *MEA* encodes a SET-domain containing Polycomb group (PcG) protein, which is required for stable inheritance of expression patterns through cell division and regulates cell proliferation in animals. Comparative expression analysis of maternal and paternal *MEA* alleles showed that only the maternal *MEA* allele was required for proper endosperm development and only the maternal *MEA* mRNA was detected in the endosperm while both maternal and paternal *MEA* alleles were expressed in the embryo (Kinoshita et al., 1999). Further functional analysis of *FIS1*, which was found to be identical to *MEA*, and *FIS2* genes suggested that altered regulation of these genes and their homologs might play a role in apomictic seed development (Luo et al., 1999). Mutation of *MEA* led to embryo arrest in 50% of seeds after fertilization, which suggested a key role of *MEA* in seed development. *F644*, a gene also identified from an *Arabidopsis* mutant and found to be identical to *MEA*, also allowed fertilization-independent endosperm development and therefore further confirmed the function of *MEA* during endosperm development (Kiyosue et al., 1999). Characterization of *FERTILIZATION-INDEPENDENT ENDOSPERM* (*FIE*), isolated from *Arabidopsis*, revealed that *FIE* was a homolog of the WD motif-containing PcG proteins which were isolated from animals (Ohad et al., 1999). Loss-of-function mutant *fie* was able to initiate endosperm development without fertilization (Ohad et al., 1996) suggesting the function of FIE Polycomb protein as a suppressor for endosperm development before fertilization in sexually reproducing plants. Taken together, the results suggest that autonomous endosperm development in apomicts may require inactivation or modified regulation of the PcG complexes.

Dosage Effect and Endosperm in Apomicts

In terms of maternal and paternal contributions to the endosperm, endosperm in an apomict is different from that in the sexually reproduced seed. In sexual plants, endosperm develops from the fertilization of the reduced central cell, which contains two reduced nuclei, by the reduced sperm. Therefore, the ratio of maternal and paternal genomic contribution is 2m:1p in a diploid, which seems to be critical for viable seed development in many sexually reproducing plants. However, in apomictic plants the relative genetic maternal and paternal contributions to endosperm are variable because of the unreduced nature and variable number of central cell nuclei in different apomictic species (Ozias-Akins, 2006). The dosage effect in the endosperm is a possible barrier for introducing apomixis into sexual relatives because of the seed abortion problem (Morgan et al., 1998). Since the apomicts still produce viable seeds, the requirement of fixed maternal and paternal genomic ratios in the endosperm may be less stringent in apomicts (Grimanelli, 1997) or the balance is restored by modifications such as DNA methylation (Spielman et al., 2003). Genome-wide demethylation analysis of *Arabidopsis* endosperm with application of the Illumina Genome Analyzer platform to quantify DNA methylation showed that the entire endosperm genome is demethylated (Hsieh et al., 2009). Expression of genes with reduced DNA methylation was compared between endosperm and embryo to determine the effect of methylation changes in endosperm gene expression. Results showed that genes exhibiting reduced methylation within 1 kb of the 5' upstream of transcription were preferentially expressed in the endosperm whereas genes demethylated near the 3' end did not show a significant change in expression. Thus it is suggested that endosperm demethylation likely reinforces gene silencing in the embryo (Hsieh et al., 2009).

Polyploidy and Apomixis

Genetic analysis of apomixis is further complicated by the polyploid nature of apomicts. Gametophytic apomicts, including both aposporous and diplosporous apomicts, are almost exclusively polyploids (Nogler, 1984). The theory that expression of apomixis may depend on the polyploid genome has been proposed with the example from *Paspalum* (Quarin et al., 2001). However, diploid gametophytic apomicts have been reported (Bicknell, 1997; Kojima, 1997; Kantama et al., 2007), which suggests that polyploidy is not absolutely required for the expression of apomixis. Nevertheless, the frequency of unreduced embryo sacs and seed formation in these diploid plants was much lower compared to that in the polyploids, which suggests that the higher ploidy level may enhance the expression of apomixis rather than ensure its presence (Bicknell and Koltunow, 2004; Ozias-Akins, 2006). Another hypothesis combining interspecific hybridization with polyploidy was postulated to play critical regulatory roles during megasporogenesis, megagametogenesis and fertilization during the occurrence of apomixis (Carman, 1997; Roche et al., 2001). The possible explanation for this mechanism is that hybridization and increased ploidy level affect the status of DNA methylation and then cause asynchronous expression of the reproductive regulatory genes. A more recent reproduction mode study in *Hypericum* by flow cytometric seed screening (FCSS) and chromosome counts showed that the relative DNA content of apomicts was significantly increased either by polyploidization or by both polyploidization and increase DNA content per chromosome (Matzk, 2003). A possible reason for the higher genetic load on the chromosomes might be accumulation of retroelements. It has been reported that simple chromosome doubling of sexual diploid *Paspalum* produced apomictic autotetraploids (Quarin et al., 2001), which suggested the genome duplication might involve in apomixis induction. Other studies in maize and *Arabidopsis* seem

to support the model that polyploidy alters the expression pattern of reproductive regulatory genes. Twenty out of 700 genes corresponding to sexual diploid *Arabidopsis* examined were suppressed in the allotetraploids (Comai et al., 2000). Dosage regulation was examined in a maize ploidy series and the expression data from 18 genes tested suggests that the absolute level of gene expression increases as structural gene dosage increases (Guo et al., 1996; Edger and Pires, 2009). However, it is clear that polyploidy alone is not sufficient for apomixis based on the fact that there are more than half of the angiosperm species that are polyploids but most of them reproduce sexually (Grimanelli et al., 2001). It has been argued that apomixis is transmitted preferentially through diploid gametes rather than haploid gametes (Grimanelli et al., 1998; Tas and Van Dijk, 1999). This would result in the establishment of polyploidy among apomicts. Therefore, the increased genome size of apomicts might be the consequence of apomixis rather than the cause for apomixis. It is still unclear about the role of polyploidy in the apomixis pathway.

High-throughput Next-generation Sequencing Technologies

In the last three decades, the Sanger enzymatic dideoxynucleotide DNA sequencing technique (Sanger et al., 1977) has been the dominant tool for deciphering complete genes and later entire genomes. Despite the continuous improvements such as the application of capillary electrophoresis systems and the decreasing costs, the Sanger sequencing protocol has been shown to have limitations for routine large sequence output. These limitations include the need for gels or polymers used to separate the fluorescently-labeled DNA fragments, the relatively low number of samples which could be analyzed in parallel and the difficulty of total automation of the sample preparation methods (Ansorge, 2009). Triggered by the human genome sequencing project, demand for faster and more affordable sequencing technologies and many

years of hard work of the scientists has led to the development of next-generation sequencing technologies.

The principle of 'sequencing-by-synthesis' technique developed at the European Molecular Biology Laboratory (EMBL), which allowed parallel sequencing by means of a sensitive charge-coupled device (CCD) camera, was one of the original developments applied extensively in today's high-throughput DNA sequencers (Ansorge, 2009). Beginning in 2004, the next-generation sequencers have become commercially available and their capability of producing millions of DNA sequence reads in a single run is having a major impact on genome-wide biological research.

The commercially available next-generation sequencers share some common features (Mardis, 2008). One of the important features is the ability to process millions of sequence reads at a time. Furthermore, the sequence reads are produced from fragment 'libraries' without the conventional vector-based cloning and *Escherichia coli*–based amplification stages. This will avoid, to a certain degree, the bias of representation in sequencing projects. Moreover, relatively little amount of starting material is needed for sequencing. This feature makes transcriptome or DNA content analysis of specific cell types much easier and more feasible. However, current commercially available next-generation sequencers all produce shorter read lengths than the conventional capillary sequencers, which may impact the assembly of the reads, especially for *de novo* sequencing projects. Currently there are three major platforms for massively parallel DNA sequencing in widespread use, which are the Roche/454 FLX Pyrosequencer, the Illumina/Solexa Genome Analyzer and the Applied Biosystems SOLiD™.

The Roche/454 FLX Pyrosequencer

The first revolution for DNA sequencing was the pyrosequencing technology developed by 454 Life Sciences based on the principle of 'sequencing-by-synthesis'. The key step for this platform is that the pyrophosphate molecule released on incorporation of a nucleotide by DNA polymerase triggers a series of downstream enzymatic reactions to ultimately produce light (Margulies et al., 2005). The Roche/454 FLX Pyrosequencer avoids the labor intensive cloning step of the Sanger sequencing by taking advantage of emulsion PCR. Instead of sequencing in individual tubes or in microtiter plate wells, the DNA fragments for sequencing are amplified on the surface of agarose beads. Individual DNA fragments for sequencing are ligated to adaptors, whose sequences are complementary to the oligomers attached to the surface of the agarose beads. Then each template-containing bead is subsequently transferred into a well containing the PCR reactants and the DNA fragments are amplified. Imaging of the light flashes from luciferase activity records the sequential flow of nucleotide added during the amplification of the DNA template. The signals recorded during the run for each bead on the 454 picotiter plate are translated into a sequence read. The raw reads are processed and then screened to remove poor-quality sequences, mixed sequences (more than one initial DNA fragment per bead), and sequences without the initiating adaptor sequence prior to the final sequencing readout. The current 454 instrument, the GS-FLX Titanium, produces an average read length of 400bp per bead, with a combined throughput of 400-600 Mb of sequence data in a 10-h run (Pettersson et al., 2009).

The Illumina/Solexa Genome Analyzer

Based on a same principle of 'sequencing-by-synthesis' as the Roche/454 FLX Pyrosequencer, the Solexa sequencing platform utilizes fluorescently labeled nucleotides to

detect incorporation of bases into DNA and is capable of producing sequence reads from tens of millions of surface-amplified DNA fragments simultaneously (Bentley, 2006; Morozova and Marra, 2008; Ansorge, 2009). The Illumina process starts from a mixture of single-stranded and adaptor-ligated DNA fragments. One end of the DNA fragment is attached to the solid surface of a glass flow cell and the other end subsequently bends over and hybridizes to a complementary adaptor, thereby forming the bridge-template for synthesis of the complementary strand. Bridge amplification of fragments happens by incubation with reactants on the surface. A reaction mixture containing primers, four nucleotides each labeled with a different fluorescent dye and polymerase are added simultaneously onto the surface of flow cell channels. Because the 3′-OH group of all the four labeled nucleotides is chemically inactivated, only a single base is incorporated per cycle. Each base incorporation cycle at each cluster is detected and identified by the CCD camera via the fluorescent dye. After the image recording step, the fluorescent group of the incorporated nucleotide is removed and the 3′-OH group is reactivated by a chemical reaction followed by another synthesis cycle. The average sequence read length of the current system, the Illumina Genome Analyzer IIx, offers a combination of 100 bp read length and >300 Mb per flow cell in about ten days (http://www.illumina.com/systems/genome_analyzer.ilmn).

The Applied Biosystems SOLiD System

Similar to other next-generation sequencing platforms, the Sequencing by Oligo Ligation and Detection (SOLiD) platform starts with an adaptor-ligated fragment library (Mardis, 2008; Morozova and Marra, 2008; Ansorge, 2009). Each fragment binds to one magnetic bead coated with complementary oligos to the adaptors and becomes amplified by emulsion PCR. The beads with the amplification products are covalently attached onto a glass surface for sequencing.

Different from other sequencing platforms, the SOLiD System uses DNA ligase and a unique approach to sequence the amplified fragments. Ligase-mediated sequencing starts with the annealing of a complementary sequencing primer to the adaptor. Then mixture of octamer oligonucleotides, in which the fourth and fifth bases are fluorescently labeled, and ligation mixture are added. When an octamer hybridizes to the DNA fragment sequence adjacent to the sequencing primer 3′ end, DNA ligase seals the phosphate backbone followed by fluorescence detection to determine bases 4 and 5 in the sequence. After the fluorescence detection, the ligated octamers are cleaved off after the fifth base by a chemical cleavage step and the fluorescent groups are removed. The hybridization and ligation process repeats to determine bases 9 and 10 in the DNA fragment sequence, and so on. An entire run can produce 3-4 Gb of DNA sequence data with an average read length of 25-35 bp in about 5 days.

Genetic Analysis Applications of Next-generation Sequencing

The next-generation sequencing technologies are capable of producing millions of DNA sequence reads in a single run. The bacterial cloning step in traditional sequencing approaches is avoided by sequencing the amplified single-strand fragment library. However, high-throughput comes at the expense of read lengths, which provides potential problems for sequence assembly particularly in areas rich in repeat sequences. Moreover, the huge amount of data generated by the high-throughput sequencing run in the form of short reads is a considerable challenge for processing and analysis of the data. Nevertheless, this technology has found broad sequence-based applications, in which determining the sequence of the whole DNA molecule is not essential.

Genomic sequencing

The high-throughput, accuracy and lower cost offered by next-generation sequencing approaches have tremendously facilitated genomic sequencing as originally expected. Application of the new sequencing technologies in human genome sequencing is successful and it makes personal genomics closer to reality (Wheeler et al., 2008). Except for the original whole genome sequencing purpose, high-throughput sequencing has been shown to be highly sensitive in detection of variations like single nucleotide polymorphisms (SNPs) in specific genomic regions or genome-wide sequences (Thomas et al., 2006; Barbazuk et al., 2007; Imelfort et al., 2009). Also, sequencing the bisulfite DNA by using next-generation technology has great potential in the research of genomic DNA methylation profiling (Taylor et al., 2007).

Chromatin Immunoprecipitation (ChIP) sequencing

Chromatin immunoprecipitation (ChIP) has been a common approach for studying regulatory DNA–protein binding interactions since its first introduction in 1988 (Solomon et al., 1988). Basically, the DNA fragments binding to the precipitated proteins are released and subsequently identified either by Southern blotting or by quantitative PCR (qPCR). For the purpose of genome-wide analysis of DNA–protein binding sites, an advanced version of the ChIP approach, ChIP-chip, was developed by combining ChIP with microarray analysis (Ren et al., 2000; Horak and Snyder, 2002). The ChIP-chip approach is a useful tool for analyzing the DNA associated with a protein of interest by comparison to a set of reference sequences on a microarray. It greatly facilitates the investigation of gene regulatory networks by enabling large-scale analysis of transcription factor binding sites. However, dependency of ChIP-chip on an available set of reference sequences limits its applications in non-model organisms. Development of next-generation sequencing technologies will address this limitation by

providing more available genomic reference sequences and more importantly, replacing the microarray-based readout with direct DNA sequencing of the released fragments from ChIP experiments, which is known as ChIP sequencing (ChIP-seq) approach (Impey et al., 2004; Robertson et al., 2007). It is likely that ChIP-seq will significantly contribute to our understanding of regulation of protein binding sites and improve the annotation of binding sites in genome-wide fashion (Mardis, 2008; Morozova and Marra, 2008).

Gene expression analysis aided by next-generation sequencing

Serial analysis of gene expression (SAGE) is a powerful approach for genome-wide analysis of gene expression and it has been applied in transcriptome analysis of different organisms (Velculescu et al., 1995; Fregene et al., 2004). However, the application of SAGE is limited by the cost of DNA sequencing, which is part of the SAGE protocol. High throughput, high accuracy and lower cost features offered by next-generation sequencing instruments have been shown to be the perfect combination with SAGE technique for gene expression studies (Cheung et al., 2006; Weber et al., 2007; Sharbel et al., 2009).

In the last two decades, expressed sequence tags (ESTs) have offered opportunities for gene discovery and examination of differences in gene expression across a variety of samples such as tissue types and life-cycle stages in different organisms, including human (Adams et al., 1991; Park et al., 1993; Umeda et al., 1994; Lim et al., 1996; Li et al., 2009). Large sets of expressed sequence tags (ESTs) have been applied for microarray analyses to provide information on the relative expression levels of large numbers of gene transcripts (Li et al., 2009). However, microarray analysis has typically focused on the study of model organisms due to the lack of sequence data for non-model organisms. With application of next-generation sequencing technologies, the recent availability of large sets of ESTs offers the opportunity to

consider designing and applying microarray technology to a larger and more diverse set of species, and more and more ESTs will be available for genome-wide mRNA profiling studies. The successful examples include *Arabidopsis thaliana* (Weber et al., 2007; Eveland et al., 2008), *Medicago truncatula* (Cheung et al., 2006), *Zea mays* (Ohtsu et al., 2007), as well as *Drosophila melanogaster* (Torres et al., 2008). Another application of next-generation sequencing in gene expression analysis is in discovery of noncoding RNAs (ncRNAs). Research of ncRNAs including microRNAs (miRNAs) has been reported to provide new insights into gene regulation in both plant and animal (Andersen and Panning, 2003; Chen, 2005; Du and Zamore, 2005; Pillai, 2005; Garzon et al., 2006; Winston, 2009; Zhang et al., 2009). Sequencing is the best way for noncoding RNA discovery because of the evolutionary diversity of ncRNA sequences. The high capacity and relatively low cost of next-generation sequencing technologies are promoting the discovery of ncRNAs with a rapid pace (Ruby et al., 2006; Kasschau et al., 2007). Moreover, the relatively quantitative readout from next-generation sequencers will provide insights into ncRNA characterization and therefore facilitate the annotation of sequenced genomes (Mardis, 2008).

Conclusions and Prospects

Today, the large number of short reads produced by next-generation sequencers is offering great opportunities for genetic and biological research to individual laboratories in addition to larger genome centers. They have shown great potential with broad applications in genomic, proteomic and medical research. However, more studies are needed to improve the robustness of the applications as well as lower the cost for many applications. Another aspect that deserves more efforts for improvement is the read length, which is the major drawback that limits their applications, especially in *de novo* sequencing. New technology with capability to

provide both high-throughput and longer read length will largely expand the applications in genomic and genetic research.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252:** 1651-1656

Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in Pennisetum squamulatum reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. Plant Physiol **134:** 1733-1741

Akiyama Y, Hanna WW, Ozias-Akins P (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. Theor Appl Genet **111:** 1042-1051

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

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

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

Andersen AA, Panning B (2003) Epigenetic gene regulation by noncoding RNAs. Curr Opin Cell Biol **15:** 281-289

Ansorge WJ (2009) Next-generation DNA sequencing techniques. N Biotechnol **25:** 195-203

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J **51:** 910-918

Bashaw EC, Hanna WW (1990) Apomictic reproduction. *In* GP Chapman, ed, Reproductive Versatility in the Grasses. Cambridge University Press, Cambridge, pp 100-130

Bemer M, Wolters-Arts M, Grossniklaus U, Angenent GC (2008) The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. Plant Cell **20:** 2088-2101

Bentley DR (2006) Whole-genome re-sequencing. Curr Opin Genet Dev **16:** 545-552

Bicknell RA (1997) Isolation of a diploid, apomictic plant of Hieracium aurantiacum. Sexual Plant Reproduction **10:** 168-172

Bicknell RA, Borst NK, Koltunow AM (2000) Monogenic inheritance of apomixis in two Hieracium species with distinct developmental mechanisms. Heredity **84 (Pt 2):** 228-237

Bicknell RA, Koltunow AM (2004) Understanding apomixis: recent advances and remaining conundrums. Plant Cell **16 Suppl:** S228-245

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, Custers JB, van Lookeren Campagne MM (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. Plant Cell **14:** 1737-1749

Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Linn Soc **61:** 51-94

Casson S, Spencer M, Walker K, Lindsey K (2005) Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. Plant J **42:** 111-123

Catanach AS, Erasmuson SK, Podivinsky E, Jordan BR, Bicknell R (2006) Deletion mapping of genetic regions associated with apomixis in Hieracium. Proc Natl Acad Sci U S A **103:** 18650-18655

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997) Fertilization-independent seed development in Arabidopsis thaliana. Proc Natl Acad Sci U S A **94:** 4223-4228

Chen L, Guan L, Seo M, Hoffmann F, Adachi T (2005) Developmental expression of ASG- 1 during gametogenesis in apomictic guinea grass (Panicum maximum). J Plant Physiol **162:** 1141-1148

Chen L, Miyazaki C, Kojima A, Saito A, Adachi T (1999) Isolation and characterization of a gene expressed during early embryo sac development in apomictic guinea grass (*Panicum maximum*). J Plant Physiol **154:** 55-62

Chen X (2005) MicroRNA biogenesis and function in plants. FEBS Lett **579:** 5923-5931

Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD (2006) Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. BMC Genomics **7:** 272

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B (2000) Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. Plant Cell **12:** 1551-1568

Conner J, Liu Z (2000) LEUNIG, a putative transcriptional corepressor that regulates AGAMOUS expression during flower development. Proc Natl Acad Sci U S A **97:** 12902-12907

Crane CF (2001) Classification of Apomictic Mechanisms. *In* Y Savidan, Carman, J.G., Dresselhaus, T., ed, The Flowering of Apomixis: From Mechanisms to Genetic Engineering. CIMMYT, IRD, European Commission DG VI (FAIR), Mexico, D.F., pp 24-43

Crawford BC, Yanofsky MF (2008) The formation and function of the female reproductive tract in flowering plants. Curr Biol **18:** R972-978

Davies B, Schwarz-Sommer Z (1994) Control of floral organ identity by homeotic MADS-box transcription factors. Results Probl Cell Differ **20:** 235-258

de Meeus T, Prugnolle F, Agnew P (2007) Asexual reproduction: genetics and evolutionary aspects. Cell Mol Life Sci **64:** 1355-1372

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

Du T, Zamore PD (2005) microPrimer: the biogenesis and function of microRNA. Development **132:** 4645-4652

Dujardin M, Hanna WW (1984) Cytogenetics of double cross hybrids between Pennisetum americanum -P. purpureum amphiploids and P. americanum X Pennisetum squamulatum interspecific hybrids. Theoretical and Applied Genetics **69:** 97-100

Dujardin M, Hanna WW (1989) Developing apomictic pearl millet-characterization of a BC_3 plant. J Genet Breed **43:** 145-151

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res **17:** 699-717

Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQ, Gerentes D, Perez P, Smyth DR (1996) AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. Plant Cell **8:** 155-168

Eveland AL, McCarty DR, Koch KE (2008) Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. Plant Physiol **146:** 32-44

Franklin-Tong VE (1999) Signaling and the modulation of pollen tube growth. Plant Cell **11:** 727-738

Fregene M, Matsumura H, Akano A, Dixon A, Terauchi R (2004) Serial analysis of gene expression (SAGE) of host-plant resistance to the cassava mosaic disease (CMD). Plant Mol Biol **56:** 563-571

Gaj MD, Zhang S, Harada JJ, Lemaux PG (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. Planta **222:** 977-988

García R, Asíns MJ, Forner J, Carbonell EA (1999) Genetic analysis of apomixis in Citrus and Poncirus by molecular markers. Theor Appl Genet **99:** 511-518

Garzon R, Fabbri M, Cimmino A, Calin GA, Croce CM (2006) MicroRNA expression and function in cancer. Trends Mol Med **12:** 580-587

Goel S, Chen Z, Conner JA, Akiyama Y, Hanna WW, Ozias-Akins P (2003) Delineation by fluorescence in situ hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in Pennisetum squamulatum and Cenchrus ciliaris. Genetics **163:** 1069-1082

Golubovskaya I, Avalkina N, Sheridan WF (1997) New insights into the role of the maize ameiotic1 locus. Genetics **147:** 1339-1350

Golubovskaya I, Grebennikova ZK, Avalkina NA, Sheridan WF (1993) The role of the ameiotic1 gene in the initiation of meiosis and in subsequent meiotic events in maize. Genetics **135:** 1151-1166

Grimanelli D, Hernández, M., Perotti, E., and Yves Savidan (1997) Dosage effects in the endosperm of diplosporous apomictic Tripsacum (Poaceae). Sexual Plant Reproduction **10:** 279-282

Grimanelli D, Leblanc O, Espinosa E, Perotti E, Gonzalez de Leon D, Savidan Y (1998) Non-Mendelian transmission of apomixis in maize-Tripsacum hybrids caused by a transmission ratio distortion. Heredity **80 (1):** 40-47

Grimanelli D, Leblanc O, Perotti E, Grossniklaus U (2001) Developmental genetics of gametophytic apomixis. Trends Genet **17:** 597-604

Grossniklaus U, Moore JM, Gagliano WB (1998) Molecular and genetic approaches to understanding and engineering apomixis: *Arabidopsis* as a powerful tool. *In* EAS S.S. Virmani, K. Muralidharan, ed, Advances in hybrid rice technology: proceedings of the 3rd International Symposium on Hybrid Rice. International Rice Research Institute, Los Baños, pp 187-211

Grossniklaus U, Nogler GA, van Dijk PJ (2001) How to avoid sex: the genetic control of gametophytic apomixis. Plant Cell **13:** 1491-1498

Grossniklaus U, Schneitz K (1998) The molecular and genetic basis of ovule and megagametophyte development. Semin Cell Dev Biol **9:** 227-238

Grossniklaus U, Spillane C, Page DR, Kohler C (2001) Genomic imprinting and seed development: endosperm formation with and without sex. Curr Opin Plant Biol **4:** 21-27

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science **280:** 446-450

Guo M, Davis D, Birchler JA (1996) Dosage effects on gene expression in a maize ploidy series. Genetics **142:** 1349-1355

Hanna WW (1995) Use of apomixis in cultivar development. Adv Agron **54:** 333–350

Hanna WW, Bashaw EC (1987) Apomixis: its identification and use in plant breeding. Crop Science **27:** 1136–1139

Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt ED, Boutilier K, Grossniklaus U, de Vries SC (2001) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol **127:** 803-816

Hong QB, Xiang SQ, Chen KL, Chen LG (2001) [Two complementary dominant genes controlling apomixis in genus Citrus and poncirus]. Yi Chuan Xue Bao **28:** 1062-1067

Horak CE, Snyder M (2002) ChIP-chip: a genomic approach for identifying transcription factor binding sites. Methods Enzymol **350:** 469-483

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D (2009) Genome-wide demethylation of Arabidopsis endosperm. Science **324:** 1451-1454

Imelfort M, Duran C, Batley J, Edwards D (2009) Discovering genetic polymorphisms in next-generation sequencing data. Plant Biotechnol J **7:** 312-317

Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH (2004) Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. Cell **119:** 1041-1054

Jessup RW, Burson BL, Burow GB, Wang YW, Chang C, Li Z, Paterson AH, Hussey MA (2002) Disomic inheritance, suppressed recombination, and allelic interactions govern apospory in buffelgrass as revealed by genome mapping. Crop Science **42:** 1688-1694

Kantama L, Sharbel TF, Schranz ME, Mitchell-Olds T, de Vries S, de Jong H (2007) Diploid apomicts of the Boechera holboellii complex display large-scale chromosome substitutions and aberrant chromosomes. Proc Natl Acad Sci U S A **104:** 14026-14031

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol **5:** e57, 0479-0493

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell **11:** 1945-1952

Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB, Fischer RL (1999) Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. Proc Natl Acad Sci U S A **96:** 4186-4191

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

Kojima A, Nagato, Y. (1997) Discovery of highly apomictic and highly amphimictic dihaploids in Allium tuberosum. Sexual Plant Reproduction **10:** 8-12

Koltunow AM (1993) Apomixis: Embryo Sacs and Embryos Formed without Meiosis or Fertilization in Ovules. Plant Cell **5:** 1425-1437

Koltunow AM, Bicknell RA, Chaudhury AM (1995) Apomixis: Molecular Strategies for the Generation of Genetically Identical Seeds without Fertilization. Plant Physiol **108:** 1345-1352

Koltunow AM, Grossniklaus U (2003) Apomixis: a developmental perspective. Annu Rev Plant Biol **54:** 547-574

Kumlehn J, Kirik V, Czihal A, Altschmied L, Matzk F, Lörz H, Bäumlein H (2001) Parthenogenetic egg cells of wheat: cellular and molecular studies Sexual Plant Reproduction **14:** 239-243

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

Leblanc O, Armstead I, Pessino SC, Ortiz JP, Evans C, Valle CD, Hayward MD (1997) Non-radioactive mRNA fingerprinting to visualise gene expression in mature ovaries of *Brachiaria* hybrids derived from *B. brizantha*, an apomictic tropical forage. Plant Science **126:** 49-58

Li H, Cao H, Wang Y, Pang Q, Ma C, Chen S (2009) Proteomic analysis of sugar beet apomictic monosomic addition line M14. J Proteomics **73(2):** 297-308

Li W, Olohan L, Williams D, Hughes M, Gracey A, Cossins A (2009) Application of ESTs in microarray analysis. Methods Mol Biol **533:** 289-309

Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwang I, Cho MJ (1996) Expressed sequence tags of Chinese cabbage flower bud cDNA. Plant Physiol **111:** 577-588

Lohe AR, Chaudhury A (2002) Genetic and epigenetic processes in seed development. Curr Opin Plant Biol **5:** 19-25

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM (1999) Genes controlling fertilization-independent seed development in Arabidopsis thaliana. Proc Natl Acad Sci U S A **96:** 296-301

Malho R, Trewavas AJ (1996) Localized Apical Increases of Cytosolic Free Calcium Control Pollen Tube Orientation. Plant Cell **8:** 1935-1949

Malik MR, Wang F, Dirpaul JM, Zhou N, Polowick PL, Ferrie AM, Krochko JE (2007) Transcript profiling and identification of molecular markers for early microspore embryogenesis in Brassica napus. Plant Physiol **144:** 134-154

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet **24:** 133-141

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet **9:** 387-402

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:** 376-380

Martinez EJ, Urbani MH, Quarin CL, Ortiz JP (2001) Inheritance of apospory in bahiagrass, Paspalum notatum. Hereditas **135:** 19-25

Matzk F, Hammer, K., Schubert, I. (2003) Coevolution of apomixis and genome size within the genus Hypericum. Sexual Plant Reproduction **16:** 51-58

Matzk F, Meister A, Brutovska R, Schubert I (2001) Reconstruction of reproductive diversity in Hypericum perforatum L. opens novel strategies to manage apomixis. Plant J **26:** 275-282

Matzk F, Prodanovic S, Baumlein H, Schubert I (2005) The Inheritance of apomixis in Poa pratensis confirms a five locus model with differences in gene expressivity and penetrance. Plant Cell **17:** 13-24

Mayer KF, Schoof H, Haecker A, Lenhard M, Jurgens G, Laux T (1998) Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. Cell **95:** 805-815

Mercier R, Armstrong SJ, Horlow C, Jackson NP, Makaroff CA, Vezon D, Pelletier G, Jones GH, Franklin FC (2003) The meiotic protein SWI1 is required for axial element formation and recombination initiation in Arabidopsis. Development **130:** 3309-3318

Mizukami Y, Ma H (1992) Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. Cell **71:** 119-131

Morcillo F, Gallard A, Pillot M, Jouannic S, Aberlenc-Bertossi F, Collin M, Verdeil JL, Tregear JW (2007) EgAP2-1, an AINTEGUMENTA-like (AIL) gene expressed in meristematic and proliferating tissues of embryos in oil palm. Planta **226:** 1353-1362

Morgan RN, Ozias-Akins P, Hanna W (1998) Seed set in an apomictic BC3 pearl millet. International Journal of Plant Sciences **159:** 89-97

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics **92:** 255-264

Naumova TN, Laak JVD, Osadtchiy J, Matzk F, Kravtchenko A, Bergervoet J, Ramulu KS, Boutilier K (2004) Reproductive development in apomictic populations of Arabis holboellii (Brassicaceae). Sexual Plant Reproduction **14:** 195-200

Nogler GA (1984) Gametophytic apomixis. *In* BM Johri, ed, Embryology of Angisoperms. Springer-Verlag, Berlin, pp 475–518

Noyes RD, Rieseberg LH (2000) Two independent loci control agamospermy (Apomixis) in the triploid flowering plant Erigeron annuus. Genetics **155:** 379-390

Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL (1996) A mutation that allows endosperm development without fertilization. Proc Natl Acad Sci U S A **93:** 5319-5324

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL (1999) Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell **11:** 407-416

Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, Chen T, Zhang X, Timmermans MC, Beck J, Buckner B, Janick-Buckner D, Nettleton D, Scanlon MJ, Schnable PS (2007) Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). Plant J **52:** 391-404

Otto SP, Lenormand T (2002) Resolving the paradox of sex and recombination. Nat Rev Genet **3:** 252-261

Ozias-Akins P (2006) Apomixis: Developmental Characteristics and Genetics. Critical Reviews in Plant Sciences **25:** 199-214

Ozias-Akins P, Roche D, Hanna WW (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. Proc Natl Acad Sci U S A **95:** 5127-5132

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

Pagnussat GC, Yu HJ, Sundaresan V (2007) Cell-fate switch of synergid to egg cell in Arabidopsis eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene BLH1. Plant Cell **19:** 3578-3592

Park YS, Kwak JM, Kwon OY, Kim YS, Lee DS, Cho MJ, Lee HH, Nam HG (1993) Generation of expressed sequence tags of random root cDNA clones of Brassica napus by single-run partial sequencing. Plant Physiol **103:** 359-370

Passarinho P, Ketelaar T, Xing M, van Arkel J, Maliepaard C, Hendriks MW, Joosen R, Lammers M, Herdies L, den Boer B, van der Geest L, Boutilier K (2008) BABY BOOM target genes provide diverse entry points into cell proliferation and cell growth pathways. Plant Mol Biol **68:** 225-237

Pawlowski WP, Wang CJ, Golubovskaya IN, Szymaniak JM, Shi L, Hamant O, Zhu T, Harper L, Sheridan WF, Cande WZ (2009) Maize AMEIOTIC1 is essential for multiple early meiotic processes and likely required for the initiation of meiosis. Proc Natl Acad Sci U S A **106:** 3603-3608

Pessino SC, Espinoza F, Martinez EJ, Ortiz JP, Valle EM, Quarin CL (2001) Isolation of cDNA clones differentially expressed in flowers of apomictic and sexual Paspalum notatum. Hereditas **134:** 35-42

Pettersson E, Lundeberg J, Ahmadian A (2009) Generations of sequencing technologies. Genomics **93:** 105-111

Pierson ES, Miller DD, Callaham DA, Shipley AM, Rivers BA, Cresti M, Hepler PK (1994) Pollen tube growth is coupled to the extracellular calcium ion flux and the intracellular calcium gradient: effect of BAPTA-type buffers and hypertonic media. Plant Cell **6:** 1815-1828

Pillai RS (2005) MicroRNA function: multiple mechanisms for a tiny RNA? RNA **11:** 1753-1761

Quarin CL, Espinoza F, Martinez EJ, Pessino SC, Bovo OA (2001) A rise of ploidy level induces the expression of apomixis in *Paspalum notatum*. Sexual plant reproduction **13:** 243-249

Ravi M, Marimuthu MP, Siddiqi I (2008) Gamete formation without meiosis in Arabidopsis. Nature **451:** 1121-1124

Ray A, Robinson-Beers K, Ray S, Baker SC, Lang JD, Preuss D, Milligan SB, Gasser CS (1994) Arabidopsis floral homeotic gene BELL (BEL1) controls ovule development through negative regulation of AGAMOUS gene (AG). Proc Natl Acad Sci U S A **91:** 5761-5765

Reiser L, Fischer RL (1993) The Ovule and the Embryo Sac. Plant Cell **5:** 1291-1301

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. Science **290:** 2306-2309

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones

S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods **4:** 651-657

Robinson-Beers K, Pruitt RE, Gasser CS (1992) Ovule Development in Wild-Type Arabidopsis and Two Female-Sterile Mutants. Plant Cell **4:** 1237-1249

Roche D, Hanna W, Ozias-Akins P (2001) Is supernumerary chromatin involved in gametophytic apomixis of polyploid plants? Sexual Plant Reproduction **13:** 343-349

Rodrigues JC, Cabral GB, Dusi DM, de Mello LV, Rigden DJ, Carneiro VT (2003) Identification of differentially expressed cDNA sequences in ovaries of sexual and apomictic plants of Brachiaria brizantha. Plant Mol Biol **53:** 745-757

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell **127:** 1193-1207

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:** 5463-5467

Savidan Y (2000) Apomixis: genetics and breeding. Plant Breed Rev **18:** 13–85

Schmidt ED, Guzzo F, Toonen MA, de Vries SC (1997) A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. Development **124:** 2049-2062

Schneitz K (1999) The molecular and genetic control of ovule development. Curr Opin Plant Biol **2:** 13-17

Schoof H, Lenhard M, Haecker A, Mayer KF, Jurgens G, Laux T (2000) The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. Cell **100:** 635-644

Sharbel TF, Voigt ML, Corral JM, Thiel T, Varshney A, Kumlehn J, Vogel H, Rotter B (2009) Molecular signatures of apomictic and sexual ovules in the Boechera holboellii complex. Plant J **58:** 870-882

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

Singh M, Burson BL, Finlayson SA (2007) Isolation of candidate genes for apomictic development in buffelgrass (Pennisetum ciliare). Plant Mol Biol **64:** 673-682

Skinner DJ, Hill TA, Gasser CS (2004) Regulation of ovule development. Plant Cell **16 Suppl:** S32-45

Sogo A, Jaffre T, Tobe H (2004) Pollen-tube growth and fertilization mode in Gymnostoma (Casuarinaceae): their characteristics and evolution. J Plant Res **117:** 249-251

Solomon MJ, Larsen PL, Varshavsky A (1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell **53:** 937-947

Spielman M, Vinkenoog R, Scott RJ (2003) Genetic mechanisms of apomixis. Philos Trans R Soc Lond B Biol Sci **358:** 1095-1103

Srinivasan C, Liu Z, Heidmann I, Supena ED, Fukuoka H, Joosen R, Lambalk J, Angenent G, Scorza R, Custers JB, Boutilier K (2007) Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (Nicotiana tabacum L.). Planta **225:** 341-351

Tas IC, Van Dijk PJ (1999) Crosses between sexual and apomictic dandelions (Taraxacum). I. The inheritance of apomixis. Heredity **83:** 707-714

Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, Caldwell CW, Shi H (2007) Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res **67:** 8511-8518

Theissen G, Saedler H (1995) MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. Curr Opin Genet Dev **5:** 628-639

Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, Yuza Y, Garraway LA, LaFramboise T, Lee JC, Shah K, O'Neill K, Sasaki H, Lindeman N, Wong KK, Borras AM, Gutmann EJ, Dragnev KH, DeBiasi R, Chen TH, Glatt KA, Greulich H, Desany B, Lubeski CK, Brockman W, Alvarez P, Hutchison SK, Leamon JH, Ronan MT, Turenchalk GS, Egholm M, Sellers WR, Rothberg JM, Meyerson M (2006) Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med **12:** 852-855

Torres TT, Metta M, Ottenwalder B, Schlotterer C (2008) Gene expression profiling by massively parallel sequencing. Genome Res **18:** 172-177

Tucker MR, Araujo AC, Paech NA, Hecht V, Schmidt ED, Rossell JB, De Vries SC, Koltunow AM (2003) Sexual and apomictic reproduction in Hieracium subgenus pilosella are closely interrelated developmental pathways. Plant Cell **15:** 1524-1537

Umeda M, Hara C, Matsubayashi Y, Li HH, Liu Q, Tadokoro F, Aotsuka S, Uchimiya H (1994) Expressed sequence tags from cultured cells of rice (Oryza sativa L.) under stressed conditions: analysis of transcripts of genes engaged in ATP-generating pathways. Plant Mol Biol **25:** 469-478

Van Dijk PJ, Tas IC, Falque M, Bakx-Schotman T (1999) Crosses between sexual and apomictic dandelions (Taraxacum). II. The breakdown of apomixis. Heredity **83:** 715-721

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science **270:** 484-487

Wang HZ, Hu B, Chen GP, Shi NN, Zhao Y, Yin QC, Liu JJ (2008) Application of Arabidopsis AGAMOUS second intron for the engineered ablation of flower development in transgenic tobacco. Plant Cell Rep **27:** 251-259

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol **144:** 32-42

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature **452:** 872-876

Willemse MTM, van Went, J.L. (1984) The female gametophyte. *In* BM Johri, ed, Embryology of Angiosperms. Springer-Verlag, Berlin, pp 159-196

Winston F (2009) A transcription switch toggled by noncoding RNAs. Proc Natl Acad Sci U S A **106:** 18049-18050

Yang WC, Ye D, Xu J, Sundaresan V (1999) The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev **13:** 2108-2117

Yu H, Ito T, Wellmer F, Meyerowitz EM (2004) Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. Nat Genet **36:** 157-161

Zhang S, Sun L, Kragler F (2009) The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation. Plant Physiol **150:** 378-387

Zhu H, Bi YD, Yu LJ, Guo DD, Wang BC (2009) Comparative proteomic analysis of apomictic monosomic addition line of Beta corolliflora and Beta vulgaris L. in sugar beet. Mol Biol Rep **36:**2093–2098

Zuo J, Niu QW, Frugis G, Chua NH (2002) The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. Plant J **30:** 349-359

CHAPTER 3

IDENTIFICATION OF ALIEN EXPRESSED GENES FROM APOSPOROUS OVULES BY COMPARATIVE SEQUENCING OF DONOR PARENT AND BACKCROSS TRANSCRIPTOMES

Abstract

Apomixis, asexual seed production in plants, holds great potential for agriculture as a means to fix hybrid vigor. Apospory is a form of apomixis where the embryo develops from an unreduced egg that is derived from a somatic nucellar cell, the aposporous initial, via mitosis. Understanding the molecular mechanism regulating aposporous initial specification will be the most critical step toward elucidation of apomixis. Toward this end, two transcriptomes, which were derived from ovules of an apomictic donor parent and its apomictic backcross derivative at the stage of apospory initiation, were sequenced using 454-FLX technology. Comparison of the two transcriptome sequences identified up to 61 transcripts linked to the ASGR-carrier chromosome. Only one of these alien expressed genes could be assigned to the ASGR by screening a limited number of apomictic and sexual F_1s. These identified candidate genes, if mapped to the ASGR, could provide significant insight into the regulation of apospory initiation. Our results suggested that the strategy of comparative sequencing of donor parent and backcross transcriptomes was an efficient method to identify alien expressed genes in a recurrent parent background.

Introduction

Apomixis, asexual reproduction through seed, is widespread among flowering plant families, but low in its frequency of occurrence (Nogler, 1984). Different from sexual reproduction, apomictically derived embryos develop autonomously from unreduced ovular cells instead of through fertilization of a reduced egg by a sperm. Therefore, the progeny of an apomictic plant are genetically identical to the maternal plant (Koltunow, 1993; Grimanelli et al., 2001). This trait can be used as an advanced breeding tool in agriculture since it can enable fixation of hybrid vigor and seed propagation of desired genotypes (Hanna, 1995; Koltunow et al., 1995; Savidan, 2000; van Dijk and van Damme, 2000). However, no major agriculturally important crop possesses this trait (Carman, 1997; Bicknell and Koltunow, 2004; Ozias-Akins, 2006). Furthermore, introgression of apomixis into crops through crossing has been impeded by certain factors such as polyploidy and incompatibility (Bicknell and Koltunow, 2004). Therefore, discovery of genetic mechanisms underlying apomixis will be crucial for manipulation of apomixis to allow its introduction into target crops.

Apomixis has been classified into two types and three developmental pathways: gametophytic apomixis, including apospory and diplospory, and sporophytic apomixis, which is also known as adventitious embryony (Koltunow, 1993). In sporophytic apomixis, an embryo forms directly from an ovular cell and coexists with the zygotic embryo. For gametophytic apomixis, the embryo develops from an unreduced egg in an embryo sac derived through mitosis of either a somatic nucellar cell (apospory) or the megaspore mother cell (diplospory). In apospory, meiosis either does not complete or its products degenerate while aposporous initials (AIs) develop from one or more somatic nucellar cells. Both genotypes chosen for the present study are aposporous with the trait conferred by genetic elements from *Pennisetum*

squamulatum. Aposporous *P. squamulatum* has four-nucleate embryo sacs that lack antipodals (Ozias-Akins, 2006). Apospory in this species was reported to be inherited as a dominant Mendelian trait (Ozias-Akins et al., 1998) associated with an approximately 50 Mb, heterochromatic and hemizygous chromosomal region designated the aposporous specific genomic region (ASGR), (Goel et al., 2003; Akiyama et al., 2004).

Differential gene expression analysis is one approach for discovery of regulatory mechanisms and downstream effects associated with apomixis. Differential display is a method that has been successfully applied to detect apomixis- or sex-specific transcripts by comparison of gene expression between apomictic and sexual genotypes. In *Brachiaria*, ovaries at anthesis yielded two apomixis-specific fragments (Leblanc et al., 1997). In a later study with *Brachiaria* that focused on the earlier sporogenesis and gametogenesis stages, eleven differentially expressed fragments were identified (Rodrigues et al., 2003). Three expressed sequence tags (ESTs) from *Paspalum notatum*, all highly similar in sequence, showed differential expression in flowers between apomictic and sexual F_1 individuals after apospory initiation (Pessino et al., 2001). A related method, cDNA-AFLP has also been applied to isolate apomixis candidate genes in *Poa pratensis* where 179 transcript fragments from spikelets showed quantitatively different expression between apomictic and sexual genotypes (Albertini et al., 2004). The full length sequences of two genes of interest, *PpSERK* (SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE) and *APOSTART* were obtained by RACE (rapid amplification of cDNA ends) and their temporal and spatial expression patterns were assessed by reverse transcription polymerase chain reaction (RT-PCR) and in situ hybridization, respectively (Albertini et al., 2005). However, neither one of these two candidate genes showed apomictic- or sexual-specific expression, only quantitative differences in expression between apomictic and

sexual genotypes were observed (Albertini et al., 2005). cDNA library screening is another strategy used for identification of apomixis candidate genes. One apomict specific gene was identified from a *Panicum maximum* ovule cDNA library and shown to be expressed in both aposporous initials and embryos at four days after anthesis (Chen et al., 1999; Chen et al., 2005). By suppression subtractive hybridization two transcripts expressed in embryo sacs were isolated from *Pennisetum ciliare* ovary libraries showing either higher expression in an apomictic than sexual genotype or specific expression in apomixis (Singh et al., 2007). Even though many differentially expressed transcripts have been identified from apomictic species, none of the candidate genes has been shown to be genetically linked to the trait of apomixis.

The initiation of sexual and apomictic pathways likely is activated by different signals (Ozias-Akins and van Dijk, 2007). Formation of apospory initials is the first and most critical event for occurrence of apospory. Therefore, understanding the molecular mechanism underlying apospory initiation can provide insight into developmental regulation and downstream signaling that results in apomixis. In order to discover candidates for regulating aposporous initial specification in *P. squamulatum*, we compared two transcriptomes derived from microdissected live ovules at the stage of AI formation between the apomictic donor parent, *P. squamulatum*, and its apomictic derivative backcross 8 (BC_8) containing a single *P. squamulatum* chromosome. BC_8 was generated during the process of introducing apomixis from apomictic *P. squamulatum* into sexual tetraploid *P. glaucum* (pearl millet). Initially, a tetraploid *P. glaucum* x *P. squamulatum* F_1 was crossed with a *P. glaucum* x *P. purpureum* F_1 and hybrid apomictic individuals with good male fertility were selected (Dujardin, 1984). Subsequent backcrosses with tetraploid *P. glaucum* (Dujardin and Hanna, 1989) yielded a BC_8 line that was shown by FISH to contain only one chromosome from *P. squamulatum*. This single

chromosome common to both apomictic BC_8 and *P. squamulatum* was the ASGR-carrier chromosome based on the transmission of the trait of apomixis and linked molecular markers. We hypothesize that candidate genes for regulating aposporous initial specification localize to the ASGR, should function in both Ps26 and BC_8 at the same developmental stage, and would be identical in sequence since they are related by descent.

Recently, the development and commercialization of new massively parallel sequencing platforms have made transcriptome sequencing faster and more affordable. One platform, developed by 454 Life Sciences Corporation, the 454 GS-FLX sequencer, is capable of producing 100 Mb of sequence data with an average read length of 250 bp per bead in a 7-h run (Droege and Hill, 2008). Successful applications of these high-throughput sequencing technologies to transcriptome analysis have been reported (Emrich et al., 2007; Jones-Rhoades et al., 2007; Weber et al., 2007; Vera et al., 2008; Meyer et al., 2009; Wang et al., 2009). Here we present expressed sequence tags (ESTs) generated by Roche 454 high-throughput sequencing technology from dissected ovule tissues staged for aposporous initial formation from two apomictic lines chosen for their common features of apospory and single shared chromosome. Alien chromosome (ASGR-carrier chromosome) expressed transcripts were identified and are being tested as candidates for regulating apospory initial development.

Materials and Methods

Plant Materials

Pennisetum squamulatum (P_S26; PI 319196, 2n=56) and line 58 of backcross 8 (BC_8) were used for ovule collection. Compared with the BC_7 line which was used in previous studies (Goel et al., 2003), the BC_8 line contains only one alien chromosome from P_S26, the ASGR-carrier chromosome (Singh et al. 2009). *P. glaucum* (IA4X), *P. purpureum* (N37) and a small

segregating population of BC_8 (BC_8 is facultative thus it produces ~ 18% sexually derived offspring; (Singh et al. 2009) were used for assigning the candidate transcript fragments to the ASGR-carrier chromosome. A limited number of individuals from a segregating F_1 population between *P. squamulatum* and *P. glaucum* were used for mapping the transcript fragments to the ASGR.

RNA Isolation

Young florets were dissected from small inflorescence sections whose anthers were at stages between premeiosis and prophase, as determined by acetocarmine staining of anther squashes. One group of florets was stored in RNA*Later*® solution (Ambion, Austin, TX, USA) at 4°C while the other group was processed for ovary clearing by methyl salicylate (Young et al., 1979) to screen for the ovary developmental stage. Ovules from thirty cleared florets were examined for each group. If the cleared sample showed AIs in less than 30% of the ovaries and the remaining ovaries were at an earlier developmental stage, then florets stored in RNA*Later*® solution from the same section of inflorescence were used for ovule dissection. About 40 ovules per sample were collected and total RNA was extracted from the ovules with RNAqueous®-Micro Micro Scale RNA Isolation Kit (Ambion). RNA integrity and quantity were analyzed with an Agilent 2100 Bioanalyser (Agilent, Santa Clara, CA, USA) at the Interdisciplinary Center for Biotechnology Research (ICBR) of the University of Florida.

RNA Amplification and ds-cDNA Synthesis for Roche 454 Sequencing

With total RNA as starting material, mRNA was amplified by T7-based in vitro transcription following the manual of TargetAmp™ 2-Round aRNA Amplification Kit 2.0 (Epicentre, Madison, WI, USA). Size range and quantity of the amplified mRNA were measured by both gel electrophoresis and Agilent 2100 Bioanalyser analysis (Agilent). For each sample,

an equal amount of amplified mRNA from the three biological replicates was pooled for ds-cDNA synthesis following the protocol developed by the Schnabel lab (Nakazono et al., 2003). Size range and quantity of ds-cDNA were also analyzed by both gel electrophoresis and using the Agilent 2100 Bioanalyser (Agilent) before submitting the samples for sequencing.

454 Sequencing and Processing

About 6 μg of ds-cDNA from both Ps26 and BC_8 was submitted to the Genome Sequencing Center at Washington University for 454-FLX sequencing. Samples of cDNA were subjected to mechanical shearing (nebulization), size selected, and blunt-end fragments were ligated to short adaptors, which provided primer target sites for both amplification and sequencing. The Multifunctional Inertial Reference Assembly (MIRA) program (Chevreux et al., 2004) was used to process and assemble the sequences from each library. Adaptor sequences and low quality sequence reads were removed prior to assembly. The assembly was run as a *de novo*, 454 EST project with accurate assembly and polyA/T clipping. Each library of contig assemblies from Ps26 and BC_8 was converted to a database and analyzed with the BlastN program provided by the RCC (Research Computing Center) at the University of Georgia (http://rcc.uga.edu/index.html). With Ps26 library sequences as the query and BC_8 library as the database, BlastN analysis was performed with an E-value cutoff of ≤ 10e-100. The BlastN output was then parsed using an internal script such that only contigs showing 100% identity were selected for further analysis.

BLAST Analysis of the Selected Contigs

BlastX was used to analyze sequences that showed 100% identity over a region of at least 100 bp between Ps26 and BC_8 by searching against the NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/) databases. When there were no

significant hits of the sequences by BlastX, BlastN was then conducted to search for similar ESTs from other species, and then these ESTs were used as BlastX queries to search for putative encoding proteins.

Mapping of Identical Contigs to the Alien Chromosome and/or ASGR

Using internal scripts, ACE and Fasta files containing the contig sequences of interest from both P_S26 and BC_8 libraries were generated. The subsets of contig sequences were assembled *de novo* with the ContigExpress program in Vector NTI.10 (Invitrogen, Carlsbad, CA, USA). Reassembled contigs from this sequence subset were used as queries with BlastN against the respective Ps26 and BC_8 MIRA-assembled databases at an E-value cutoff of 10e-25. The BlastN results were parsed and used to help estimate the 'uniqueness' of the contig within the transcriptome. Primers were designed based on the overlapping region of Ps26 and BC_8 contigs, in some cases including further 3' sequences for primer design when the contigs were unique in both databases. When there were multiple contigs from each database showing high similarity to each other, primers were designed based on the region with the best polymorphisms to distinguish one from another. Primers were first tested for amplification with Ps26, IA4X, N37 and a small segregating population of BC_8 progenies. Those primers which did not amplify with IA4X and sexual BC_8 individuals were used for further screening with apomictic and sexual F_1s to test for linkage to the ASGR.

Expression Pattern Analysis of Transcripts Mapped to the Alien Chromosome

Total RNA was extracted from a panel of BC_8 tissues including vegetative (leaf, root), and reproductive tissues at anthesis but before pollination (anther and ovary) with QIAGEN DNeasy® Plant Mini kit (QIAGEN, Valencia, CA, USA) following the manufacturer's protocol. First-strand cDNA was synthesized following the manufacturer's protocol of First-strand cDNA

Synthesis kit (Invitrogen). RT-PCR reactions were performed with the primer pairs used in genomic DNA screening in a total volume of 20 µl containing 1 µl of first-strand cDNA, 1 µM of each primer, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 1 unit of JumpStart™ *Taq* DNA polymerase (Sigma, St. Louis, MO, USA). Amplification of contaminating genomic DNA was tested by the inclusion of controls that omitted the reverse transcriptase enzyme from the cDNA synthesis reaction, e.g. no RT controls. The PCR reaction was denatured at 94°C for 5 min followed by 35 cycles of 94°C denaturation for 30 seconds, annealing for 30 seconds at respective temperatures, and 72°C extension for 1 min. RT-PCR products were separated on a 1.5% agarose gel and stained with ethidium bromide. Gel images were captured with the Molecular Imager Gel Doc XR System (Bio-Rad Laboratories, CA).

cDNA Library Construction

Ovaries and anthers collected from apomictic BC_8 around anthesis but prior to fertilization were frozen in liquid nitrogen. Total RNA was extracted with the RNeasy® Plant Mini kit (QIAGEN) and then poly A^+ RNA was purified from total RNA with Oligotex® mRNA Mini kit (QIAGEN) following the manufacturer's protocols. Yield of mRNA was quantified with a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). mRNA was used for double-stranded cDNA synthesis with ZAP-cDNA® Synthesis Kit following the manufacturer's protocol (Stratagene, La Jolla, CA, USA). Ligations, packaging, titering of the packaging reactions, and plaque lifts were conducted following the manufacturer's protocol of ZAP-cDNA® Gigapack® III Gold Cloning Kit (Stratagene).

cDNA library Screening for Target Genes

The apomictic BC_8 ovary and anther-enriched cDNA library was screened with α-^{32}P labeled probes for the transcripts mapping to the ASGR-carrier chromosome. The PCR

fragments amplified from apomictic BC_8 genomic DNA with the primers used for assigning a fragment to the ASGR-carrier chromosome were diluted and labeled with α-^{32}P by PCR in a total volume of 20 µl. The labeling reaction contained ~0.1 ng primary PCR fragment, 1.25 unit Jumpstart *Taq* DNA polymerase (Sigma), 0.25 µM of each primer, 0.5 mM dATP/dTTP/dGTP mixture, 5 µl of α-^{32}P-labeled dCTP (3000 Ci/mmol) and 1× PCR buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$). Probes were purified by passing through homemade Sephadex G-50 (Sigma) columns, which were assembled with Ultrafree®-MC Centrifugal Filter Units (Millipore, Bedford, MA, USA). Pre-hybridization of the membranes in hybridization buffer (0.5 M sodium phosphate, 7% SDS, 1 mM EDTA, pH 8.0) containing 0.1 mg ml^{-1} salmon sperm DNA, which was denatured in boiling water for 10 minutes and cooled on ice before adding to the hybridization solution, was conducted at 65°C for 4 h before addition of the labeled, denatured probe. Hybridization was conducted at 65°C overnight followed by three washes at the same temperature for 30 min each with the following buffers: 1) 1× SSC, 0.1% SDS; 2) 0.5× SSC, 0.1% SDS; 3) 0.1× SSC, 0.1% SDS. After the final wash, membranes were wrapped with plastic film and exposed to x-ray film overnight at -80°C prior to manually developing with Kodak® GBX Developer and Fixer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). Autoradiographs were aligned with the respective plates to recover hybridizing plaques with sterile glass pipettes. Recovered plaques were released in tubes containing 1.0 ml SM phage buffer (according to the formula in the manual of ZAP-cDNA® Gigapack® III Gold Cloning Kit) and 20 µl chloroform (Sigma). After overnight elution at 4°C, 1 µl SM buffer of each recovered sample was used for PCR to verify positive signals. Since the primary screening was carried out with a high density of plaque clones, the recovered positive plaques were purified after

secondary and tertiary screens at much lower densities. Single plaques showing positive hybridization signals were recovered in 500 ml SM buffer with 10 µl chloroform (Sigma) at 4°C.

Sequencing and Mapping of Candidate cDNA Clones to the ASGR

In vivo excision of single plaque clones was conducted using ExAssist® helper phage with SOLR® strain following the protocol in the manual of ZAP-cDNA® Gigapack® III Gold Cloning Kit (Stratagene). Single colonies containing the pBluescript double-stranded phagemid with the cloned cDNA insert were isolated and cultured in liquid Luria-Bertani (LB) media containing 100 µg mL^{-1} ampicillin at 37°C overnight. An aliquot of each culture was further grown in freeze broth containing 100 µg mL^{-1} ampicillin at 37°C overnight and then stored at -80°C before sending out for sequencing. Sequencing was conducted with M13 primers at the Laboratory for Genomics and Bioinformatics (University of Georgia, Athens, GA). Vector and bad quality sequences were trimmed from the original sequences with VectorNTI Advanced 8 (Invitrogen) and primers were designed with VectorNTI using the high quality cDNA sequences. Primers were then tested with apomictic and sexual F_1s for linkage to the ASGR as described above.

Results

Aposporous Ovule-enriched cDNA Samples for Sequencing

Ovules from Ps26 and BC_8 at the stage of aposporous initial formation (Figure 3.1) were manually dissected from pistils (Figure 3.2). The yield of total RNA from each collection of 40 ovules was approximately 20 ng; 15 ng were used for one-round of T7 RNA polymerase-based RNA amplification. The average yield from one round of amplification was 90 µg. Three biological replicates were collected and amplified for both Ps26 and BC_8. Equal amounts of amplified RNA from each replicate were combined and 15 µg amplified RNA was used for ds-

cDNA synthesis for each sample. The majority of the ds-cDNA synthesized from amplified RNA was distributed in a size range from 200 bp to 1000 bp (Figure 3.3).

Assembly of Sequences from Ps26 and BC_8 Aposporous Ovules

Two aposporous ovule transcriptomes, one from Ps26 and the other from BC_8, were sequenced using the high-throughput 454-FLX sequencer. The Ps26 transcriptome library contained 332,567 reads with an average read length of 147 base pair (bp) and the BC_8 transcriptome library contained 363,637 reads with an average read length of 142 bp. By the Multifunctional Inertial Reference Assembly (MIRA) program, 33,977 contigs were assembled from the Ps26 ovule transcriptome library and 26,576 contigs from the BC_8 ovule transcriptome library. The number of reads per contig ranged from 1 to 759 in Ps26 assemblies and 1 to 1661 in BC_8 assemblies with the majority having less than 30 reads per assembly in both cases. The numbers of singletons in Ps26 library and BC_8 are 176 and 78 respectively.

Identification of Alien Expressed Transcripts

When MIRA-assembled contigs from the two libraries were analyzed by BlastN with Ps26 sequences as queries and BC_8 sequences as the database, a total of 118 contigs from each library with 100% sequence identity in an overlapping region ≥100 bp between Ps26 and BC_8 was obtained. We created a subset fasta file containing the 236 Ps26/BC_8 contigs (118 from each library) for further overlap analysis. The 236 Ps26/BC_8 contigs were assembled using the Contig Express program in Vector NTI suite 10. Sixty-one inter-genotype contigs with no mismatches were assembled (Table 3.1). The average overlapping regions of the 61 inter-genotype contigs was 241 bp with a range from 181 bp to 419 bp and the average number of sequence reads in the assembled contigs was 13. The remaining contigs from the 236 Ps26/BC_8 contigs, while initially identified by BlastX as having 100% identity over a region >100 bp, did

not continue to share sequence similarity outside this region and therefore were not assembled together by Contig Express. BlastX searches against NCBI databases were carried out for the 61 Ps26/BC_8 inter-genotype contigs generated by Contig Express and best protein hits for 26 contigs were summarized in Table 3.1. Because the sequences are 3' biased, a BlastN analysis against the expressed sequence tags (ESTs) database at NCBI with the remaining 34 Ps26/BC8 contigs was done to find potential orthologs from other species. At an E-value cutoff of e-40, 20 contigs had EST hits (Table 3.1). Two additional contigs had EST hits with a less stringent E-value. A BlastX was performed using the other EST sequences to determine if tentative protein functions could be obtained. The best hits of the BlastX were listed in table 3.1. The remaining 13 (21%) contigs did not have hits by either BlastX or BlastN; therefore, they were considered orphan genes.

If needed, up to four primer pairs were used to test for linkage of the 61 contigs to the ASGR-carrier chromosome, primer pairs were designed based on the 61 Ps26 contig sequences (Table 3.2). After screening by PCR against Ps26, IA4X, N37 and a small number of progeny from apomictic BC_8 segregating for mode of reproduction, 46 contigs showed specific amplification from Ps26 and apomictic BC_8 but no amplification from IA4X or sexual BC_8 individuals (Figure 3.4, Table 3.3). This amplification pattern establishes linkage of these 46 contigs to the ASGR-carrier chromosome. For the remaining 15 contigs which have not yet been linked to the ASGR-carrier chromosome, other methods of mapping, such as single-strand conformation polymorphism analysis (SSCP) or high resolution melting (HRM) will be applied. The 46 ASGR-carrier chromosome-linked contigs were then screened with a limited number of apomictic and sexual F_1s for mapping to the ASGR. This resulted in one ASGR-linked contig, Ps26_c9369, since the primers only amplified with apomictic F_1s but not sexual F_1s (Figure 3.5,

Table 3.3). Test for linkage of the remaining ASGR-carrier chromosome candidate genes to the ASGR, will need to be done with application of other, more sensitive approaches of mapping.

Expression Profiles of Alien Expressed Transcripts by RT-PCR

To date, RT-PCR with RNA extracted from apomictic BC_8 leaf, root, anther and ovary tissues has been completed for 36 candidate genes which were mapped to the ASGR-carrier chromosome. Thirty-four were expressed in all four organ types examined (Figure 3.6a, Table 3.4). However, one putative MADS-domain containing transcription factor, corresponding to contig Ps26_c33813, showed reproductive tissue specific amplification (Figure 3.6b, Table 3.4) and contig Ps26_c10535, a putative Lon protease, showed expression in all organs except anther (Table 3.4).

Phage cDNA Library Construction and Screening

Since only one out of 46 ASGR-carrier chromosome linked expressed genes could be localized to the ASGR, more complete gene sequence will be needed to increase the probability of finding sequence differences among gene family members/homeologs/alleles. In order to obtain longer sequences of the candidate genes to design primers for testing their linkage to the ASGR, a cDNA library containing about 300,000 phage clones was constructed from apomictic BC_8 mature ovary and anther RNA. This source of RNA was chosen because of the observed gene expression patterns and because it would be very difficult to collect enough immature ovule RNA for cDNA library construction due to small ovule size and tedious microdissection techniques. Screening of the cDNA library with probes amplified from 27 contigs resulted in detection of hybridization signals from 24 probes and recovery of primary plaques. PCR screening with contig-specific primers identified phage clones containing 17 of the desired transcript sequences (Figure 3.7, Table 3.5). Purified individual phage clones for each of the 17

candidate genes were recovered upon secondary screening and confirmed by PCR with gene-specific primers (Table 3.5).

Sequencing of the cDNA Clones

The insert sizes of cDNA clones were determined by PCR with M13 primers and most were larger than 1.0 kb (Figure 3.8). To date, good quality sequences have been obtained corresponding to 14 clones whereas the remaining 3 clones returned with bad quality sequences. After removal of vector, the sequences were assembled using ContigExpress in the VectorNTI Advanced 8 package. Blast analysis of the assemblies showed that full length cDNA of two genes showing high similarity to ubiquitin-conjugating enzyme E2 N (NP_001148361.1; GI: 226491078) and MADS-box transcription factor 8 (ACG32919.1; GI: 195622178) were obtained. The blast hits were consistent with the blast results of the corresponding Ps26/BC8 inter-genotype contigs, Ps26_c2785/ BC8_c8847 and Ps26_c33813/ BC8_c2708. Primers will be designed with the longer sequence information to test the linkage of the candidate genes to the ASGR in segregating F_1s.

Discussion

454 Transcriptome Sequencing for Identification of Alien Expressed Transcripts

Transcriptome profiling has been extensively used for gene discovery in plants because the absence of introns greatly enhances the information content of the data set and eases data interpretation (Busch and Lohmann, 2007; Malik et al., 2007; Spencer et al., 2007; Hoecker et al., 2008). Combined with 454 high-throughput sequencing technology in transcriptome sequencing, it has become a more promising approach to understand molecular events at the gene expression level on a genome-wide scale. Compared to the standard dideoxy chain-termination (Sanger et al.) sequencing method, 454-sequencing is faster, more cost-effective and

most importantly, it is capable of providing hundreds of thousands of sequences with ultra-high single-read accuracies of more than 99.5% (Droege and Hill, 2008). Successful applications of 454 sequencing technology in transcriptome sequencing and single nucleotide polymorphism (SNP) discovery have been reported. A total of 184,599 unique sequences generated from a normalized cDNA library of *Medicago truncatula* with a single 454 GS20 sequencing run had hits to more genes than a comparable amount of sequence from the TIGR *Medicago truncatula* Gene Index (MtGI). The data suggested the 454 DNA sequencing technology was effective in revealing transcripts (Cheung et al., 2006). Complementary DNAs derived from laser capture microdissection (LCM)-collected maize shoot apical meristems (SAM) were sequenced by 454 sequencing technology. Analysis of the LCM-SAM-454 sequences revealed that retrotransposon-related sequences accounted for about 14% of the total ESTs, consistent with the microarray results, which demonstrated the accuracy of 454 DNA sequencing technology (Ohtsu et al., 2007). When the transcriptomes of SAMs isolated from two inbred lines of maize were analyzed, more than 36,000 putative SNPs were detected and over 85% of a sample of the putative SNPs were validated by Sanger sequencing (Barbazuk et al., 2007). Application of 454 sequencing technology for expression profiling in *Drosophila melanogaster* also indicated its great potential in comparison of expression profiles across species (Torres et al., 2008) A study of soybean cyst nematode biotypes by 454 sequencing also revealed that the quality of the sequence data was sufficient for de novo SNP identification without a reference genome (Bekal et al., 2008).

In contrast to other work aimed at identifying genes involved in apomictic reproduction by searching for differences between apomictic and sexual genotypes, our study compared two apomictic lines for identical transcripts. The characteristically high accuracy of sequences

generated by 454 high-throughput technology allowed this goal to be accomplished. Comparison of the two ovule transcriptomes from the apospory initiation stage resulted in identification of 61 putative ASGR-carrier chromosome candidate expressed genes, of which 46 have been confirmed. Most of the 61 candidates contained less than 20 sequence reads in the assemblies. Since the number of sequence reads in the candidates provides some insight into the relative levels of transcription (Morozova and Marra, 2008), we conclude that most of the candidates are low-abundance or rare transcripts, further supporting the use of 454-sequencing technology for enhancing sequence depth. Combining the high accuracy and efficiency of detecting low-abundance or rare transcripts, 454 sequencing was shown to be an effective approach for identifying the identical transcripts between two transcriptomes.

However, the increase in throughput comes at the expense of read length. The current 454 GS FLX sequencer is capable of producing an average read length of 200-300 bases, which is still much shorter than the read length of capillary sequencing (650-800 bp) (Droege and Hill, 2008). Thus, fragment assembly of genomes sequenced by 454 could be challenging with the limitation of read-length although this concern is less for EST analysis since transcriptomes contain much less repetitive sequence. However, improved technology known as 'GS FLX Titanium' allows for reads longer than 400 bp is now available to researchers (www.454.com). Also application of paired-end reads strategy in parallel sequencing was shown to be able to identify and characterize tandem duplications, inverted duplications, as well as obtain high-resolution copy-number information (Campbell et al., 2008). These developments are improving the assembly of sequences containing repetitive elements. In the two ovule transcriptomes we generated by 454 sequencing, the average read length was ~150 bp, which was shorter than expected. Two potential reasons for shorter than expected reads are the use of RNA

amplification during preparation of the samples and subsequent shearing of the cDNAs, a standard part of the library preparation procedure. The T7-based RNA amplification yielded relatively 3'-enriched transcriptome, which actually was an advantage for identification of unique genes, but also resulted in shorter cDNA fragments compared to unamplified samples. Another possible factor is the species itself. It has been shown that the average read length can vary among different organisms due to differences in AT/GC content (Droege and Hill, 2008). Short sequence length was indeed a disadvantage during the mapping of candidates since the potential for designing gene- or allele- specific PCR primers was limited by the contig lengths. Mapping of the remaining 15 candidates to the alien chromosome and testing of linkage for the desired candidate genes to the ASGR would definitely require more sequence information.

Candidate Genes for Regulation of Apospory Initiation

We previously reported that the ASGR is sufficient to induce apomixis in sexual pearl millet (Ozias-Akins et al. 1998; Goel et al. 2003); therefore, the trait of apomixis in BC_8 is conferred by the ASGR-carrier chromosome from Ps26 (Singh et al. 2009). In the present study, we have attempted to identify candidate genes regulating the first step of apomixis, aposporous initial development, by transcriptome analysis of ovules from both Ps26 and BC_8. The ovules were collected at the stage of aposporous initial development, which ranged from no apparent apospory initials (~70%) to distinct aposporous initials observed (~30%). By pooling ovules over this range of development our objective was to minimize the chance of missing genes involved in the pathway of apomixis initiation since we would predict transcription prior to, and perhaps beyond, apospory initial formation. The 61 candidate genes identified from our study included genes with putative functions, genes with unknown functions and orphan genes. Based on the estimated size ratio of the ASGR to the whole chromosome (~25%) and the

heterochromatic character of the ASGR (Akiyama et al., 2004), it is unlikely that more than ¼ of these candidate genes are within the ASGR. Presently, only one gene could be mapped to the ASGR, and further mapping will require application of methods that enable polymorphism detection among related sequences in polyploids and the presence of a single-dose (or at most double-dose) allele on the ASGR-carrier chromosome. Near full-length cDNA sequences isolated from the cDNA library will facilitate the discovery of polymorphisms. Since the ovule-specific sequences at the stage of AI development may be absent from or under-represented in the library derived from a mix of mature ovary and anther, given that no hybridization signals were observed for three contigs during primary screening, we may not be able to isolate cDNA clones for some candidates from the cDNA library. For these genes, other approaches such as RACE (rapid amplification of cDNA ends) may be an alternative for deriving longer or full-length cDNA sequences.

Based on the hypothesis that apomixis arose through de-regulation of the sexual developmental pathway (Koltunow, 1993; Grimanelli et al., 2001), up- and down-regulated genes, sexual- and apomictic-specific genes, or heterochronically expressed genes between sexual and apomictic genotypes are all possible candidates for switching the mode of reproduction. By means of SuperSAGE, over 4,000 differentially expressed mRNA tags between sexual and apomictic ovules at the stage of apomeiosis were identified from *Boechera* (Sharbel et al., 2009), of which 543 tags were heterochronically expressed, 39 tags were down-regulated and 20 tags were up-regulated in apomictic ovules. In order to elucidate the function of any of the candidates, gene-by-gene functional analysis will have to be conducted. It is critical but also difficult to determine which genes are highest priorities for knock-out or knock-down for functional analysis. In contrast to the large number of candidates resulting from the

study of contrasting transcriptomes between apomictic and sexual genotypes, our strategy of transcriptome comparison between two apomictic lines gave rise to a much smaller group of candidates. The disadvantage of this strategy is that we may lose the down-regulated candidate genes in apomictic ovules since we were unable to include the sexual genotype in this study.

Without functional analysis of candidate apomixis genes identified from ovule transcriptomes, the genetic mechanism regulating apospory initiation remains unclear. However, the ovule transcriptomes provide significant new data for studying early ovule development and potential insight into reproductive pathways. Our data show that the combination of selecting specific reproductive tissues and sequencing with 454 high-throughput sequencing technology is a promising approach for identification of genes involved in different developmental events.

References

Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in Pennisetum squamulatum reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. Plant Physiol **134:** 1733-1741

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

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

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J **51:** 910-918

Bekal S, Craig JP, Hudson ME, Niblack TL, Domier LL, Lambert KN (2008) Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. Mol Genet Genomics **279:** 535-543

Bicknell RA, Koltunow AM (2004) Understanding apomixis: recent advances and remaining conundrums. Plant Cell **16 Suppl:** S228-245

Busch W, Lohmann JU (2007) Profiling a plant: expression analysis in Arabidopsis. Curr Opin Plant Biol **10:** 136-141

Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurles ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet **40:** 722-729

Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Linn Soc **61:** 51-94

Chen L, Guan L, Seo M, Hoffmann F, Adachi T (2005) Developmental expression of ASG- 1 during gametogenesis in apomictic guinea grass (Panicum maximum). J Plant Physiol **162:** 1141-1148

Chen L, Miyazaki C, Kojima A, Saito A, Adachi T (1999) Isolation and characterization of a gene expressed during early embryo sac development in apomictic guinea grass (*Panicum maximum*). J Plant Physiol **154:** 55-62

Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD (2006) Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. BMC Genomics **7:** 272

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res **14:** 1147-1159

Droege M, Hill B (2008) The Genome Sequencer FLX System-longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotechnol **136:** 3-10

Dujardin M, Hanna WW (1984) Cytogenetics of double cross hybrids between Pennisetum americanum -P. purpureum amphiploids and P. americanum X Pennisetum squamulatum interspecific hybrids. Theoretical and Applied Genetics **69:** 97-100

Dujardin M, Hanna WW (1989) Developing apomictic pearl millet-characterization of a BC_3 plant. J Genet Breed **43:** 145-151

Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res **17:** 69-73

Goel S, Chen Z, Conner JA, Akiyama Y, Hanna WW, Ozias-Akins P (2003) Delineation by fluorescence in situ hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in Pennisetum squamulatum and Cenchrus ciliaris. Genetics **163:** 1069-1082

Grimanelli D, Leblanc O, Perotti E, Grossniklaus U (2001) Developmental genetics of gametophytic apomixis. Trends Genet **17:** 597-604

Hanna WW (1995) Use of apomixis in cultivar development. Adv Agron **54:** 333–350

Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008) Comparison of maize (Zea mays L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics **179:** 1275-1283

Jones-Rhoades MW, Borevitz JO, Preuss D (2007) Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. PLoS Genet **3:** 1848-1861

Koltunow AM (1993) Apomixis: Embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell **5:** 1425-1437

Koltunow AM, Bicknell RA, Chaudhury AM (1995) Apomixis: Molecular strategies for the generation of genetically identical seeds without fertilization. Plant Physiol **108:** 1345-1352

Leblanc O, Armstead I, Pessino SC, Ortiz JP, Evans C, Valle CD, Hayward MD (1997) Non-radioactive mRNA fingerprinting to visualise gene expression in mature ovaries of Brachiaria hybrids derived from B. brizantha, an apomictic tropical forage. Plant Science **126:** 49-58

Malik MR, Wang F, Dirpaul JM, Zhou N, Polowick PL, Ferrie AM, Krochko JE (2007) Transcript profiling and identification of molecular markers for early microspore embryogenesis in Brassica napus. Plant Physiol **144:** 134-154

Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis BL, Matz MV (2009) Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. BMC Genomics **10:** 219

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics **92:** 255-264

Nakazono M, Qiu F, Borsuk LA, Schnable PS (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. Plant Cell **15:** 583-596

Nogler GA (1984) Gametophytic apomixis. *In* BM Johri, ed, Embryology of Angisoperms. Springer-Verlag, Berlin, pp 475–518

Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, Chen T, Zhang X, Timmermans MC, Beck J, Buckner B, Janick-Buckner D, Nettleton D, Scanlon MJ, Schnable PS (2007) Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). Plant J **52:** 391-404

Ozias-Akins P (2006) Apomixis: Developmental Characteristics and Genetics. Critical Reviews in Plant Sciences **25:** 199-214

Ozias-Akins P, Roche D, Hanna WW (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. Proc Natl Acad Sci U S A **95:** 5127-5132

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

Pessino SC, Espinoza F, Martinez EJ, Ortiz JP, Valle EM, Quarin CL (2001) Isolation of cDNA clones differentially expressed in flowers of apomictic and sexual Paspalum notatum. Hereditas **134:** 35-42

Rodrigues JC, Cabral GB, Dusi DM, de Mello LV, Rigden DJ, Carneiro VT (2003) Identification of differentially expressed cDNA sequences in ovaries of sexual and apomictic plants of Brachiaria brizantha. Plant Mol Biol **53:** 745-757

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:** 5463-5467

Savidan Y (2000) Apomixis: genetics and breeding. Plant Breed Rev **18:** 13–85

Sharbel TF, Voigt ML, Corral JM, Thiel T, Varshney A, Kumlehn J, Vogel H, Rotter B (2009) Molecular signatures of apomictic and sexual ovules in the Boechera holboellii complex. Plant J **58:** 870-882

Singh M, Burson BL, Finlayson SA (2007) Isolation of candidate genes for apomictic development in buffelgrass (Pennisetum ciliare). Plant Mol Biol **64:** 673-682

Singh M, Conner JA, Zeng Y, Hanna WW, Johnson VE, Ozias-Akins P Characterization of apomictic BC7 and BC8 pearl millet: Meiotic chromosome behavior and construction of an ASGR-carrier chromosome-specific library. Crop Science (in press)

Spencer MW, Casson SA, Lindsey K (2007) Transcriptional profiling of the Arabidopsis embryo. Plant Physiol **143:** 924-940

Torres TT, Metta M, Ottenwalder B, Schlotterer C (2008) Gene expression profiling by massively parallel sequencing. Genome Res **18:** 172-177

van Dijk P, van Damme J (2000) Apomixis technology and the paradox of sex. Trends Plant Sci **5:** 81-84

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol **17:** 1636-1647

Wang W, Wang Y, Zhang Q, Qi Y, Guo D (2009) Global characterization of Artemisia annua glandular trichome transcriptome using 454 pyrosequencing. BMC Genomics **10:** 465

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol **144:** 32-42

Young BA, Sherwood RT, Bashaw EC (1979) Cleared-pistil and thicksectioning techniques for detecting aposporous apomixis in grasses. Can J Bot **57:** 1668-1672

Table 3.1 Contigs sharing 100% identity between Ps26 and BC_8 ovule transcriptomes and the best hit of the assembled inter-genotype contigs to protein (BlastX) or nucleotide (BlastN) sequences in NCBI databases.

BLAST query	**Sequence # in Ps26 contigs**	**BLAST hit**	**Sequence # in BC_8 contigs**	**Overlap length (bp)**	**BlastX**	**BlastN (E Value)**	**BlastX of EST hit in BlastN column**
Ps26_ c8378	10	BC_8_ c5674	7	321	trafficking protein particle complex subunit 3 NP_001151598.1 GI:226490977		
Ps26_ c10331	6	BC_8_ c7991	8	241	no hit	RCRST0_005870 Foxtail millet EC612643.1 GI:149362118 (3e-55)	no hit
Ps26_ c11544	4	BC_8_ c10325	5	228	no hit	no hit	
Ps26_ c13157	4	BC_8_ c5112	13	227	no hit	no hit	
Ps26_ c13655	5	BC_8_ c24571	2	192	no hit	pPAP_06_E02 Apomictic pistil BM084376.1 GI:27532285 (8e-24)	putative 26S proteasome non-ATPase regulatory subunit 3 ACG34075.1 GI:195624490
Ps26_ c1372	28	BC_8_ c12789	9	326	no hit	CCGC4364.g1 CCGC *Panicum virgatum* early floral buds + reproductive tissue FL750787.1 GI:198007657 (e-174)	NADH-ubiquinone oxidoreductase 51 kDa subunit NP_001148767.1 GI:226532265
Ps26_ c13922	5	BC_8_ c12833	3	212	no hit	no hit	
Ps26_ c2448	22	BC_8_ c12858	5	225	no hit	pPAP_10_F04 Apomictic pistil FL813942.1 GI:198086024 (2e-57)	ankyrin protein kinase-like NP_001152470.1 GI:226495939

Ps26_ c30691	2	BC_8_ c10294	5	206	no hit	no hit	
Ps26_ c3546	14	BC_8_ c8622	6	295	no hit	no hit	
Ps26_ c5080	7	BC_8_ c12542	3	212	hypothetical protein OsJ_24918 EEE67490.1 GI:222637358		
Ps26_ c583	29	BC_8_ c6141	9	223	no hit	6X_JF-rd_A11 pAPO *Cenchrus ciliaris* EB652936.1 GI:164107582 (6e-127)	SRC2 protein kinase C - phospholipids ACG40316.1 GI:195641696]
Ps26_ c8165	7	BC_8_ c5964	9	185	no hit	84Z_JF_G03 pAPO *Cenchrus ciliaris* EB661430.1 GI:164123871 (7e-70)	TPA: AT-hook motif nuclear localized protein 2 FAA00302.1 GI:119657406
Ps26_ c9369	7	BC_8_ c3452	14	190	hypothetical protein OsJ_30933 EAZ15525.1 GI:125574241		
Ps26_ c2339	15	BC_8_ c7917	6	264	no hit	CCGG12847.g1 CCGG *Panicum virgatum* late flowering buds FL812358.1 GI:198084376 (e-23)	ESP4 (ENHANCED SILENCING PHENOTYPE 4) NP_195760.1 GI:15240970
Ps26_ c1279	20	BC_8_ c8634	6	243	ENT domain containing protein ACG36577.1 GI:195629872		
Ps26_ c7587	8	BC_8_ c11918	3	202	ATPNG1 (*Arabidopsis Thaliana* PEPTIDE-N-GLYCANASE 1) NP_199768.1 GI:15240508		

Ps26_c2785	12	BC_8_c8847	6	273	ubiquitin-conjugating enzyme E2 N NP_001148361.1 GI:226491078		
Ps26_c194	44	BC_8_c2920	18	304	no hit	no hit	
Ps26_c17388	2	BC_8_c6454	10	208	histone 4 BAG68513.1 GI:195972757		
Ps26_c3455	11	BC_8_c8607	5	193	no hit	26X_JF_C01 pAPO *Cenchrus ciliaris* EB655151.1 GI:164198597 (e-102)	putative condensing XP_002529162.1 GI:255576542
Ps26_c1312	26	BC_8_c3757	15	313	no hit	25X_JF_D10 pAPO *Cenchrus ciliaris* EB656417.1 GI:164027660 (2e-47)	protein phosphatase 2A regulatory subunit A AAM94368.1 GI:22296816
Ps26_c338	46	BC_8_c3527	17	419	universal stress protein (USP) family protein NP_001159067.1 GI:259490110		
Ps26_c33813	2	BC_8_c2708	20	229	putative MADS-domain transcription factor CAA70485.1 GI:3851333		
Ps26_c1422	25	BC_8_c3852	14	245	no hit	no hit	
Ps26_c6131	9	BC_8_c8955	6	224	no hit	CCHY9952.g1 CCHY *Panicum virgatum* callus FL987585.1 GI:198319427 (e-49)	putative calcium-dependent protein kinase ACG46220.1 GI:195653505
Ps26_c2388	19	BC_8_c2949	19	201	no hit	6W6III_JF_H03 pAPO *Cenchrus ciliaris* EB662068.1 GI:164227478 (3e-48)	polygalacturonase inhibitor 1 precursor ACG36448.1 GI:195629614

Ps26_ c32589	5	BC_8_ c3672	17	229	no hit	71Z_JF_B09 pAPO *Cenchrus ciliaris* EB654350.1 GI:163993222 (5e-93)	putative microtubule-associated protein CAD23144.1 GI:37776903
Ps26_ c10535	6	BC_8_ c22186	2	182	no hit	no hit	
Ps26_ c2807	10	BC_8_ c12602	5	241	no hit	5X_JF_A06 pAPO *Cenchrus ciliaris* EB652848.1 GI:164180053 (6e-116)	Phosphoglucomutase/phosphomannomutase C terminal ABN08987.1 GI:124361015
Ps26_ c2838	14	BC_8_ c3538	15	183	no hit	no hit	
Ps26_ c3609	10	BC_8_ c10814	4	245	no hit	no hit	
Ps26_ c5210	14	BC_8_ c5192	10	273	hypothetical protein OsJ_25077 EEE67565.1 GI:222637433		
Ps26_ c6744	6	BC_8_ c292	93	257	ADP-ribosylation factor BAB90396.1 GI:20161472		
Ps26_ c9776	6	BC_8_ c4965	10	258	no hit	MK_7_78 *Pennisetum glaucum* seedlings CD726437.1 GI:32277284 (2e-46)	hypothetical protein SORBIDRAFT_07g010440 XP_002444160.1 GI:242078783
Ps26_ c5851	13	BC_8_ c5854	9	192	no hit	no hit	
Ps26_ c6373	8	BC_8_ c6664	9	235	no hit	1475276 CERES-197 *Zea mays* FL451677.1 GI:211043870 (2e-41)	hypothetical protein LOC100276553 NP_001143786.1 GI:226505008
Ps26_ c20942	3	BC_8_ c5841	6	209	putative splicing factor Prp8 BAD67606.1 GI:55296044		

Ps26_ c24301	12	BC_8_ c1706	26	191	no hit	CCHZ29515.g1 CCHZ *Panicum virgatum* GD049680.1 GI:198385177 (4e-92)	putative GTP-binding protein CAC39050.1 GI:14140133
Ps26_ c25664	2	BC_8_ c10631	3	202	ACT domain containing protein NP_001151043.1 GI:226530840		
Ps26_ c30198	2	BC_8_ c9466	5	220	centromere/microtubule binding protein cbf5, putative XP_002523427.1 GI:255564866		
Ps26_ c3993	18	BC_8_ c16185	3	246	fk506-binding protein, putative XP_002534360.1 GI:255587693		
Ps26_ c4364	13	BC_8_ c15332	2	181	no hit	CCHZ9541.g1 CCHZ *Panicum virgatum* GD021384.1 GI:198352214 (5e-31)	helix-loop-helix-like protein AAO72577.1 GI:29367409
Ps26_ c5781	12	BC_8_ c3084	15	214	no hit	83X_JF_E11 pAPO *Cenchrus ciliaris* EB658721.1 GI:164099005 (0)	hypothetical protein SORBIDRAFT_04g004 190 XP_002451580.1 GI:242060582
Ps26_ c6192	6	BC_8_ c6077	8	188	no hit	SS1_23_E11.b1_A012 Salt-stressed seedlings Sorghum CD231677.1 GI:30975142 (2e-81)	putative serine/threonine protein phosphatase 2A BAD67828.1 GI:55296109
Ps26_ c1472	16	$BC_8$8_ c3819	15	330	small zinc finger-like protein AAD40002.1 GI:5107180		
Ps26_ c2405	17	BC_8_ c12726	3	208	ethylene-responsive small GTP-binding protein AS88430.1		

					GI:46326983		
Ps26_ c15085	3	BC_8_ c7592	7	227	DNA cytosine methyltransferase Zmet3 ACG47300.1 GI:195655665		
Ps26_ c1406	24	BC_8_ c4551	9	221	putative anther ethylene-upregulated protein ER1 BAC79907.1 GI:33146619		
Ps26_ c1580	24	BC_8_ c6425	8	216	putative vacuolar ATP synthase AAO65147.1 GI:29164794		
Ps26_ c18163	3	BC_8_ c6626	8	290	peroxisomal membrane carrier protein NP_001152063.1 GI:226500946		
Ps26_ c1878	13	BC_8_ c7425	9	242	no hit	CCGI4193.g1 CCGI *Panicum virgatum* FL856163.1 GI:198128193(3e-69)	hypothetical protein SORBIDRAFT_02g036 200 XP_002460850.1 GI:242045958
Ps26_ c19109	3	BC_8_ c9186	4	205	no hit	no hit	
Ps26_ c22381	2	BC_8_ c547	226	185	no hit	2X6III_JF-rd_A11 pAPO *Cenchrus ciliaris* EB652659.1 GI:164076750 (7e-58)	APx2 - Cytosolic Ascorbate Peroxidase ACG41151.1 GI:195643366
Ps26_ c28392	2	BC_8_ c12100	4	230	no hit	no hits	
Ps26_ c3656	11	BC_8_ c2239	27	199	no hit	74X_JF_A05 pAPO *Cenchrus ciliaris* EB653126.1 GI:164132997 (2e-171)	Zea CEFD homolog1 NP_001151142.1 GI:226501670

Ps26_c4150	14	BC_8_c3261	11	276	rRNA-processing protein EBP2, putative XP_002526440.1 GI:255570978		
Ps26_c704	40	BC_8_c1322	27	368	26S protease regulatory subunit, Putative XP_002526219.1 GI:255570523		
Ps26_c2552	57	BC_8_c1808	56	384	40S ribosomal protein S6 ACG31980.1 GI:195620300		
Ps26_c14318	4	BC_8_c14583	4	366	triose phosphate/phosphate translocator ACG33816.1 GI:195623972		
Ps26_c21597	2	BC_8_c14882	2	226	RNA polymerase III large subunit AAS79689.1 GI:45935335		

Table 3.2 Primers designed for mapping.

Oligo Name	Primer #	Sequence (5'-3')	Tm (°C)
Ps26_c10331_F	1476	ACTAGGGGAAAGAAAGTCATCCTC	60
Ps26_c10331_R	1477	AAGCGTGTATCACTTCACGAAC	
Ps26_c11544_F	1478	TACTTTTGGTGTGGTTTCGACC	56
Ps26_c11544_R	1479	CATACATGGACTACTGATGGCT	
Ps26_c13157_F	1480	GGCTAGAACCGTCGAGCAT	60
Ps26_c13157_R	1481	TTCAGAGAATTAGGCCGAGGAG	
Ps26_c13655_F	1482	GCTGCTGCTTCCTAGTTTCATT	60
Ps26_c13655_R	1483	CTAATTTTCAGGGCCAGGACAG	
Ps26_c1372_F	1484	GCCATTGCGCTTCTGTAATAGT	60
Ps26_c1372_R	1485	AGCATACATGGTACATTGATCGAG	
Ps26_c13922_F	1486	AGATCATTAGCCACTAGGATCGTC	60
Ps26_c13922_R	1487	AGAACAAGAACCTCCTGCTACAA	
Ps26_c20942_F	1488	GAAGCATCTGTTCCGTTCACTC	60
Ps26_c20942_R	1489	TCCTGTGAATCAACAAGTTCAGC	
Ps26_c24301_F	1490	CGTCCTGCTGCTCTTCTT	58
Ps26_c24301_R	1491	TGCTATTTCGGCAATTAACAGA	
Ps26_c2448_F	1492	TATCGTGCCTGCTGGTGATAG	60
Ps26_c2448_R	1493	CTGTACTCCTTGATACGCAAGC	
Ps26_c25664_F	1494	CAAAGTTGAGGACCCTGACAT	58
Ps26_c25664_R	1495	CATGCTTCGTTTTGGTCCATC	
Ps26_c30198_F	1496	CATCACTACCCTCTTGATCTTTGC	60
Ps26_c30198_R	1497	AAGGACTCTGCTGTGAATGCTA	
Ps26_c30691_F	1498	AAATGGACTTAGGCTTGCGTTG	60
Ps26_c30691_R	1499	ACAATACATCAATACCCAGTTGCC	
Ps26_c3546_F	1500	CCGCGACAACATATCAACACAT	60
Ps26_c3546_R	1501	CGGATGAGAAAACAACAGCCAT	
Ps26_c3993_F	1502	CCGTGGAAAACTTACTGATGGG	60
Ps26_c3993_R	1503	AGCAACAAGTTCCGTGTCAAAT	
Ps26_c4364_F	1504	CAGATCACCCAGATCACCAC	60
Ps26_c4364_R	1505	AAGCTAAATCCTGGTGGAAAGA	
Ps26_c5080_F	1506	TTAGGGACTGATGACACAAGGG	61
Ps26_c5080_R	1507	GGCATCAACCAGGACAATACTC	
s26_c5781_F	1508	TGTTGTAGTTCGGTTTCTGCTC	60
s26_c5781_R	1509	TCCCATTTCAACCTTTATGCCAG	
Ps26_c583_F	1510	CTACGACGACGACTACTGATCC	62
Ps26_c583_R	1511	ATTTTGCAGTGCTCAACTCCTC	
Ps26_c8165_F	1512	TTCTTCTTCTTCACCACCAAGG	60
Ps26_c8165_R	1513	GCTCTTGCTGTTAGGTGTTGTT	
Ps26_c9369_F	1514	GGCATACCAGCAAACCAAGTTA	61
Ps26_c9369_R	1515	ATCATCACATGCTTCAAACAGC	

Ps26_c2339_F	1528	TAAAAAAGGGGAGGATGA	55
Ps26_c2339_R	1529	AGAAAAGCAGGATTTATAGTGT	
Ps26_c1279_F	1530	TTTAAGTACCTTCTGTTTGGAG	55
Ps26_c1279_R	1531	TGGCTTTTACCTTGGTCT	
Ps26_c7587_F	1532	TGATGAATTATTGGTGGTCA	55
Ps26_c7587_R	1533	GTTGTTTATTCATGTCCTTAATT	
Ps26_c2785_F	1534	ATGAGATATATATGTAACC	50
Ps26_c2785_R	1535	GTAATAATCTTCGTGTTG	
Ps26_c17388_F	1538	TCATCATCTTCGTCCTGG	55
Ps26_c17388_R	1539	AGAAGATCGACACACACAC	
Ps26_c3455_F	1540	GGATGTGCTGTTGCCCTG	55
Ps26_c3455_R	1541	CAAGCAGCAGTTAACAGACTT	
Ps26_c1312_F1	1542	CTGTCGCTGTTCAACTAC	55
Ps26_c1312_R1	1543	GAGCTAGACAACACATCA	
Ps26_c6192_F	1546	GCGCCTTGTGCCAAAGTC	60
Ps26_c6192_R	1547	TTCTCACTCCCTCCAGCC	
Ps26_c338_F1	1548	GCTTTGGCATTGTTTGTG	55
Ps26_c338_R1	1549	AGGTATTGCAATGCAAAGA	
Ps26_c33813_F	1565	AATAAACTAGTACTGCAATCTGC	55
Ps26_c33813_R	1566	GTACGGTGTTCAAATAATGTG	
Ps26_c1422_F	1567	GTTAGATTGCTGTTTTGATG	55
Ps26_c1422_R	1568	ATGAGCGCAATTTGAACA	
Ps26_c6131_F2	1571	CTCGGCAAAGCAACAAGCGG	61
Ps26_c6131_R2	1572	CGTATTGGGAAACAACACGC	
Ps26_c1472_F	1573	TCCATAAAACTAAAGGCTATGT	54
Ps26_c1472_R	1574	TTTGATGGATTACTATGTAAACC	
Ps26_c2388_F	1575	TCGAGCTGTTAAAGGCGTGTCA	61
Ps26_c2388_R	1576	CCTTTGATAGTACATCAGGCCG	
Ps26_c2405_F	1577	CATCTGTCGTCTCCGTGTC	61
Ps26_c2405_R	1578	CTCTTGCTGATGAGTATGGG	
Ps26_c15085_F	1579	CATTATTTTAGAATTCTTGACTCC	54
Ps26_c15085_R	1580	CAAATATCATTCACAGTTCAACG	
Ps26_c2009_F	1581	AACGCCAATTGATCACACGAGC	61
Ps26_c2009_R	1582	TCAGGCTGCAGCTGCTGATG	
Ps26_c1406_F	1583	CCGATGGGGGTGATTTAATG	61
Ps26_c1406_R	1584	CGGTGACCATCTAGTAACAGGA	
Ps26_c194_F2	1604	AAACATCATATAGATCCGG	53
Ps26_c194_R2	1605	ATTTTCCCTAATTCTCAGAG	
Ps26_c1580_F	1628	TGGAGTCATTGTGGGCATCA	60
Ps26_c1580_R	1629	CATGGAGTCTCGGCAGTAGG	
Ps26_c10535_F	1630	CACTTGTTGGTTCTTTCTGTTGG	57
Ps26_c10535_R	1631	TTAAGTGGTTCGTCGGCATA	
Ps26_c18163_F	1632	ACAGAGGCAAGCTCCATAGC	59

Ps26_c18163_R	1633	AGGGTGCTCGACTATTGGTT	
Ps26_c1878_F	1634	ACCATTCCGTCATCTCTCGG	59
Ps26_c1878_R	1635	ATTATCGGCGTGTAAGGCCA	
Ps26_c19109_F	1636	GGCTTCAGTTTAGCGTTTGCT	60
Ps26_c19109_R	1637	AAGAATCATGCCAACACGCC	
BC8_c547_F	1638	CACAAAGCTAAGAAGCGGTGAA	60
BC8_c547_R	1639	CACAAGGCCCACCAATGTTTAT	
Ps26_c2807_F	1640	ATTACCGGCGTGTCACTTATCT	60
Ps26_c2807_R	1641	CGCATGGAAGCAGAAGGTTTAT	
Ps26_c2838_F	1642	TTACCGGACATTGTGAGGAACG	58
Ps26_c2838_R	1643	CCAGAGTATTTGGTTCCCATCT	
Ps26_c28392_F	1644	CTGTAGTACGGCACGAGTGG	60
Ps26_c28392_R	1645	CAAGTTCTCCGACTGCTTGAAA	
Ps26_c3609_F	1646	GCAGACAGTTGGATGACATTGG	58
Ps26_c3609_R	1647	ACTAAAGGAGAGGCGCTGTAAA	
Ps26_c3656_F	1648	TCGGAAACCAGCAGTCAATCTA	59
Ps26_c3656_R	1649	CATGTGTTGTATTTGGCGTTCC	
Ps26_c4150_F	1650	GAGTCAGTCAAGAAGTGGAGGA	60
Ps26_c4150_R	1651	TTTAGCCCTTTACGACCACCAT	
Ps26_c5210_F	1652	GGGTTTCAGGTTTGATTTCGGT	60
Ps26_c5210_R	1653	GCAAGACCGTCATACTTCATTCC	
Ps26_c5851_F	1654	CAGAACAGCCACTTTGCTCTAA	59
Ps26_c5851_R	1655	GCTCCATTGGGACATACACATC	
Ps26_c6373_F	1656	CTGACACCATAAACAGCAACCC	60
Ps26_c6373_R	1657	CTTCTACAACGACAAGAACCCG	
Ps26_c6744_F	1658	TGCAACCTGTTAGGTGTCTTCT	60
Ps26_c6744_R	1659	CCAGACGGTGTACTAATCACGA	
Ps26_c704_F	1660	CGTGATGACAAAGGACGAGTTC	58
Ps26_c704_R	1661	GGACGCCTTCCTTCTTCTTGTA	
Ps26_c8378_F	1662	ACAGCAAATGAGCATCCAAACA	60
Ps26_c8378_R	1663	TAGGCATTGGAGATGGTGTCAA	
Ps26_c9776_F	1664	TGTGATATGTGTGCCTAGCTGT	60
Ps26_c9776_R	1665	AATACGGCAAATGATGCCAAGT	
Ps26_c14318_F	1666	TTGGAGTGCAATGGCTTCAAAT	60
Ps26_c14318_R	1667	ACTGACTTTGACACCCTCTGTT	
Ps26_c21597_F	1668	TTGTGCTCATACGTTGGTGACG	60
Ps26_c21597_R	1669	AACGGCAAGATTCAGAAACACC	
Ps26_c2552_F	1670	GAAGAGGATTGCCAAGAAGCAG	58
Ps26_c2552_R	1671	TCGACTGATCCAACACAGAACA	
Ps26_c1406_f1	1680	TCAGTTAGGCAGTGACCTGT	59
Ps26_c1406_r1	1681	TGAAAGTCGAACTATAGAAGTCCCA	
Ps26_c1472_f1	1682	AGTCGGACAGCTTCTTGGAA	60
Ps26_c1472_r1	1683	ATTTCTCTCCCGAACACCCC	

Ps26_c150585_f1	1684	CAGTCCATGCAACAACCTGG	60
Ps26_c150585_r1	1685	GCTGTCGGAGCAAACTCAAA	
Ps26_c1580_f1	1686	TTCCCCATTCATCTCGCTGT	60
Ps26_c1580_r1	1687	AACCAGCACACATGGAGTCT	
Ps26_c18163_f1	1688	TGATCTAGGTCTTTGCAAGTATTAGT	55
Ps26_c18163_r1	1689	CCTAACAGAGGCAAGCTCCA	
Ps26_c1878_f1	1690	TAGCCAGCAAAAGGTAATCTCC	55
Ps26_c1878_r1	1691	TTAGCGAGATTATCGGCGTG	
Ps26_c19109_f1	1692	TGCTGCTTTTTGTCAGAGTTCG	55
Ps26_c19109_r1	1693	GTGCTTTCACCATTATCGTCCA	
Ps26_c20942_f1	1694	GGCTCAGAAAAAGAAGCATCTG	55
Ps26_c20942_r1	1695	CACGGGTTTCAAATTGAAATTG	
Ps26_c22381_f1	1696	GTGAGGCAGGCCATATTTTGT	55
Ps26_c22381_r1	1697	AAATCTTCGTCGAGACGTCGA	
Ps26_c2405_f1	1698	AAGACCAGTGATTCCACCAA	55
Ps26_c2405_r1	1699	CAGCAAGATTTGGTGTTACTGC	
Ps26_c24301_f1	1700	TCAGTTAATGTAAACTTTGTTGCACTG	55
Ps26_c24301_r1	1701	TGAAAAATGCTATTTCGGCA	
Ps26_c25664_f1	1702	AGAGAGGATACGGCTAACCA	55
Ps26_c25664_r1	1703	TCATCTTCAACAACCACATGA	
Ps26_c28392_f1	1704	TCGTCCTGCTCCTGCCTCAA	55
Ps26_c28392_r1	1705	CGGATCTCATCATGCCCATA	
Ps26_c30198_f1	1706	ACCCTCTTGATCTTTGCAACAG	55
Ps26_c30198_r1	1707	TGGTGCTAAGCTTATGATCCCT	
Ps26_c704_f1	1708	TCCCCGATCAGGTAGCAAAC	60
Ps26_c704_r1	1709	GCCGCTCTTGTTTGTCAGG	
Ps26_c3656_f1	1710	CTGCCACTTTGATGTATTGACG	55
Ps26_c3656_r1	1711	CAATAATGCCCTGGCATGTC	
Ps26_c3993_f1	1712	GCGTAAACCATTTACTCAGAGC	59
Ps26_c3993_r1	1713	GATGAGTTTGCTCCAGGACG	
Ps26_c4150_f1	1714	TCAAGAACTTGTGCTCAGATCC	59
Ps26_c4150_r1	1715	GCTTCTGGTAACAACACACCA	
Ps26_c4364_f1	1716	ACAACGGTGTCACCAGATCA	60
Ps26_c4364_r1	1717	CTCAGTGCTGTCGGTGTAGT	
Ps26_c5781_f1	1718	AGGTCTTTGTCAGGTTTTAGATTGA	60
Ps26_c5781_r1	1719	TGGCAGATCAGAATACCTCCC	
Ps26_c6192_f1	1720	TTTGTTGCCTTTGTAAGCGC	55
Ps26_c6192_r1	1721	ACTCCCTCCAGCCAAGTAGAAC	

Table 3.3 Summary of mapping results. The one contig mapped to the ASGR is shown in bold. +: positive amplification; -: no amplification; N/A: not assayed.

Ps26_contig	Primer Numbers	cDNA size	PCR with Ps26	PCR with IA4X	PCR with N37	PCR with BC_8 apo : sex	PCR with F_1 apo : sex
Ps26_c10331	1476/1477	210	+	-	-	+ : -	+ : +
Ps26_c11544	1478/1479	165	+	-	+	+ : -	+ : +
Ps26_c13157	1480/1481	161	+	-	+	+ : -	+ : +
Ps26_c13655	1482/1483	214	+	-	+	+ : -	+ : +
Ps26_c1372	1484/1485	215	+	-	+	+ : -	+ : +
Ps26_c13922	1486/1487	200	+	-	-	+ : -	+ : +
Ps26_c2448	1492/1493	189	+	-	+	+ : -	+ : +
Ps26_c30691	1498/1499	206	+	-	+	+ : -	+ : +
Ps26_c3546	1500/1501	245	+	-	+	+ : -	+ : +
Ps26_c5080	1506/1507	204	+	-	-	+ : -	+ : +
Ps26_c583	1510/1511	212	+	-	+	+ : -	+ : +
Ps26_c8165	1512/1513	150	+	-	+	+ : -	+ : +
Ps26_c9369	**1514/1515**	**274**	**+**	**-**	**-**	**+ : -**	**+ : -**
Ps26_c2339	1528/1529	213	+	-	-	+ : -	+ : +
Ps26_c1279	1530/1531	228	+	-	+	+ : -	+ : +
Ps26_c7587	1532/1533	172	+	-	+	+ : -	+ : +
Ps26_c2785	1534/1535	226	+	-	-	+ : -	+ : +
Ps26_c194	1604/1605	283	+	-	-	+ : -	+ : +
Ps26_c17388	1538/1539	163	+	-	+	+ : -	+ : +
Ps26_c3455	1540/1541	102	+	-	+	+ : -	+ : +
Ps26_c1312	1542/1543	143	+	-	+	+ : -	+ : +
Ps26_c338	1548/1549	120	+	-	+	+ : -	+ : +
Ps26_c33813	1565/1566	140	+	-	+	+ : -	+ : +
Ps26_c1422	1567/1568	120	+	-	+	+ : -	+ : +
Ps26_c6131	1571/1572	179	+	-	+	+ : -	+ : +
Ps26_c2388	1575/1576	128	+	-	+	+ : -	+ : +
Ps26_c32589	1581/1582	216	+	-	+	+ : -	+ : +
Ps26_c10535	1630/1631	148	+	-	N/A	+ : -	+ : +
Ps26_c2807	1640/1641	164	+	-	N/A	+ : -	+ : +
Ps26_c2838	1642/1643	103	+	-	N/A	+ : -	+ : +
Ps26_c3609	1646/1647	150	+	-	N/A	+ : -	+ : +
Ps26_c5210	1652/1653	157	+	-	N/A	+ : -	+ : +
Ps26_c6744	1658/1659	202	+	-	N/A	+ : -	+ : +
Ps26_c9776	1664/1665	170	+	-	N/A	+ : -	+ : +
Ps26_c5851	1654/1655	179	+	-	N/A	+ : -	+ : +
Ps26_c6373	1656/1657	178	+	-	N/A	+ : -	+ : +

Ps26_c20942	1488/1489	125	+	+	+	+ : +	+ : +
Ps26_c24301	1490/1491	120	+	+	+	+ : +	+ : +
Ps26_c25664	1494/1495	193	+	+	+	+ : +	+ : +
Ps26_c30198	1496/1497	210	+	+	+	+ : +	+ : +
Ps26_c3993	1502/1713	800	+	N/A	N/A	+ : -	+ : +
Ps26_c4364	1505/1716	150	+	N/A	N/A	+ : -	+ : +
Ps26_c5781	1508/1509	156	+	+	+	+ : +	+ : +
Ps26_c6192	1720/1721	140	+	N/A	N/A	+ : -	+ : +
Ps26_c1472	1573/1574	185	+	+	+	+ : +	+ : +
Ps26_c2405	1577/1578	180	+	+	+	+ : +	+ : +
Ps26_c15085	1579/1580	120	+	+	+	+ : +	+ : +
Ps26_c1406	1583/1681	250	+	N/A	N/A	+ : -	+ : +
Ps26_c1580	1628/1629	237	+	+	N/A	+ : +	+ : +
Ps26_c18163	1632/1633	169	+	+	N/A	+ : +	+ : +
Ps26_c1878	1690/1691	157	+	N/A	N/A	+ : -	+ : +
Ps26_c19109	1692/1693	163	+	N/A	N/A	+ : -	+ : +
Ps26_c22381	1696/1697	246	+	N/A	N/A	+ : -	+ : +
Ps26_c28392	1704/1705	181	+	N/A	N/A	+ : -	+ : +
Ps26_c3656	1648/1649	152	+	+	N/A	+ : +	+ : +
Ps26_c4150	1650/1715	450	+	N/A	N/A	+ : -	+ : +
Ps26_c704	1708/1709	155	+	N/A	N/A	+ : -	+ : +
Ps26_c2552	1670/1671	243	+	+	N/A	+ : +	+ : +
Ps26_c14318	1666/1667	175	+	+	N/A	+ : +	+ : +
Ps26_c21597	1668/1669	150	+	+	N/A	+ : +	+ : +
Ps26_c8378	1662/1663	199	N/A	N/A	N/A	N/A	N/A

Table 3.4 RT-PCR results for the 36 contigs mapped to the ASGR-carrier chromosome. Leaf, root, ovary and anther tissues were collected from apomictic BC_8 plants. RT(+): RT with reverse transcriptase; RT(-): RT without reverse transcriptase as DNA contamination control; +: positive amplification; -: no amplification. The contig showing amplification only in ovary and anther is indicated in bold and the contig showing no expression in anther is indicated in italics.

Contig IDs	Leaf RT(+)	Leaf RT(-)	Root RT(+)	Root RT(-)	Ovary RT(+)	Ovary RT(-)	Anther RT(+)	Anther RT(-)
Ps26_c10331	+	-	+	-	+	-	+	-
Ps26_c11544	+	-	+	-	+	-	+	-
Ps26_c13157	+	-	+	-	+	-	+	-
Ps26_c13655	+	-	+	-	+	-	+	-
Ps26_c1372	+	-	+	-	+	-	+	-
Ps26_c13922	+	-	+	-	+	-	+	-
Ps26_c2448	+	-	+	-	+	-	+	-
Ps26_c30691	+	-	+	-	+	-	+	-
Ps26_c3546	+	-	+	-	+	-	+	-
Ps26_c5080	+	-	+	-	+	-	+	-
Ps26_c583	+	-	+	-	+	-	+	-
Ps26_c8165	+	-	+	-	+	-	+	-
Ps26_c9369	+	-	+	-	+	-	+	-
Ps26_c2339	+	-	+	-	+	-	+	-
Ps26_c1279	+	-	+	-	+	-	+	-
Ps26_c7587	+	-	+	-	+	-	+	-
Ps26_c2785	+	-	+	-	+	-	+	-
Ps26_c194	+	-	+	-	+	-	+	-
Ps26_c17388	+	-	+	-	+	-	+	-
Ps26_c3455	+	-	+	-	+	-	+	-
Ps26_c1312	+	-	+	-	+	-	+	-
Ps26_c338	+	-	+	-	+	-	+	-
Ps26_c33813	**-**	**-**	**-**	**-**	**+**	**-**	**+**	**-**
Ps26_c1422	+	-	+	-	+	-	+	-
Ps26_c6131	+	-	+	-	+	-	+	-
Ps26_c2388	+	-	+	-	+	-	+	-
Ps26_c32589	+	-	+	-	+	-	+	-
Ps26_c10535	+	-	+	-	+	-	-	-
Ps26_c2807	+	-	+	-	+	-	+	-

Ps26_c2838	+	-	+	-	+	-	+	-
Ps26_c3609	+	-	+	-	+	-	+	-
Ps26_c5210	+	-	+	-	+	-	+	-
Ps26_c6744	+	-	+	-	+	-	+	-
Ps26_c9776	+	-	+	-	+	-	+	-
Ps26_c5851	+	-	+	-	+	-	+	-
Ps26_c6373	+	-	+	-	+	-	+	-

Table 3.5 cDNA library screening results. N/A: not available.

Contig ID	Primer #	# of primary signals	# of positive primary signals after PCR screen	# of single clones sequenced
Ps26_c10331	1476/77	1	1	1
Ps26_c11544	1478/79	0	N/A	N/A
Ps26_c13157	1480/81	1	1	1
Ps26_c13655	1482/83	2	0	N/A
Ps26_c1372	1484/85	2	1	1
Ps26_c13922	1486/87	2	0	N/A
Ps26_c2448	1492/93	2	3	3
Ps26_c30691	1498/99	2	2	2
Ps26_c3546	1500/01	1	0	N/A
Ps26_c5080	1506/07	2	2	2
Ps26_c583	1510/11	2	1	1
Ps26_c8165	1512/13	2	2	1
Ps26_c9369	1514/15	2	2	1
Ps26_c2339	1528/29	0	N/A	N/A
Ps26_c1279	1530/31	1	0	N/A
Ps26_c7587	1532/33	3	0	N/A
Ps26_c2785	1534/35	4	1	1
Ps26_c17388	1538/39	1	1	1
Ps26_c3455	1540/41	0	N/A	N/A
Ps26_c1312	1542/43	1	1	1
Ps26_c338	1548/49	1	1	1
Ps26_c33813	1565/66	4	4	2
Ps26_c1422	1567/68	1	0	N/A
Ps26_c6131	1571/72	1	1	1
Ps26_c2388	1575/76	3	3	3
Ps26_c32589	1581/82	5	1	1
Ps26_c194	1604/05	1	0	N/A

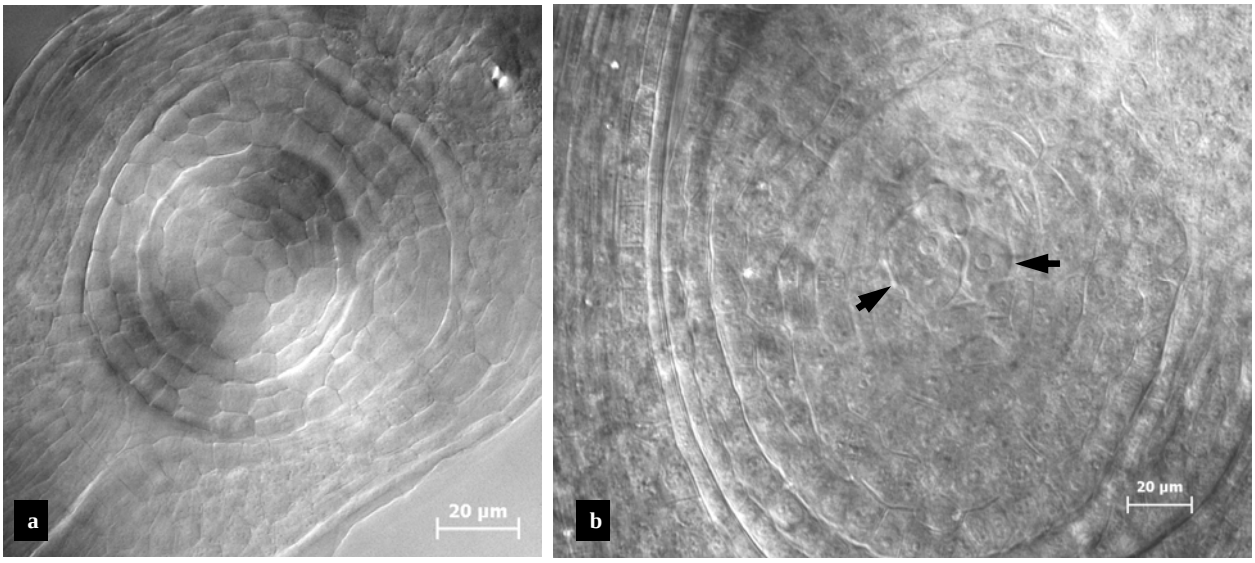


Figure 3.1 Ovary clearing results. a, cleared ovary showing no aposporous initials. b, cleared ovary showing two aposporous initials, indicated by solid arrows.

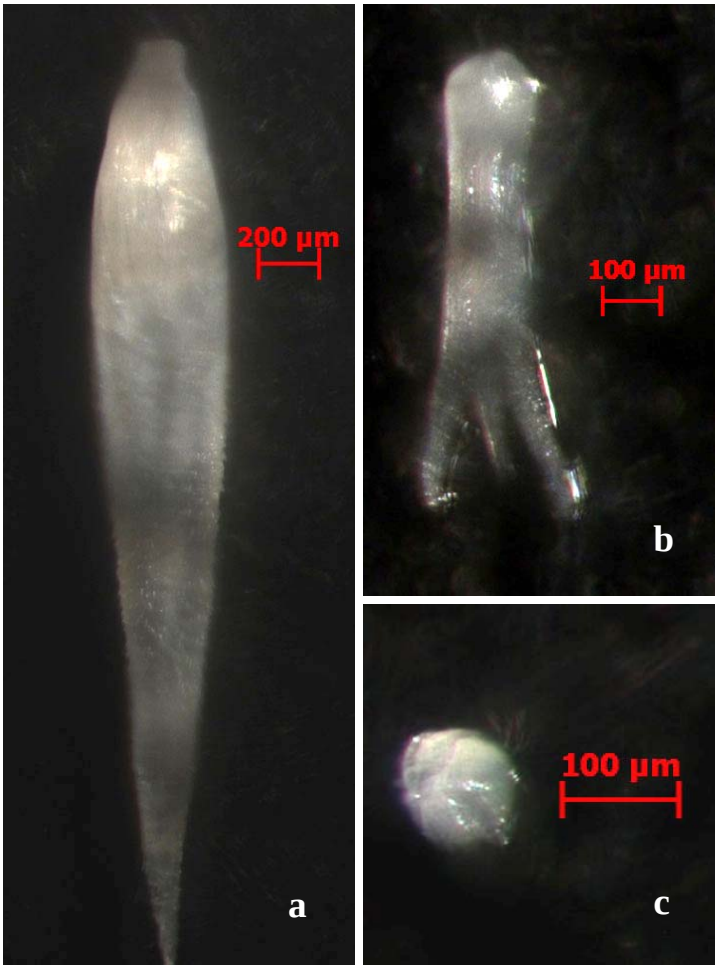


Figure 3.2 Microdissection of ovule. a, whole flower of Ps26 was dissected from the spikelet. b, intact pistil was dissected from the flower. c, ovular tissue was dissected from the pistil to get rid of most of the ovary wall.

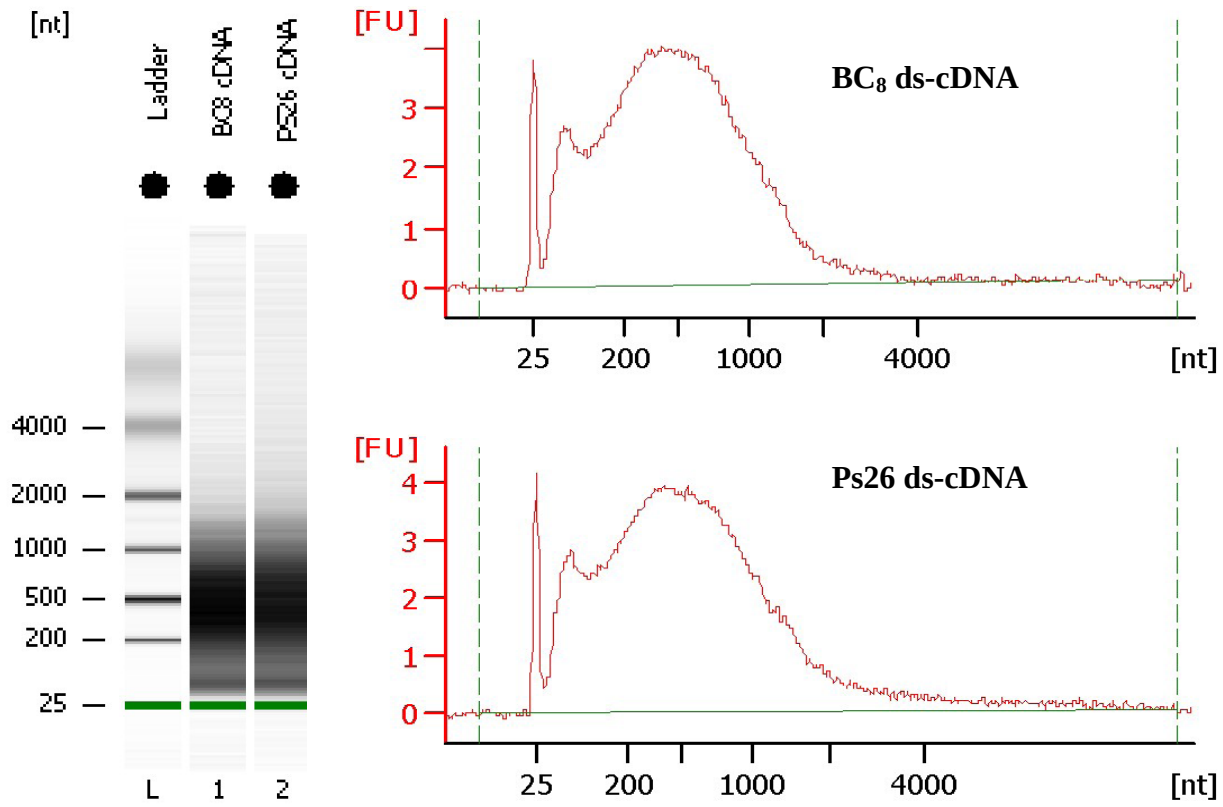


Figure 3.3 Agilent Bioanalyzer 2100 analysis result of the ds-cDNA samples.

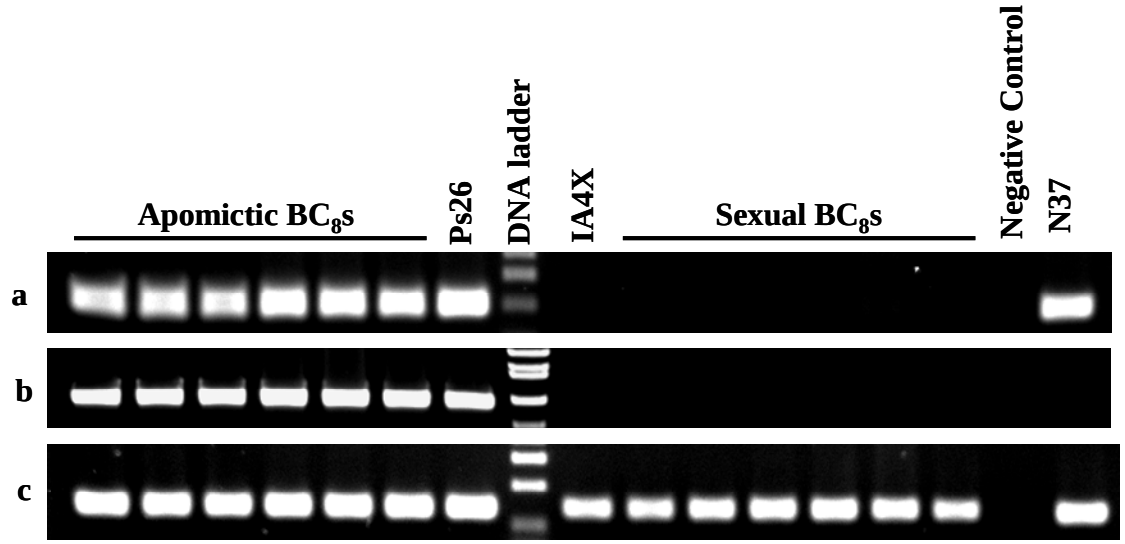


Figure 3.4 Examples for mapping of ovule-expressed transcripts to the ASGR-carrier chromosome. a, amplification from Ps26, N37 and apomictic BC_8 but not from IA4X or sexual BC_8 (P1510/1511). b, amplificationfrom Ps26 and apomictic BC_8 but not from IA4X, N37 or sexual BC_8 (P1514/1515). c, amplification from Ps26, IA4X, N37 and both apomictic and sexual BC_8 (P1504/1505).

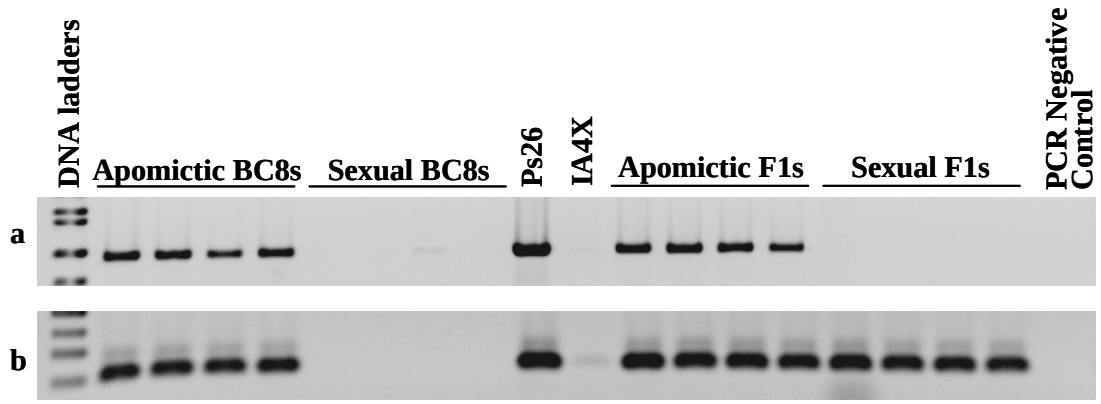


Figure 3.5 Examples for mapping of ovule-expressed transcripts to the ASGR. a, amplification of apomictic F_1s but not sexual F_1s (P1514/1515). b, amplification of both apomictic F_1s and sexual F_1s (P1506/1507).

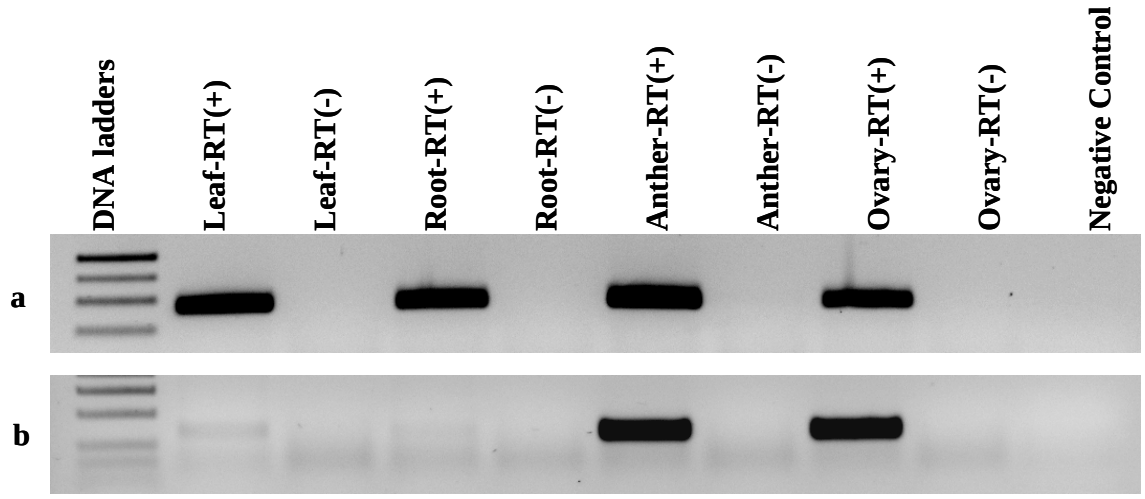


Figure 3.6 Examples of expression patterns for ASGR-carrier chromosome linked sequences. a, 34 out of 36 genes tested showed expression in all four organs tested. b, one gene (Ps26_c33813) was expressed only in ovary and anther. RT(+): reverse transcription performed with reverse transcriptase; RT(-):reverse transcription performed without reverse transcriptase; Negative Control: PCR was performed but template was replaced with distilled water.

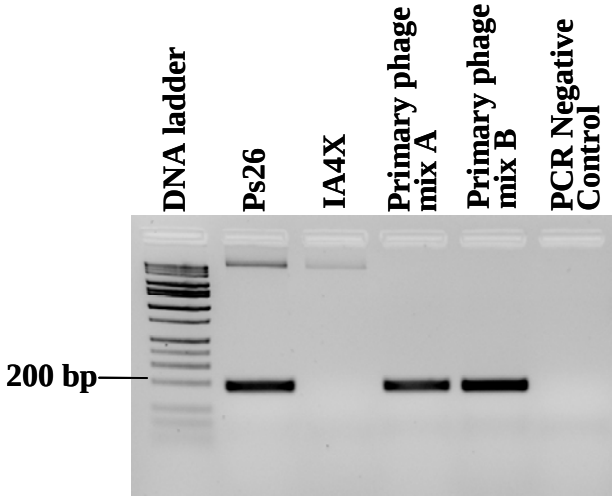


Figure 3.7 Example of PCR screening of primary phage plaques. The probe was amplified with primers P1492/P1493, primary phage mix A and B were the plaques recovered after primary screening.

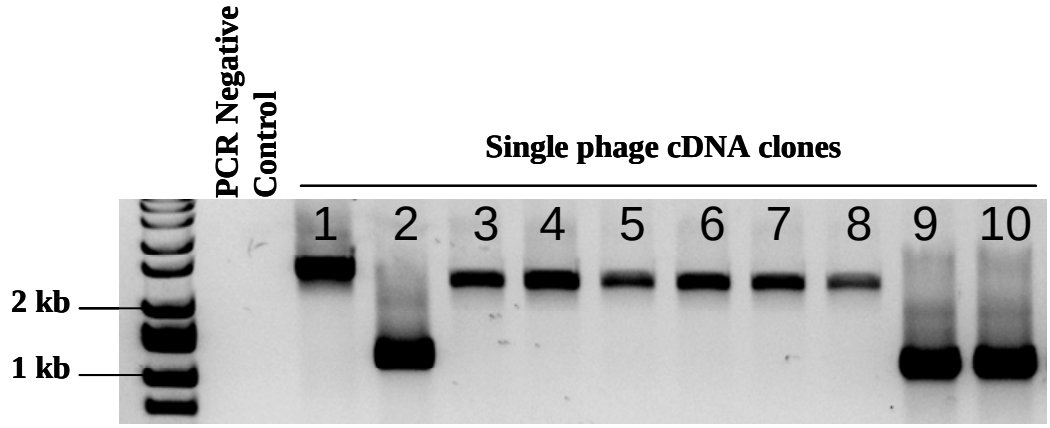


Figure 3.8 A set of recovered single phage colonies containing the desired cDNA inserts were amplified by PCR with M13 primers to test the insert size of phage cDNA clones. 3 to 8 are single phage clones corresponding to the same primary phage signal of contig Ps26_c2448. 9 and 10 are single phage clones corresponding to the same primary phage signal of contig Ps26_1312. 1 and 2 are single phage clones recovered from two different primary phage signals of contig Ps26_33813.

CHAPTER 4

COMPARATIVE ANALYSIS OF SEQUENCES AND EXPRESSION OF *BABY BOOM-LIKE* GENES IN BUFFELGRASS (*CENCHRUS CILIARIS*)

Abstract

Ectopic expression of the *Brassica napus BABY BOOM* has been reported to be sufficient for induction of embryo development from differentiated somatic cells. To investigate the function of *BABY BOOM-like* genes identified from an apospory-specific genomic region (ASGR) in *Cenchrus ciliaris*, *ASGR-BBM-like* and *non-ASGR-BBM-like* (*N-ASGR-BBM-like*) were subjected to sequence analysis and expression studies. Results demonstrated that both *ASGR-BBM-like* and *N-ASGR-BBM-like* genes were predicted to contain two AP2 domains; however, differences in the first AP2 domain and the area outside of the AP2 domains were identified. Semi-quantitative RT-PCR conducted with aposporous ovaries from different stages indicated that the transcription of *ASGR-BBM-like* genes was up-regulated upon pollination with a significant increase one day after pollination (DAP). In contrast, there was no expression of *N-ASGR-BBM-like* until one day after pollination. Moreover, the transcription level was much lower than that of *ASGR-BBM-like* at this time point. Temporal and quantitative differences in transcript level between the two suggest that *ASGR-BBM-like* may play a role in parthenogenesis in buffelgrass.

Introduction

Apomixis is a term used for asexual reproductive processes which occur during seed development (Nogler, 1984). The most important difference between apomixis and sexual

reproduction is that the embryo in an apomict is derived solely from cells in the ovule without involvement of meiosis and fertilization of the egg by a sperm (Koltunow, 1993; Grimanelli et al., 2001). Therefore, the progeny of apomictic plants are exact genetic replicas of the mother plant. Apomixis is widespread in angiosperms, having been observed in over 400 species spanning ~40 different families (Carman, 1997). However, no major agriculturally important crop has this trait. If apomixis can be introduced into crops, it could be an inexpensive way to perpetuate desired genotypes, including those demonstrating hybrid vigor, and thus be used as an advanced breeding tool in agricultural science (Hanna and Bashaw, 1987; Hanna, 1995; Koltunow et al., 1995; Savidan, 2000; van Dijk and van Damme, 2000).

Cenchrus ciliaris (syn. *Pennisetum ciliare*), commonly known as buffelgrass, is a warm season perennial forage grass predominantly reproducing by obligate aposporous apomixis. Sexual genotypes have been discovered as well (Bashaw, 1962). The number of chromosomes in different genotypes of buffelgrass can range from 32 to 54, while most genotypes are tetraploid ($2n = 4x = 36$) (Visser et al., 2000). Cytological studies of apomixis in buffelgrass have revealed that a highly hemizygous chromosome region, the ASGR (apospory-specific genomic region), is associated with the trait of apomixis (Ozias-Akins et al., 1998; Goel et al., 2003; Ozias-Akins et al., 2003; Akiyama et al., 2004; Akiyama et al., 2005). Characterization of the ASGR in buffelgrass by partial sequencing of ASGR-linked BAC (bacterial artificial chromosome) clones has led to the identification of genes with putative transcription factor or signaling related functions (Conner et al., 2008). Among these genes, one named *ASGR-BBM-like*, has high sequence similarity to the *BABY BOOM* (*BBM*) genes in *Brassica* and *Arabidopsis*, members of the AP2 (APETALA2)/EREBP multigene family (Boutilier et al., 2002).

The AP2/ERF multigene family is divided into two subfamilies: *AP2* genes with two AP2 domains and *ERF* genes with a single AP2 domain (Kim et al., 2006). The AP2 domain is a DNA-binding domain that consists of 60-70 conserved amino acid residues (Jofuku et al., 1994). Most reported genes in the AP2 subfamily are involved in the regulation of organ development in plants while the genes belong to the ERF subfamily usually function in signal transduction pathways of biotic and abiotic stress responses (Riechmann and Meyerowitz, 1998). *BBM* was one member of the AP2 subfamily identified from *Brassica napus* while investigating differentially expressed genes during the switch from pollen- to microspore-derived early embryo development (Boutilier et al., 2002). *BBM* encodes a transcription factor containing two AP2 domains and is preferentially expressed in developing embryos and seeds. Overexpression studies in *Arabidopsis* have shown that *BBM* activates signal transduction pathways leading to the induction of embryo development from differentiated somatic cells (Boutilier et al., 2002).

Although the initiation of sexual and apomictic pathways likely is activated by different signals, the two different developmental processes may converge during embryogenesis at an early stage of the signaling pathway (Ozias-Akins and van Dijk, 2007). However, sexual and apomictic pathways differ in their requirement for egg fertilization – in apomixis, the egg develops via parthenogenesis. Based on the similarity of *ASGR-BBM-like* to embryogenesis genes, we hypothesize that the apomict-specific *ASGR-BBM-like* gene may be involved in parthenogenesis. In order to test whether the expression pattern of *ASGR-BBM-like* was consistent with a putative function in parthenogenesis and different from related genes in the buffelgrass genome, we compared expression patterns between *ASGR-BBM-like* and *non-ASGR-BBM-like* (*N-ASGR-BBM-like*) genes in apomictic and sexual genotypes of buffelgrass by semi-quantitative RT-PCR. Expression of *ASGR-BBM-like* preceding *N-ASGR-BBM-like* in the

apomict suggests that *ASGR-BBM-like* is likely to be a regulator of parthenogenesis in buffelgrass.

Materials and Methods

Plant Materials

Two genotypes of buffelgrass (*Cenchrus ciliaris* or *Pennisetum ciliare*, 2n=4x=36) used in this study were maintained by vegetative propagation (Roche et al. 1999). One is the obligate aposporous B-12-9, and the other is the obligate sexual B-2s. Plants were grown in the greenhouse with a continuous temperature ranging from 24°C to 30°C.

Nucleic Acid Extraction

Genomic DNA was extracted either following the protocol previously described (Ozias-Akins et al., 1993) or with QIAGEN DNeasy® Plant Mini kit (QIAGEN, Valencia, CA, USA) following the manufacturer's protocol. DNA quantification was conducted with a spectrophotometer and measured with the absorbance at 260/280. Total RNA was extracted with QIAGEN RNeasy® Plant Mini kit and quantified with RiboGreen® RNA Quantitation Kit (Molecular Probes, Eugene, OR, USA) following the manufacturers' protocols. BAC DNA for shotgun library construction was extracted with QIAGEN Large Construct Kit.

Southern Blot Analysis

Genomic DNA (~12 μg) was digested with 50 units of *Hin*d III restriction enzyme (New England Biolabs, Beverly, MA, USA) in a 40 μl reaction volume at 37°C overnight. Digested DNA fragments were separated by electrophoresis in a 1% agarose gel with 1× TBE buffer. The gel was blotted to Genescreen Plus nylon membrane (NEN Life Sciences, Boston, MA, USA) and probed following standard protocols (Sambrook and Russell, 2001). The membrane was pre-hybridized in 30 ml hybridization buffer (6× SSC, 1% SDS, 100 μg ml^{-1} salmon sperm

DNA) at 65°C overnight. Probes were labeled with α-^{32}P following the instructions of Random Prime Labeling kit (Invitrogen, Carlsbad, CA, USA) and cleaned with Sephadex® G-50 (Sigma, St. Louis, MO, USA). Labeled probe was denatured at 95°C for 10 min and placed immediately on ice before adding into the hybridization buffer. Hybridization was conducted at 65°C overnight followed by four washes at the same temperature for 15 min each with the following buffers: 1) 2× SSC, 0.1% SDS; 2) 1× SSC, 0.1% SDS; 3) 0.5× SSC, 0.1% SDS; and 4) 0.1× SSC, 0.1% SDS. The membrane was exposed and the signal was detected with the Storm phosphor imager system (Amersham Biosciences, Pittsburgh, PA, USA).

BAC Library Screening

BAC libraries from two apomictic genotypes, a polyhaploid from *P. glaucum* X *P. squamulatum* and buffelgrass were constructed previously (Roche et al., 2002), and BAC clones containing ASGR-linked markers have been isolated and characterized (Akiyama et al., 2004; Goel et al., 2006; Conner et al., 2008). To identify BAC clones containing *ASGR-BBM-like* homologous genes (e.g., *N-ASGR-BBM-like*), PCR primers based on the sequence information of *ASGR-BBM-like* were designed to amplify the 566 bp 5'-end fragment (P1015/P1016, 5'-GCGACAAGAGCCCACACCCT-3'/5'-GTCCGCTGCCCGAAAGTGTC-3'), the 462 bp AP2 domain-containing fragment (P1017/P1018, 5'-AGGCATAGATGGACAGGAAGG-3'/5'-GTAGTAGTTGTGGTGCCCCG-3') and the 819 bp 3'-end fragment (P294/295, 5'-GGCACTCCAAAGCGTCTCAAGGAA-3'/5'-CTCCCACTGGAGCTGATTAAGCAG-3') of *ASGR-BBM-like*. PCR was carried out in a total volume of 20 µl containing ~20 ng BAC DNA p208, which contains the genomic fragment of *ASGR-BBM-like*, 1 µM of each primer, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 1 unit of *Taq* DNA polymerase (Sigma). The reaction was denatured at 94°C for 5 min and then subjected to 35 cycles of 94°C denaturation

for 30 seconds; annealing temperatures appropriate for each primer pair (61°C for P1015/P1016 and P294/P295, 58°C for P1017/P1018) for 30 seconds, and 72°C for 1 min. The final extension was 7 min at 72°C. PCR products were separated by electrophoresis and purified with QIAquick® Gel Extraction Kit (QIAGEN). Each probe was labeled with α-^{32}P by PCR labeling in a 20 µl reaction, including 25 ng DNA fragment, 1.25 U of Jumpstart Taq polymerase (Sigma), 0.25 µM of each primer, 0.5 mM dNTP mixture (excluding dCTP), 3 µl of α-^{32}P-labeled dCTP (3000 Ci/mmol) and 1× PCR buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$). Hybridization and identification of addresses for positive clones were performed as described at www.genome.clemson.edu/groups/bac/protocols. Signals were detected with the Storm phosphor imager.

Fingerprinting Analysis and Grouping of BAC Clones

BAC clones hybridized with any of the three probes of *ASGR-BBM-like* were picked and cultured in 2 ml LB medium containing 25 µg ml^{-1} chloramphenicol at 37°C overnight on a rotary shaker. Plasmid isolation was carried out with standard alkaline lysis. For each BAC clone, 7.5 µl DNA was used for a 20 µl *Hin*d III restriction enzyme digestion reaction at 37 °C overnight. The digested DNA fragments were then separated on a 1.2% agarose gel in 1X TAE buffer and stained with SYBRGreen® dye (Molecular Probes, Eugene, OR, USA). Gels were imaged with the Molecular Imager Gel Doc XR System (Bio-Rad Laboratories, Hercules, CA, USA). Digestion patterns of BAC clones were compared and BAC clones were grouped based on the similarity of digestion patterns. The gel was also blotted to Genescreen Plus nylon membrane and hybridized with α-^{32}P-labeled *ASGR-BBM-like* probes to confirm the contig result by comparing the hybridization pattern to the digestion pattern.

Shotgun Library Construction

One BAC clone from each of the two major groups was randomly selected for shotgun library construction and named as C1400 and C1500. BAC clone culture and DNA extraction was as described in Conner et al. (2008). BAC DNA (10 μg) was sheared by a GeneMachines Hydroshear DNA shearing device (Genomic Instrumentation Services, Inc., San Carlos, CA, USA) to produce fragments of 2.0-4.0 kb. DNA fragments were blunt-end repaired using a combination of T4 DNA polymerase, Klenow DNA polymerase and T4 polynucleotide kinase and purified using the Qiaquick PCR Purification Kit (QIAGEN). Vector pBluescript (SK-) (Stratagene, La Jolla, CA, USA) was digested with *Eco*RV and dephosphorylated before ligation with the DNA fragments. Ligation products were transformed into DH10B competent cells (Stratagene) and grown on solid LB medium with 100 μg ml^{-1} ampicillin at 37 ºC overnight. Single clones from each library were picked and placed into two to three 384-well plates containing freeze-broth and 100 μg ml^{-1} ampicillin. The bacteria were cultured overnight at 37ºC. Each 384-well plate was replicated onto a membrane which was put on top of solid LB medium containing 100 μg ml^{-1} ampicillin and cultured overnight at 37ºC. On the second day, the colony membranes were denatured and neutralized before baking at 80ºC for 2 h. Baked colony blots were hybridized to α-^{32}P-labeled *BBM-like* probes as described in BAC library screening. Positive subclones were sequenced using both the T7 and M13REV primers at the University of Georgia Office of Research Services.

Sequencing Data Analysis

Sequences were analyzed using VectorNTI Advanced 8 (Invitrogen). Sequences were trimmed of vector and assembled into contigs using ContigExpress in the VectorNTI package. Contig sequences from C1400 and C1500 were aligned with DNA sequences of *ASGR-BBM-like*

genes, EU559279 and EU559278, with AlignX in the VectorNTI package. The nucleotide sequences of the two *N-ASGR-BBM-like* genes were analyzed with the rice gene prediction program at RiceGAAS (http://ricegaas.dna.affrc.go.jp/usr/) for predicted transcripts and proteins. The nucleotide sequences were used to perform phylogenetic analysis using the neighbor-joining method and the Kimura 2-parameter model with 1,000 bootstrap replicates. The predicted protein sequences were used to perform phylogenetic analysis using the neighbor-joining method and the PAM matrix model with 1,000 bootstrap replicates.

Tissue Collection for RT-PCR

Root and leaf tissues

Seeds of B-12-9 and B-2s were germinated in Petri dishes with pre-wet filter paper in the growth chamber at 28°C for one week. Germinated seeds were then planted individually in small pots and allowed to grow for 6 weeks. Roots and leaves were collected in liquid nitrogen for total RNA extraction.

Ovaries and anthers

Young inflorescences from vegetatively propagated B-12-9 and B-2s grown in the greenhouse were covered with pollinating bags during the flowering season to prevent cross-pollination. The bagged heads were checked every morning to select tissue for the collection stages, which included one day before pollination (DBP), the day of pollination (DOP) and one day after pollination (DAP). At DBP, where the stigmas were fully exserted and the anthers had not yet exserted, about 50 spikelets from the middle half of the head were collected. Sixty ovaries and thirty anthers were then dissected from the hermaphroditic florets and collected in RLT buffer (RNAeasy Mini kit, QIAGEN). The remaining part of the head was checked the following morning to confirm the DBP stage. At DOP, the anthers had begun to exsert but

collection was prior to pollen shed. DAP was defined as one day after pollen shed. Each tissue type was collected with independent triplicates.

Semi-quantitative RT-PCR

Total RNA samples were treated with deoxyribonuclease I (DNase I, Invitrogen) and quantified with a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA) before first-strand cDNA synthesis following the manufacturer's protocol of First-strand cDNA Synthesis kit (Invitrogen). RT-PCR reactions were performed in a total volume of 20 µl containing 2 µl of first-strand cDNA, 1 uM of each primer, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 1 unit of Taq DNA polymerase (Sigma). The reaction was denatured at 94°C for 5 min and then subjected to different cycles of 94°C denaturation for 30 seconds, 53°C (P779/P780: 5'-TATGTCACGACAAGAATATG-3' / 5'-TGTAACCATAACTCTCAGCT-3'; P1333/P1334: 5'- AAAGCAGCGGGTTTTCTC-3' / 5'- GATTCAGCACCGTAGCCA-3') or 59°C (P1554/P1555: 5'- AGGGTGGTGCCAAGAAGGTTA-3' / 5'-GTAGCCCCACTCGTTGTCGTA-3') annealing for 30 seconds, and 72°C extension for 1 min. To determine the cycle numbers of the linear range of the PCR for different primers, PCR products were collected after 22, 25, 28, 31, 34, 37 and 40 cycles and separated on agarose gel. Twenty-one cycles and 37 cycles were determined for the internal control glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and for both *ASGR-BBM-like* and *N-ASGR-BBM-like* genes. Within each experiment and for each gene analyzed, the complete set of samples was processed in parallel in single PCR using aliquots of the same master mix. RT-PCR experiments were carried out on six types of tissue from the two genotypes of buffelgrass. Each set of RT-PCR reactions was performed in triplicate and each time point/tissue was processed with biological triplicates. RT-PCR products were separated on a 1.5% agarose gel and stained with ethidium

bromide. Gel images were taken with the Molecular Imager Gel Doc XR System (Bio-Rad Laboratories, CA) and the pixels of each band were quantified with ImageQuant software.

Cloning and Sequencing of RT-PCR Products

The RT-PCR products from B-12-9 DOP ovaries with *ASGR-BBM-like* and *N-ASGR-BBM-like* specific primers were separated with 1.5% agarose gels (Invitrogen) and purified with QIAquick PCR purification kit (QIAGEN). Purified fragments were ligated with the PCR4-TOPO vector (Invitrogen) and transformed into One Shot TOP10 *E. coli* DH5α competent cells via the heat-shock method following the manufacturer's instructions (Invitrogen). Single colonies were selected for sequencing. Sequencing was carried out with a CEQ 8000 Genetic Analysis System.

Light Microscopy Observation of Cleared Pistils

Pistils at stages of DBP, DOP and DAP were fixed in FAA and cleared with methyl salicylate according to Young's protocol (Young et al., 1979). Whole cleared ovaries were mounted in the same clearing media and observed with differential interference contrast (DIC) optics of an Axioskop 2 plus microscope (Zeiss, Jena, Germany).

Statistical Analysis

Analysis of variances (ANOVA) was performed to determined differences in expression levels at different stages. The significant level was set at $\alpha = 0.05$. All statistical analysis was carried out with SAS 9.1 (SAS Institute Inc., Cary, NC, USA).

Results

Identification of *N-ASGR-BBM-like* Genes in Buffelgrass

A BAC library of apomictic buffelgrass B-12-9 was screened with the probes containing 5'-end, AP2 domain and 3'-end sequences of previously identified *ASGR-BBM-like* to identify

the other possible *BBM-like* genes existing in buffelgrass. Twenty-seven BAC clones in total hybridized with individual probes while eighteen clones hybridized to all the three probes, indicating that these BAC clones possibly contained the full sequences of *BBM-like* genes (Table 4.1). Six out of eighteen were the previously identified BAC clones containing the *ASGR-BBM-like* genes; therefore, the remaining twelve BAC clones should contain genes similar to the *ASGR-BBM-like* gene in sequence but located outside of the ASGR. These genes were named *non-ASGR-BBM-like* (*N-ASGR-BBM-like*).

Southern blot analysis was conducted to confirm the existence of *N-ASGR-BBM-like* genes and estimate their copy numbers in buffelgrass. The AP2 domain containing probe, which also was used in BAC library screening, hybridized to two common fragments in sexual B-2s and apomictic B-12-9 buffelgrass and to an additional unique fragment in B-12-9 (Fig. 4.1). *ASGR-BBM-like* genes exist only in apomictic B-12-9 which is confirmed by PCR with *ASGR-BBM-like* specific primers P779/P780; therefore, the unique fragment corresponded to ASGR-*BBM-like* and the signals found in both genotypes were classified as *N-ASGR-BBM-like* signals.

The majority of 21 *N-ASGR-BBM-like* containing BAC clones were further subdivided into two groups based on *Hin*dIII fingerprinting (Fig. 4.2a). The putative BAC contigs were confirmed by the hybridization patterns of *Hin*dIII-digested BAC DNA to the mixture of the three *ASGR-BBM-like* probes (Fig. 4.2b) and *Eco*RV-digested BAC DNA to the AP2 domain containing probe (Fig. 4.2c). In order to obtain the sequences of the *N-ASGR-BBM-like* genes, one BAC was picked from each of the two major groups for shotgun library construction and named as C1400 (original BAC address: C091M03) and C1500 (original address: C164L13), respectively (shown in Fig. 4.2b with arrows). Shotgun library screening with the three *ASGR-BBM-like* probes identified 20 positive subclones from C1400 and 23 positive subclones from

C1500, respectively. Sequencing of the positive subclones yielded 55 and 49 high quality sequences from C1400 and C1500, correspondingly. Two contigs were assembled from C1400 sequences with lengths of 3825 bp and 1396 bp while one contig with a length of 5777 bp was assembled from C1500 subclone sequences. Based on the DNA sequence similarity with the coding region of *CcASGR-BBM-like1* and *CcASGR-BBM-like2* sequences, two *N-ASGR-BBM-like* genes were identified: the *CcASGR-BBM-like1400* gene derived from BAC C1400 contains 3222 bp and the *CcASGR-BBM-like1500* gene derived from BAC C1500 contains 3355 bp.

Comparative Sequence Analysis between *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* Genes

Alignment of the four *BBM-like* genes from buffelgrass is shown in Figure 4.3 and the similarity among these four genes is shown in Table 4.2. *CcN-ASGR-BBM-like* genes shared 70% (C1400) and 68-69% (C1500) nucleotide sequence identity with *CcASGR-BBM-like* while there is 95% identity between the two *CcN-ASGR-BBM-like* genes. A phylogenetic tree of the four *C. ciliaris BBM-like* genes was derived based on the nucleotide sequence similarity (Figure 4.4).

When analyzed with the rice gene prediction program at RiceGAAS (http://ricegaas.dna.affrc.go.jp/usr/), both of the two *N-ASGR-BBM-like* genes were predicted to contain seven exons (highlighted in red in Fig. 4.3) and two AP2 domains (highlighted with pink rectangles in Fig. 4.5), similar to the two *CcASGR-BBM-like* genes (Fig. 4.3 and Fig. 4.5). The two *N-ASGR-BBM-like* genes differ in both predicted coding and non-coding regions with the majority of the difference found in the non-coding region and the seventh exon. One nucleotide change in the seventh exon of *CcN-ASGR-BBM-like1400* caused a potential stop codon (highlighted in pink rectangle in Fig. 4.3) resulting in a shorter protein. The predicted splicing

pattern for each gene is shown in Figure 4.5. *CcN-ASGR-BBM-like1400* and *CcASGR-BBM-like1500* encode 573 amino acid and 627 amino acid proteins, respectively, and they share 89% identity. Alignment of the predicted amino acid sequences from all four *C. ciliaris BBM-like* genes showed that the similarity between *CcN-ASGR-BBM-like* and *CcASGR-BBM-like1* were 64% (C1500) and 60% (C1400) while they shared 56% (C1500) and 52% (C1400) identify with *CcASGR-BBM-like2* (Fig. 4.6 and Table 4.3). Predicted splice site changes in exons 2, 3, 4 and 5 caused the main difference between *CcN-ASGR-BBM-like* and *CcASGR-BBM-like* predicted proteins. The structure of the phylogenetic tree of the predicted amino acid sequences was similar to that of the genomic sequences with the two *ASGR-BBM-like* genes classed in one subgroup separate from the two *N-ASGR-BBM-like* genes (Fig. 4.7).

When the two AP2 domains were analyzed, we found that the second AP2 domain was highly conserved among the four genes with only one amino acid change between *CcN-ASGR-BBM-like* and *CcASGR-BBM-like* (Fig. 4.6) and this domain is identical between the two *CcN-ASGR-BBM-like* proteins as well as between *Cc-ASGR-BBM-like1* and *CcASGR-BBM-like2*. The case with the first AP2 domain is quite different. Predicted splicing site changes caused a deletion of part of the first AP2 domain in *CcASGR-BBM-like2* protein and an insertion in the middle of the first AP2 domain in *CcN-ASGR-BBM-like* proteins (Fig. 4.6).

Expression Patterns of *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* Genes

Semi-quantitative RT-PCR with *CcASGR-BBM-like*/*CcN-ASGR-BBM-like* specific primers showed that both *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* genes were expressed in roots and anthers with similar expression levels (Figure 4.8a and 4.8b). There was very low expression of Cc*N-ASGR-BBM-like* in leaves of B-2s while no expression of *CcASGR-BBM-like* in leaf tissue of B-12-9 was detected (Figure 4.8a and 4.8b). In B-12-9 ovaries, expression of

CcASGR-BBM-like was up-regulated upon pollination with more than 3-fold increase at DAP, whereas there was no significant difference in expression between DBP and DOP (Table 4.4). The expression of *CcN-ASGR-BBM-like* in B-12-9 ovaries at the three stages analyzed was quite different in comparison to the expression of *CcASGR-BBM-like* (Figure 4.8c). There was no detectable expression at DBP and DOP and even at DAP the transcript level was only about 7% of *CcASGR-BBM-like* (Table 4.4). In B-2s ovaries, the expression of *CcN-ASGR-BBM-like* was significantly up-regulated after pollination as well (Figure 4.8d); the expression at DAP was about 3 times more than that at DBP or DOP with no significant difference between DBP and DOP (Table 4.4). The average value of three biological replicates also showed that the expression of *CcASGR-BBM-like* in B-12-9 ovaries was higher that that of *CcN-ASGR-BBM-like* in B-2s ovaries at each stage we analyzed (Table 4.4).

Small scale sequencing of RT-PCR products amplified from B-12-9 DAP ovary with primers P779/P780 and P1333/P1334 returned with 16 and 13 good quality sequences amplified from *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* specific primers, respectively. After alignment with the predicted cDNA sequences, we found that 3 out of 16 *CcASGR-BBM-like* sequences actually matched the predicted cDNA of *CcASGR-BBM-like2* and the other 13 matched *CcASGR-BBM-like1*. In the group of *CcN-ASGR-BBM-like* sequences, 3 out of 13 had a 100% match with *CcNASGR-BBM-like1400* and the remaining 10 sequences were from *CcN-ASGR-BBM-like1500*.

Embryo Sac Observation in Ovaries

Fifty whole-mounted cleared pistils of B-12-9 at stage DOP were observed for the number of embryo sacs present in each ovule and the state of the egg cells. The number of

embryo sacs in each ovule varied from one to four with an average of 1.9. Conspicuous egg cells with vacuoles and a clear nucleus were observed in each of the ovules examined.

Discussion

Sequence Analysis of *BBM-like* Genes

Analyzed with the gene prediction program at RiceGAAS, the four BBM-like genes from buffelgrass were all predicted to produce translatable transcripts with predicted proteins containing two AP2 domains. Different splice site predictions caused differences in the first AP2 domain among the four genes analyzed. The predicted first AP2 domain in *Cc-ASGR-BBM-like1* had the highest similarity to the AP2 domain in *BnBBM1* and *BnBBM2* genes and contained most of the conserved amino acid residues. Thus the differences in splice sites could either be real and biologically significant or prediction errors. Prediction artifact seems more likely based on our unpublished data from the *PsASGR-BBM-like1* (EU559277) gene. The full-length cDNA of *PsASGR-BBM-like1* gene obtained by rapid amplification of cDNA ends (RACE) had an additional small exon of 9 bp (3 amino acids) compared to the predicted cDNA. To confirm the authenticity of the predictions, the cDNA sequences covering the first AP2 domain will have to be obtained by RACE.

Gene Expression during Embryogenesis

In apomixis the unreduced embryo spontaneously develops from the unreduced egg cell without the fertilization of the egg by a sperm, i.e., via parthenogenesis (Koltunow, 1993; Koltunow and Grossniklaus, 2003; Ozias-Akins, 2006). With the increasing intensity of research towards unveiling the genetic control of apomixis, a hypothesis has been brought to developmental biologists that apomixis arises through deregulation of the sexual developmental

pathway. Thus, the genes functioning in signaling pathways for zygotic embryogenesis may also be relevant to parthenogenesis.

Attempts to identify genes controlling embryogenesis have led to the isolation of a number of genes that play important roles in this developmental process, including *LEAFY COTYLEDON* (*LEC*) (Meinke et al., 1994; Fambrini et al., 2006; Braybrook et al., 2008), *SOMATIC EMBRYOGENESIS RECEPTOR KINASE* (*SERK1*) (Hecht et al., 2001; Nolan et al., 2009), *WUSCHEL* (*WUS*) (Zuo et al., 2002; Su et al., 2009) and *BABY BOOM* (*BBM*) (Boutilier et al., 2002; Malik et al., 2007; Passarinho et al., 2008). Overexpression of *BnBBM* induced somatic embryo formation on leaf tissues of *Arabidopsis* (Boutilier et al., 2002). *BnBBM1* was one of the embryo-expressed genes transcribed in microspore-derived embryos of *Brassica napus* during the induction and development stages and also in developing zygotic seeds (Malik et al., 2007). Ectopic *BBM* expression in tobacco also activated cell proliferation pathways (Srinivasan et al., 2007). Characterization of one *BBM-like* gene *EgAP2-1*, isolated from oil palm, suggested that it was an important regulator in both zygotic and somatic embryo development (Morcillo et al., 2007). Identification of target genes directly activated by *BBM* in Arabidopsis seedlings suggested that *BBM* played a role during cell proliferation and growth by activating a complex network of developmental pathways (Passarinho et al., 2008). All these functional analyses of *BBM* suggest that it plays a key role during embryo development in plants. The apomixis linked candidate gene *ASGR-BBM-like* is an AP2 domain containing gene and its nucleotide sequence and predicted protein sequence showed high similarity to *BBM* genes from rice, *Medicago*, *Brassica* and *Arabidopsis*, which also contain two AP2 domains. Semi-quantitative RT-PCR showed that *CcASGR-BBM-like* had continuous expression in aposporous ovaries before and after pollination with a significant increase after pollination. The sequence

similarity with conserved domains and the expression pattern of *CcASGR-BBM-like* indicate its potential to function during parthenogenesis.

Differences in Transcription between *ASGR-BBM-like* and *N-ASGR-BBM-like* Genes

Allelic variation can cause dramatic phenotypic differences. One example is *fw2.2*, a fruit weight key quantitative trait locus (QTL) for domestication of tomatoes, modulating the tremendous fruit size change and shape variation between wild and domesticated tomato species (Cong et al., 2002; Nesbitt and Tanksley, 2002; Tanksley, 2004). It has been indicated by comparative sequence analysis that the fruit weight variation attributable to *fw2.2* is likely modulated through the transcriptional variation associated with the changes in gene regulation rather than in the FW2.2 protein itself (Nesbitt and Tanksley, 2002). A pair of nearly isogenic lines was subjected to detailed developmental analysis to investigate the consequences of regulatory difference in *fw2.2* alleles on fruit size change (Cong et al., 2002). The results showed that alleles corresponding to large and small fruit differed in timing of peak expression by about one week. Moreover, the total transcript level differed between the alleles as well. Thus it was suggested that the heterochronic allelic variation, combined with quantitative differences in total transcript levels, accounted for the major difference in tomato fruit weight (Cong et al., 2002).

The expression of both *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* in root and anther was strong, similar to *BnBBM1*, which was also highly expressed in root tissue (Malik et al., 2007). This suggests a potential role in stimulating cell proliferation for both *ASGR-BBM-like* and *N-ASGR-BBM-like*, which was reported for the *BBM* genes in *Arabidopsis* and tobacco (Boutilier et al., 2002; Srinivasan et al., 2007). However, expression in ovaries was quite different between *CcASGR-BBM-like* and *CcN-ASGR-BBM-like*. In B-12-9, transcript levels of

CcASGR-BBM-like were low at stages DBP and DOP, but the transcript level increased about four-fold at DAP compared to the earlier stages. In contrast, there was no detectable transcription of *CcN-ASGR-BBM-like* at DBP and DOP until stage DAP, in which it started to show a very low level of transcription. In aposporous ovaries transcription of *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* differs in both the timing and the amount, e.g. the difference was both qualitative and quantitative. When the expression pattern of *CcN-ASGR-BBM-like* in the three stages of B-2s (sexual) ovaries was analyzed, we found that it was somewhat similar to that of *CcASGR-BBM-like* in B-12-9 (aposporous) ovaries. There was low expression at the first two stages with an increase at stage DAP, although the increase was much slower for *CcN-ASGR-BBM-like* in B-2s. Whether the transcription differences are caused by protein interactions or heterochronic regulatory changes is still unknown. The sequence variation between *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* was sufficient to classify them into two different branches of the phylogenetic tree and the predicted splicing changes also resulted in protein differences, especially in the first AP2 domain region.

Correlations between Gene Expression and Parthenogenesis in Buffelgrass

Based on semi-quantitative RT-PCR data, expression of *CcASGR-BBM-like* in apomictic ovaries was earlier and more abundant than that of *CcN-ASGR-BBM-like*. In order to check the possible correlation between the *CcASGR-BBM-like* expression and precocious aposporous embryo development, ovaries at DOP were cleared and observed under a microscope. Although there were no well-formed proembryos observed in the ovules, varied numbers of conspicuous egg cells (one to several) with clear nuclei were observed in each ovule examined. Parthenogenetic development of the egg cell in unpollinated ovules of buffelgrass has been reported while the variable frequencies of proembryos observed suggested that parthenogenesis

could be initiated at different times (Vielle, et al., 1995). Since the frequency of precocious embryo formation could vary from 7% to 27% between two different genotypes studied, our observation could be result of different genotypes. The cytoplasmically rich egg cells had cytological characteristics of activated egg cells and could have initiated parthenogenesis.

In aposporous buffelgrass ovaries, transcription of *CcASGR-BBM-like* was intiated before pollination and up-regulated upon pollination with a significant increasement at one day after pollination; however, the expression of *CcN-ASGR-BBM-like* was not detected before pollination and the transcription level at one day after pollination was significantly lower than that of *CcASGR-BBM-like* at the same stage. In contrast to that, the expression of *CcN-ASGR-BBM-like* in sexual buffelgrass ovaries was detected before pollination and the expression pattern was similar to that of *CcASGR-BBM-like* in aposporous ovaries with a lower transcription level at each of the stage analyzed. With respect to the important role of *BBM* genes in induction of embryonic cell development in sexually reproducing plants and the different expression patterns of *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* in aposporous ovules, it seemed that *CcASGR-BBM-like* may affect the expression of *CcN-ASGR-BBM-like* in aposporous ovules and involve in the initiation and/or maintenance of parthenogenesis of aposporous embryo. The expression data of *CcN-ASGR-BBM-like* in sexual buffelgrass, B-2s, indicated its potential function in zygotic embryogenesis. Functional knock down experiments by RNA interference (RNAi) also suggested that *ASGR-BBM-like* promoted aposporous embryo initiation and development in *Pennisetum squamulatum* (unpublished data); however, additional functional analysis of both *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* in buffelgrass will be necessary to confirm this hypothesis.

References

Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in Pennisetum squamulatum reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. Plant Physiol **134:** 1733-1741

Akiyama Y, Hanna WW, Ozias-Akins P (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. Theor Appl Genet **111:** 1042-1051

Bashaw EC (1962) Apomixis and sexuality in buffelgrass. Crop Science **2:** 412-415

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, Custers JB, van Lookeren Campagne MM (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. Plant Cell **14:** 1737-1749

BrayBrook SA, Harada JJ (2008) LECs go crazy in embryo development. Trends in Plant Science **13:** 624-630

Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Linn Soc **61:** 51-94

Cong B, Liu J, Tanksley SD (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc Natl Acad Sci U S A **99:** 13606-13611

Conner JA, Goel S, Gunawan G, Cordonnier-Pratt MM, Johnson VE, Liang C, Wang H, Pratt LH, Mullet JE, Debarry J, Yang L, Bennetzen JL, Klein PE, Ozias-Akins P

(2008) Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of Pennisetum and Cenchrus. Plant Physiol **147:** 1396-1411

Fambrini M, Durante C, Cionini G, Geri C, Giorgetti L, Michelotti V, Salvini M, Pugliesi C (2006) Characterization of LEAFY COTYLEDON1-LIKE gene in Helianthus annuus and its relationship with zygotic and somatic embryogenesis. Dev Genes Evol **216:** 253-264

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

Goel S, Chen Z, Conner JA, Akiyama Y, Hanna WW, Ozias-Akins P (2003) Delineation by fluorescence in situ hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in Pennisetum squamulatum and Cenchrus ciliaris. Genetics **163:** 1069-1082

Grimanelli D, Leblanc O, Perotti E, Grossniklaus U (2001) Developmental genetics of gametophytic apomixis. Trends Genet **17:** 597-604

Hanna WW (1995) Use of apomixis in cultivar development. Adv Agron **54:** 333–350

Hanna WW, Bashaw EC (1987) Apomixis: its identification and use in plant breeding. Crop Science **27:** 1136–1139

Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt ED, Boutilier K, Grossniklaus U, de Vries SC (2001) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol **127:** 803-816

Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. Plant Cell **6:** 1211-1225

Sangtae Kim, Pamela S. Soltis, Kerr Wall, Douglas E. Soltis (2006) Phylogeny and domain evolution in the APETALA2-like gene family. Mol Biol Evol **23:**107–120

Koltunow AM (1993) Apomixis: Embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell **5:** 1425-1437

Koltunow AM, Bicknell RA, Chaudhury AM (1995) Apomixis: Molecular strategies for the generation of genetically identical seeds without fertilization. Plant Physiol **108:** 1345-1352

Koltunow AM, Grossniklaus U (2003) Apomixis: a developmental perspective. Annu Rev Plant Biol **54:** 547-574

Malik MR, Wang F, Dirpaul JM, Zhou N, Polowick PL, Ferrie AM, Krochko JE (2007) Transcript profiling and identification of molecular markers for early microspore embryogenesis in Brassica napus. Plant Physiol **144:** 134-154

Meinke DW, Franzmann LH, Nickle TC, Yeung EC (1994) Leafy cotyledon mutants of Arabidopsis. Plant Cell **6:** 1049-1064

Morcillo F, Gallard A, Pillot M, Jouannic S, Aberlenc-Bertossi F, Collin M, Verdeil JL, Tregear JW (2007) EgAP2-1, an AINTEGUMENTA-like (AIL) gene expressed in meristematic and proliferating tissues of embryos in oil palm. Planta **226:** 1353-1362

Nesbitt TC, Tanksley SD (2002) Comparative Sequencing in the Genus Lycopersicon: Implications for the Evolution of Fruit Size in the Domestication *of* Cultivated Tomatoes. Genetics **162:** 365-379

Nogler GA (1984) Gametophytic apomixis, In: Embryology of Angiosperms. Johri, B.M., Ed., Springer-Verlag, Berlin, pp: 475-518

Nolan KE, Kurdyukov S, Ray JR (2009) Expression of the SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 (SERK1) gene is associated with developmental change in the life cycle of the model legume Medicago truncatula. Journal of Experimental Botany **60**: 1759-1771

Ozias-Akins P (2006) Apomixis: Developmental characteristics and genetics. Critical Reviews in Plant Sciences **25:** 199-214

Ozias-Akins P, Akiyama Y, Hanna WW (2003) Molecular characterization of the genomic region linked with apomixis in Pennisetum/Cenchrus. Funct Integr Genomics **3:** 94-104

Ozias-Akins P, Lubbers EL, Hanna WW, McNay JW (1993) Transmission of the apomictic mode of reproduction in Pennisetum: co-inheritance of the trait and molecular markers. Theor Appl Genet **85:** 632-638

Ozias-Akins P, Roche D, Hanna WW (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. Proc Natl Acad Sci U S A **95:** 5127-5132

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

Passarinho P, Ketelaar T, Xing M, van Arkel J, Maliepaard C, Hendriks MW, Joosen R, Lammers M, Herdies L, den Boer B, van der Geest L, Boutilier K (2008) BABY BOOM target genes provide diverse entry points into cell proliferation and cell growth pathways. Plant Molecular Biology **68:** 225-237

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem **379:** 633-646

Roche D, Conner A, Budiman A, Frisch D, Wing R, Hanna W, Ozias-Akins P (2002) Construction of BAC libraries from two apomictic grasses to study the microcolinearity of their apospory-specific genomic regions. Theor Appl Genet **104:** 804-812

Sambrook J, Russell DW (2001) Molecular cloning, Ed 3 Vol 3. Cold Spring Harbor Laboratory Press, NY

Savidan Y (2000) Apomixis: genetics and breeding. Plant Breed Rev **18:** 13–85

Sherwood RT, Young BA, Bashaw EC (1980) Facultative pomixis in buffelgrass. Crop Science **20:** 375-379

Srinivasan C, Liu Z, Heidmann I, Supena ED, Fukuoka H, Joosen R, Lambalk J, Angenent G, Scorza R, Custers JB, Boutilier K (2007) Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (Nicotiana tabacum L.). Planta **225:** 341-351

Su YH, Zhao XY, Liu YB, Zhang CL, O'Neill SD, Zhang XS (2009) Auxin-induced WUS expression is essential for embryonic stem cell renewal during somatic embryogenesis in Arabidopsis. Plant J **59:** 448-460

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell **16:** S181-S189

van Dijk P, van Damme J (2000) Apomixis technology and the paradox of sex. Trends Plant Sci **5:** 81-84

Vielle JP, Burson BL, Bashaw EC, Hussey MA (1995) Early fertilization events in the sexual and aposporous egg apparatus of Pennisetum ciliare (L.) Link. Plant J **8:** 309-316

Visser NC, Spies JJ, Venter HJT (2000) Apomictic embryo sac development in Cenchrus ciliaris (Panicoideae). Bothalia **30:** 103-110

West M, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. Plant Cell **6:** 1731-1745

Young BA, Sherwood RT, Bashaw EC (1979) Cleared-pistil and thick-sectioning techniques for detecting aposporous apomixis in grasses. Can J Bot **57:** 1668-1672

Zuo J, Niu QW, Frugis G, Chua NH (2002) The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. Plant J **30:** 349-359

Table 4.1 BAC library screening result.
Hybridization of buffelgrass BAC library with 3 different probes containing 5'-end, AP2 domain and 3'-end sequences of *ASGR-BBM-like* identified 27 BAC clones, 6 of which were previously identified and confirmed to contain *ASGR-BBM-like*. We assumed that 21 other BAC clones potentially contained other *BBM-like* gene(s) located outside of ASGR, which were designated as *N-ASGR-BBM-like*. Note: 1 stands for positive and 0 stands for negative.

probes			calling address	BAC ID	notes
5'-end	AP2	3'-end			
1	1	1	C026A16	C100	*ASGR-BBM-like*
1	1	1	C096L01	C101	*ASGR-BBM-like*
1	1	1	C096N18	C102	*ASGR-BBM-like*
1	1	1	C127L23	C103	*ASGR-BBM-like*
1	1	1	C148O15	C104	*ASGR-BBM-like*
1	1	1	C167L21	C106	*ASGR-BBM-like*
1	1	1	C140D11		
1	1	1	C104G20		
1	1	1	C111M03		
1	1	1	C123G07		
1	1	1	C135I16		
1	1	1	C106K01		
1	1	1	C164L13	C1500	
1	1	1	C167O06		
1	1	1	C004A05		
1	1	1	C057D17		
1	1	1	C091M03	C1400	
1	1	1	C065L24		
0	1	1	C070G19		
0	1	1	C149G22		
0	1	0	C020K14		
0	1	0	C087F23		
0	1	0	C114E08		
0	1	0	C127E07		
0	1	0	C165H03		
0	1	0	C067O01		
0	1	0	C099L17		

Table 4.2 Similarity of genomic sequences (from predicted start codon to stop codon) among the *BBM-like* genes.

	CcASGR-BBM-like2	*CcASGR-BBM-like1*	*CcN-ASGR-BBM-like1400*	*CcN-ASGR-BBM-like1500*
CcASGR-BBM-like2	100	98	70	68
CcASGR-BBM-like1		100	70	69
CcN-ASGR-BBM-like1400			100	95
CcN-ASGR-BBM-like1500				100

Table 4.3 Similarity among the *BBM-LIKE* genes FGENESH predicted transcripts.

	CcASGR-BBM-like2	*CcASGR-BBM-like1*	*CcN-ASGR-BBM-like1400*	*CcN-ASGR-BBM-like1500*
CcASGR-BBM-like2	100	89	52	56
CcASGR-BBM-like1		100	60	64
CcN-ASGR-BBM-like1400			100	89
CcN-ASGR-BBM-like1500				100

Table 4.4 *BBM-like* genes expression level by the ratio to *GAPDH*. Values denote means and the standard errors are in parentheses.

	B-12-9		B-2s
	CcASGR-BBM-like	*CcN-ASGR-BBM-like*	*CcN-ASGR-BBM-like*
DBP	0.11 (0.02)	0.00 (0)	0.05 (0.01)
DOP	0.10 (0.03)	0.00 (0)	0.04 (0.01)
DAP	0.43 (0.02)	0.03 (0.01)	0.15 (0.02)

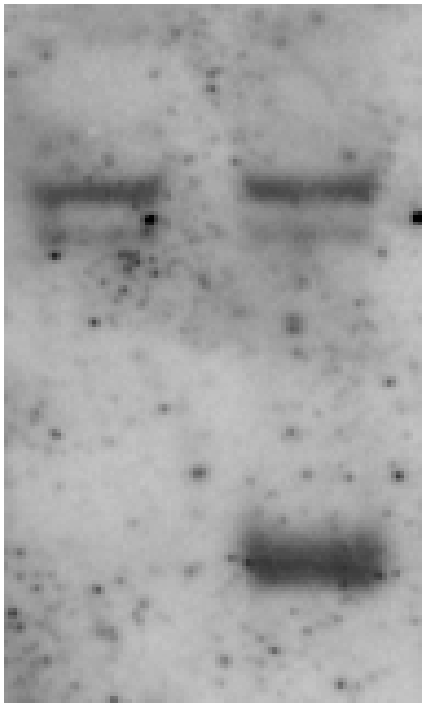


Figure 4.1 Southern blot analysis. With *Eco*RV digested genomic DNA and AP2 domain containing probe, a Southern blot showed that there were two common bands between B-2s and B-12-9 while there was one specific band only showing in B-12-9. Left lane is B-2s and right lane is B-12-9.

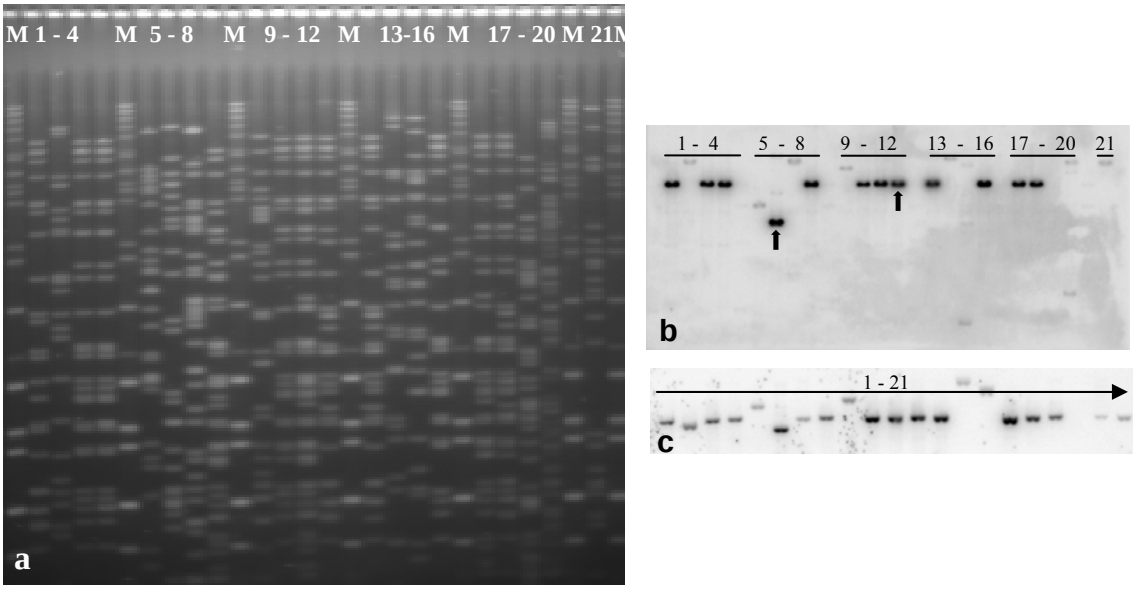
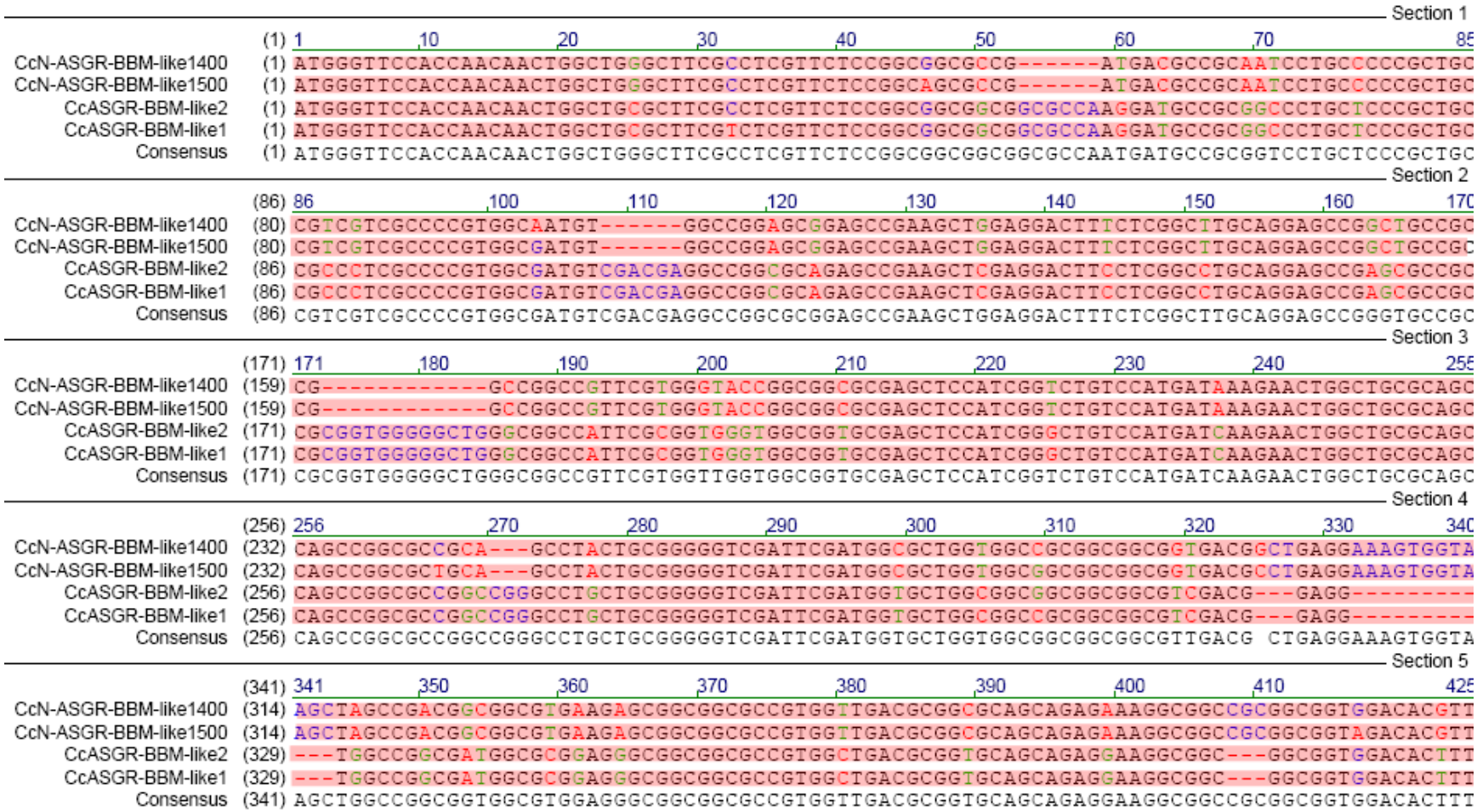
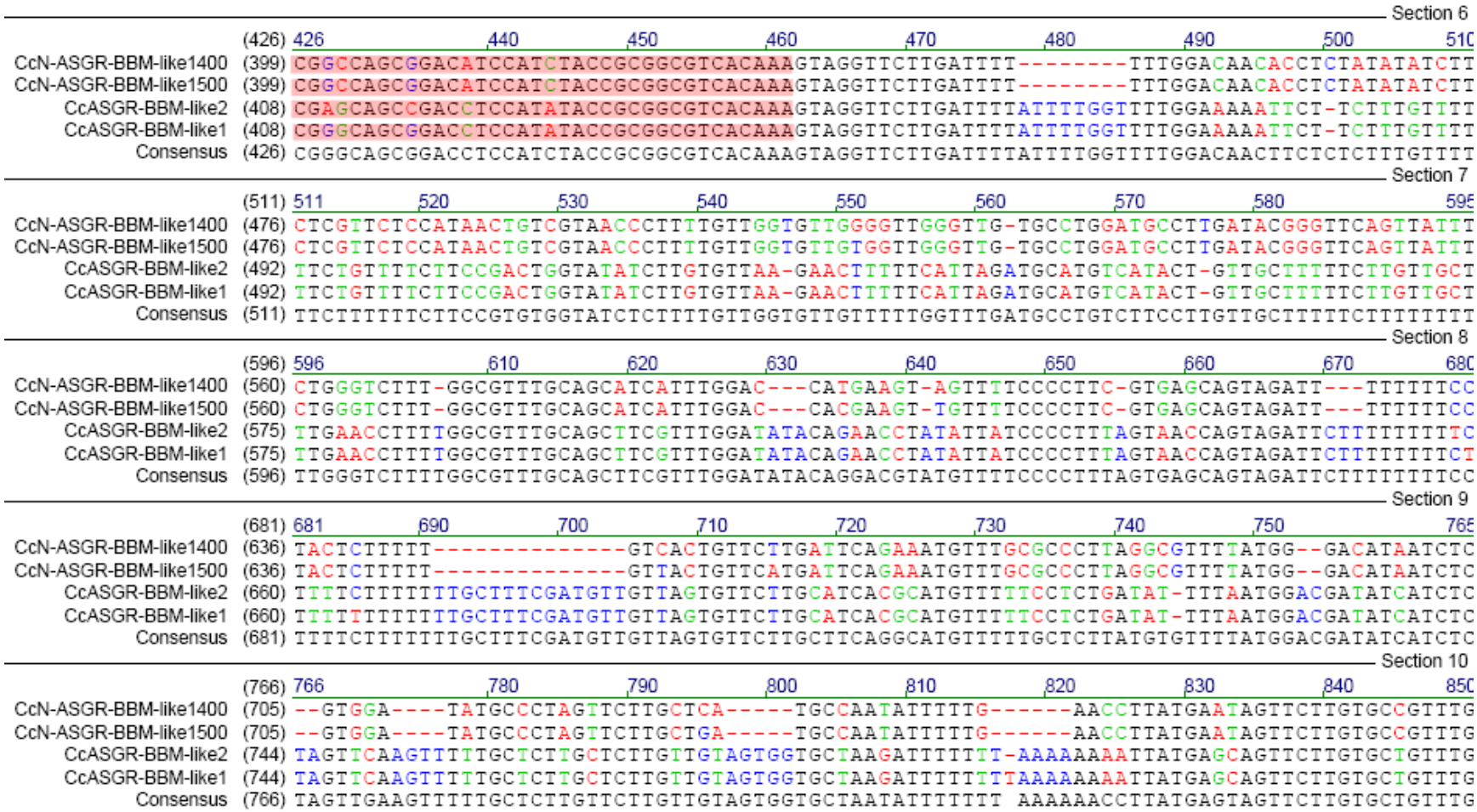


Figure 4.2 Fingerprinting analysis. 2a, *Hin*dIII-digested BAC fingerprinting image; 2b, *Hin*dIII-digested BAC fingerprinting gel hybridized to the mixture of the three probes; 2c, *Eco*RV-digested BAC clones hybridized to the probe containing the AP2 domain. M: DNA ladders, which is a mix of the Hi-Lo Marker, *Hin*dIII digested lambda DNA, and *Sal*I and *Stu*I digested lambda DNA; 1-4: C065L24, C165H03, C004A05, C106K01; 5-8: C087F23, C091M03, C127E07, C135I16; 9-12: C114E08, C140D11, C104G20, C164L13; 13-16: C167O06, C020K14, C067O01, C123G07; 17-20: C111M03, C057D17, C149G22, C099L17; 21: C070G19.

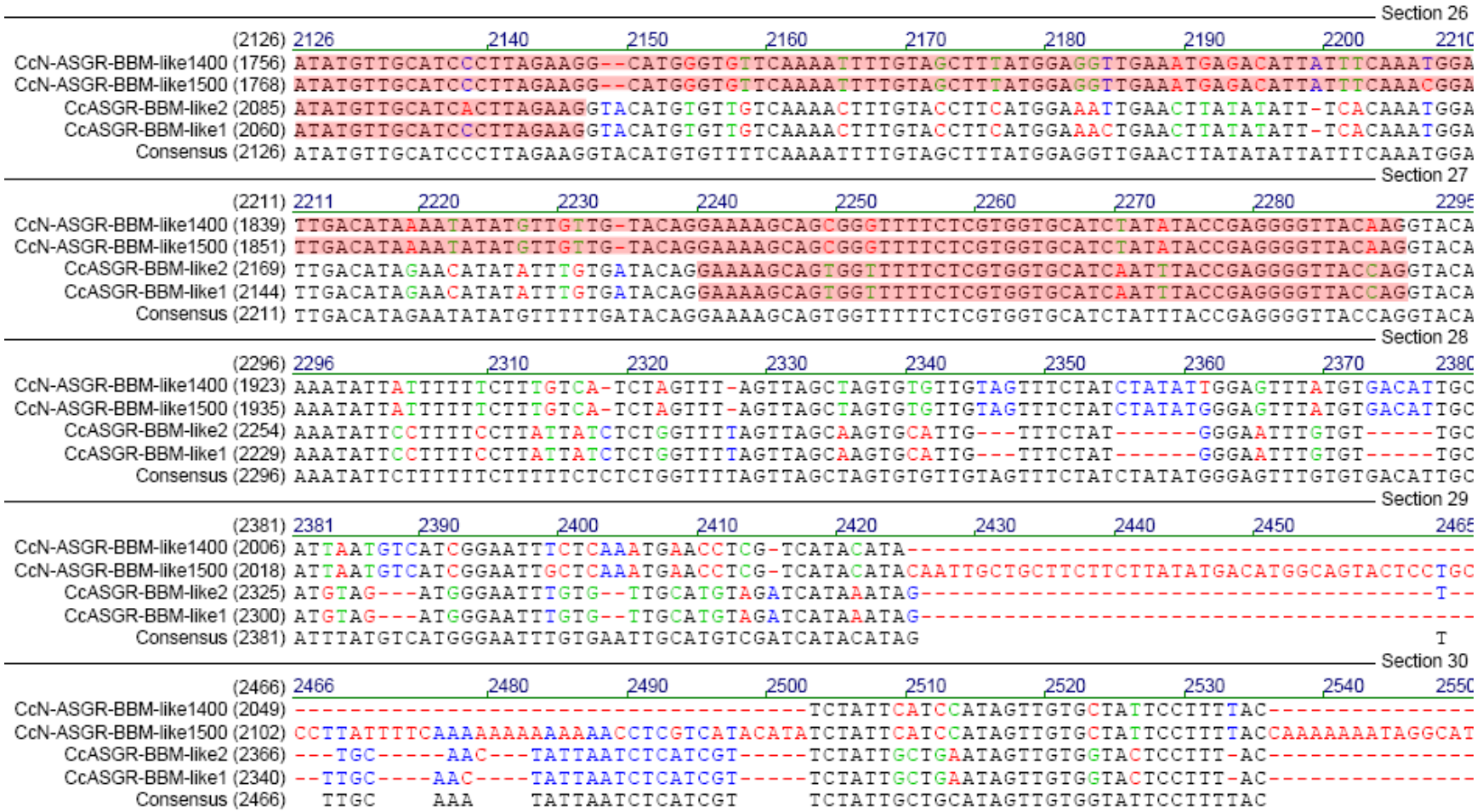

Section 1
(1) 1 10 20 30 40 50 60 70 85
CcN-ASGR-BBM-like1400 (1) ATGGGTTCCACCAACAACTGGCTGGGCTTCGCCTCGTTCTCCGGCGGCGCCG------ATGACGCCGCAATCCTGCCCCCGCTGC
CcN-ASGR-BBM-like1500 (1) ATGGGTTCCACCAACAACTGGCTGGGCTTCGCCTCGTTCTCCGGCAGCGCCG------ATGACGCCGCAATCCTGCCCCCGCTGC
CcASGR-BBM-like2 (1) ATGGGTTCCACCAACAACTGGCTGCGCTTCGCCTCGTTCTCCGGCGGCGGCGGCGCCAAGGATGCCGCGGCCCTGCTCCCGCTGC
CcASGR-BBM-like1 (1) ATGGGTTCCACCAACAACTGGCTGCGCTTCGTCTCGTTCTCCGGCGGCGGCGGCGCCAAGGATGCCGCGGCCCTGCTCCCGCTGC
Consensus (1) ATGGGTTCCACCAACAACTGGCTGGGCTTCGCCTCGTTCTCCGGCGGCGGCGGCGCCAATGATGCCGCGGTCCTGCTCCCGCTGC
Section 2
(86) 86 100 110 120 130 140 150 160 170
CcN-ASGR-BBM-like1400 (80) CGTCGTCGCCCCGTGGCAATGT------GGCCGGAGCGGAGCCGAAGCTGGAGGACTTTCTCGGCTTGCAGGAGCCGGCTGCCGC
CcN-ASGR-BBM-like1500 (80) CGTCGTCGCCCCGTGGCGATGT------GGCCGGAGCGGAGCCGAAGCTGGAGGACTTTCTCGGCTTGCAGGAGCCGGCTGCCGC
CcASGR-BBM-like2 (86) CGCCCTCGCCCCGTGGCGATGTCGACGAGGCCGGCGCAGAGCCGAAGCTCGAGGACTTCCTCGGCCTGCAGGAGCCGAGCGCCGC
CcASGR-BBM-like1 (86) CGCCCTCGCCCCGTGGCGATGTCGACGAGGCCGGCGCAGAGCCGAAGCTCGAGGACTTCCTCGGCCTGCAGGAGCCGAGCGCCGC
Consensus (86) CGTCGTCGCCCCGTGGCGATGTCGACGAGGCCGGCGCGGAGCCGAAGCTGGAGGACTTTCTCGGCTTGCAGGAGCCGGGTGCCGC
Section 3
(171) 171 180 190 200 210 220 230 240 255
CcN-ASGR-BBM-like1400 (159) CG----------GCCGGCCGTTCGTGGGTACCGGCGGCGCGAGCTCCATCGGTCTGTCCATGATAAAGAACTGGCTGCGCAGC
CcN-ASGR-BBM-like1500 (159) CG----------GCCGGCCGTTCGTGGGTACCGGCGGCGCGAGCTCCATCGGTCTGTCCATGATAAAGAACTGGCTGCGCAGC
CcASGR-BBM-like2 (171) CGCGGTGGGGGCTGGGCGGCCATTCGCGGTGGGTGGCGGTGCGAGCTCCATCGGGCTGTCCATGATCAAGAACTGGCTGCGCAGC
CcASGR-BBM-like1 (171) CGCGGTGGGGGCTGGGCGGCCATTCGCGGTGGGTGGCGGTGCGAGCTCCATCGGGCTGTCCATGATCAAGAACTGGCTGCGCAGC
Consensus (171) CGCGGTGGGGGCTGGGCGGCCGTTCGTGGTTGGTGGCGGTGCGAGCTCCATCGGTCTGTCCATGATCAAGAACTGGCTGCGCAGC
Section 4
(256) 256 270 280 290 300 310 320 330 340
CcN-ASGR-BBM-like1400 (232) CAGCCGGCGCCGCA---GCCTACTGCGGGGGTCGATTCGATGGCGCTGGTGGCCGCGGCGGCGGTGACGGCTGAGGAAAGTGGTA
CcN-ASGR-BBM-like1500 (232) CAGCCGGCGCTGCA---GCCTACTGCGGGGGTCGATTCGATGGCGCTGGTGGCGGCGGCGGCGGTGACGCCTGAGGAAAGTGGTA
CcASGR-BBM-like2 (256) CAGCCGGCGCCGGCCGGGCCTGCTGCGGGGGTCGATTCGATGGTGCTGGCGGCGGCGGCGGCGTCGACG---GAGG---------
CcASGR-BBM-like1 (256) CAGCCGGCGCCGGCCGGGCCTGCTGCGGGGGTCGATTCGATGGTGCTGGCGGCCGCGGCGGCGTCGACG---GAGG---------
Consensus (256) CAGCCGGCGCCGGCCGGGCCTGCTGCGGGGGTCGATTCGATGGTGCTGGTGGCGGCGGCGGCGTTGACG CTGAGGAAAGTGGTA
Section 5
(341) 341 350 360 370 380 390 400 410 425
CcN-ASGR-BBM-like1400 (314) AGCTAGCCGACGGCGGCGTGAAGAGCGGCGGCGCCGTGGTTGACGCGGCGCAGCAGAGAAAGGCGGCCGCGGCGGTGGACACGTT
CcN-ASGR-BBM-like1500 (314) AGCTAGCCGACGGCGGCGTGAAGAGCGGCGGCGCCGTGGTTGACGCGGCGCAGCAGAGAAAGGCGGCCGCGGCGGTAGACACGTT
CcASGR-BBM-like2 (329) ---TGGCCGGCGATGGCGCGGAGGGCGGCGGCGCCGTGGCTGACGCGGTGCAGCAGAGGAAGGCGGC---GGCGGTGGACACTTT
CcASGR-BBM-like1 (329) ---TGGCCGGCGATGGCGCGGAGGGCGGCGGCGCCGTGGCTGACGCGGTGCAGCAGAGGAAGGCGGC---GGCGGTGGACACTTT
Consensus (341) AGCTGGCCGGCGGTGGCGTGGAGGGCGGCGGCGCCGTGGTTGACGCGGTGCAGCAGAGGAAGGCGGCCGCGGCGGTGGACACTTT

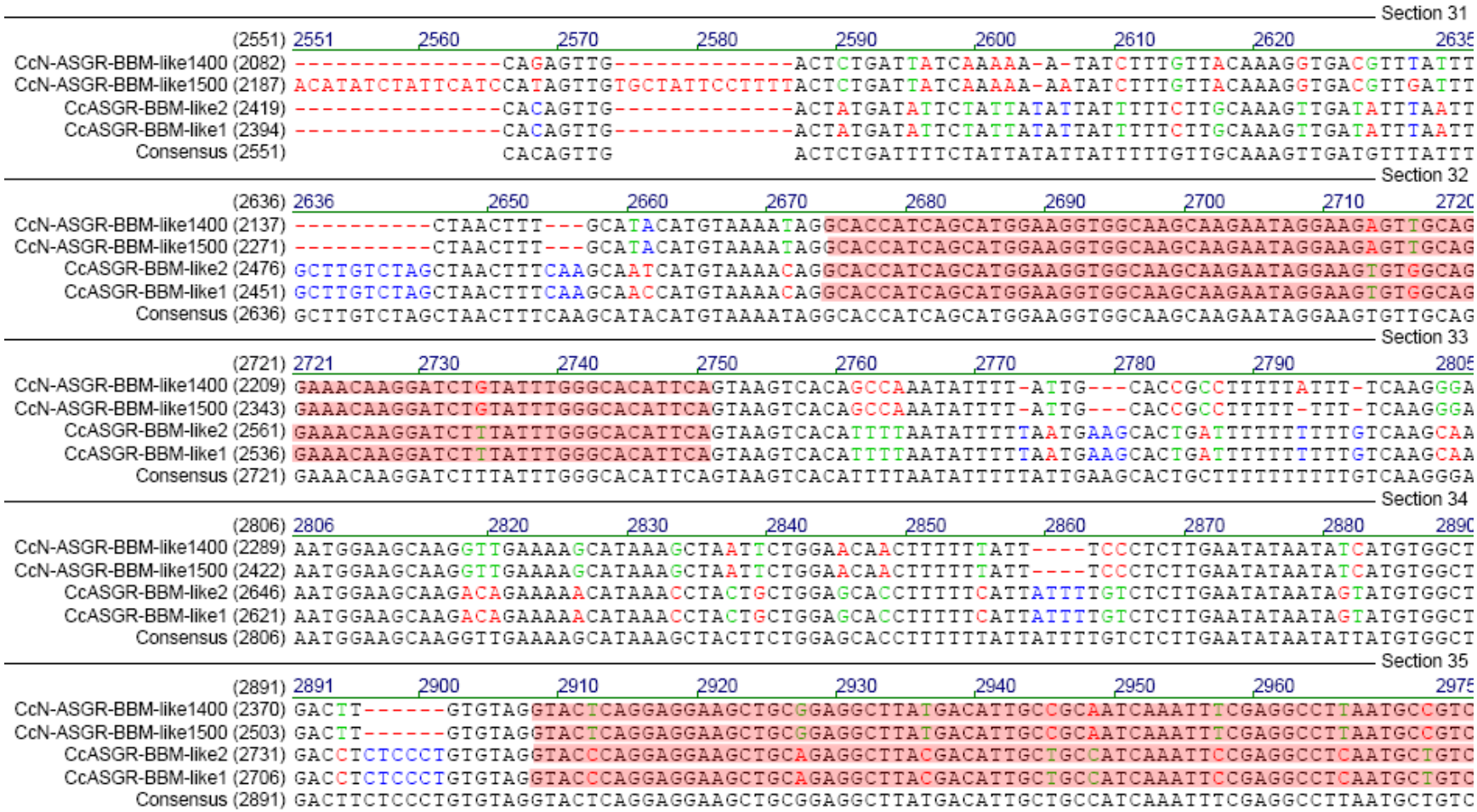

Section 6
(426) 426 440 450 460 470 480 490 500 510
CcN-ASGR-BBM-like1400 (399) CGGCCAGCGGACATCCATCTACCGCGGCGTCACAAAGTAGGTTCTTGATTTT--------TTTGGACAACACCTCTATATATCTT
CcN-ASGR-BBM-like1500 (399) CGGCCAGCGGACATCCATCTACCGCGGCGTCACAAAGTAGGTTCTTGATTTT--------TTTGGACAACACCTCTATATATCTT
CcASGR-BBM-like2 (408) CGAGCAGCCGACCTCCATATACCGCGGCGTCACAAAGTAGGTTCTTGATTTTATTTTGGTTTTGGAAAAATTCT-TCTTTGTTTT
CcASGR-BBM-like1 (408) CGGGCAGCGGACCTCCATATACCGCGGCGTCACAAAGTAGGTTCTTGATTTTATTTTGGTTTTGGAAAAATTCT-TCTTTGTTTT
Consensus (426) CGGGCAGCGGACCTCCATCTACCGCGGCGTCACAAAGTAGGTTCTTGATTTTATTTTGGTTTTGGACAACTTCTCTCTTTGTTTT
Section 7
(511) 511 520 530 540 550 560 570 580 595
CcN-ASGR-BBM-like1400 (476) CTCGTTCTCCATAACTGTCGTAACCCTTTTGTTGGTGTTGGGGTTGGGTTG-TGCCTGGATGCCTTGATACGGGTTCAGTTATTT
CcN-ASGR-BBM-like1500 (476) CTCGTTCTCCATAACTGTCGTAACCCTTTTGTTGGTGTTGTGGTTGGGTTG-TGCCTGGATGCCTTGATACGGGTTCAGTTATTT
CcASGR-BBM-like2 (492) TTCTGTTTTCTTCCGACTGGTATATCTTGTGTTAA-GAACTTTTTCATTAGATGCATGTCATACT-GTTGCTTTTTCTTGTTGCT
CcASGR-BBM-like1 (492) TTCTGTTTTCTTCCGACTGGTATATCTTGTGTTAA-GAACTTTTTCATTAGATGCATGTCATACT-GTTGCTTTTTCTTGTTGCT
Consensus (511) TTCTTTTTTCTTCCGTGTGGTATCTCTTTTGTTGGTGTTGTTTTTGGTTTGATGCCTGTCTTCCTTGTTGCTTTTTCTTTTTTT
Section 8
(596) 596 610 620 630 640 650 660 670 680
CcN-ASGR-BBM-like1400 (560) CTGGGTCTTT-GGCGTTTGCAGCATCATTTGGAC---CATGAAGT-AGTTTTCCCCTTC-GTGAGCAGTAGATT---TTTTTTCC
CcN-ASGR-BBM-like1500 (560) CTGGGTCTTT-GGCGTTTGCAGCATCATTTGGAC---CACGAAGT-TGTTTTCCCCTTC-GTGAGCAGTAGATT---TTTTTTCC
CcASGR-BBM-like2 (575) TTGAACCTTTTGGCGTTTGCAGCTTCGTTTGGATATACAGAACCTATATTATCCCCTTTAGTAACCAGTAGATTCTTTTTTTTTC
CcASGR-BBM-like1 (575) TTGAACCTTTTGGCGTTTGCAGCTTCGTTTGGATATACAGAACCTATATTATCCCCTTTAGTAACCAGTAGATTCTTTTTTTTCT
Consensus (596) TTGGGTCTTTTGGCGTTTGCAGCTTCGTTTGGATATACAGGACGTATGTTTTCCCCTTTAGTGAGCAGTAGATTCTTTTTTTTCC
Section 9
(681) 681 690 700 710 720 730 740 750 765
CcN-ASGR-BBM-like1400 (636) TACTCTTTTT--------------GTCACTGTTCTTGATTCAGAAATGTTTGCGCCCTTAGGCGTTTTATGG--GACATAATCTC
CcN-ASGR-BBM-like1500 (636) TACTCTTTTT--------------GTTACTGTTCATGATTCAGAAATGTTTGCGCCCTTAGGCGTTTTATGG--GACATAATCTC
CcASGR-BBM-like2 (660) TTTTCTTTTTTTGCTTTCGATGTTGTTAGTGTTCTTGCATCACGCATGTTTTTCCTCTGATAT-TTTAATGGACGATATCATCTC
CcASGR-BBM-like1 (660) TTTTTTTTTTTTGCTTTCGATGTTGTTAGTGTTCTTGCATCACGCATGTTTTTCCTCTGATAT-TTTAATGGACGATATCATCTC
Consensus (681) TTTTCTTTTTTTGCTTTCGATGTTGTTAGTGTTCTTGCTTCAGGCATGTTTTTGCTCTTATGTGTTTTATGGACGATATCATCTC
Section 10
(766) 766 780 790 800 810 820 830 840 850
CcN-ASGR-BBM-like1400 (705) --GTGGA----TATGCCCTAGTTCTTGCTCA-----TGCCAATATTTTTG------AACCTTATGAATAGTTCTTGTGCCGTTTG
CcN-ASGR-BBM-like1500 (705) --GTGGA----TATGCCCTAGTTCTTGCTGA-----TGCCAATATTTTTG------AACCTTATGAATAGTTCTTGTGCCGTTTG
CcASGR-BBM-like2 (744) TAGTTCAAGTTTTTGCTCTTGCTCTTGTTGTAGTGGTGCTAAGATTTTTTT-AAAAAAAATTATGAGCAGTTCTTGTGCTGTTTG
CcASGR-BBM-like1 (744) TAGTTCAAGTTTTTGCTCTTGCTCTTGTTGTAGTGGTGCTAAGATTTTTTTTAAAAAAAATTATGAGCAGTTCTTGTGCTGTTTG
Consensus (766) TAGTTGAAGTTTTTGCTCTTGTTCTTGTTGTAGTGGTGCTAATATTTTTTT AAAAAACCTTATGAGTAGTTCTTGTGCTGTTTG

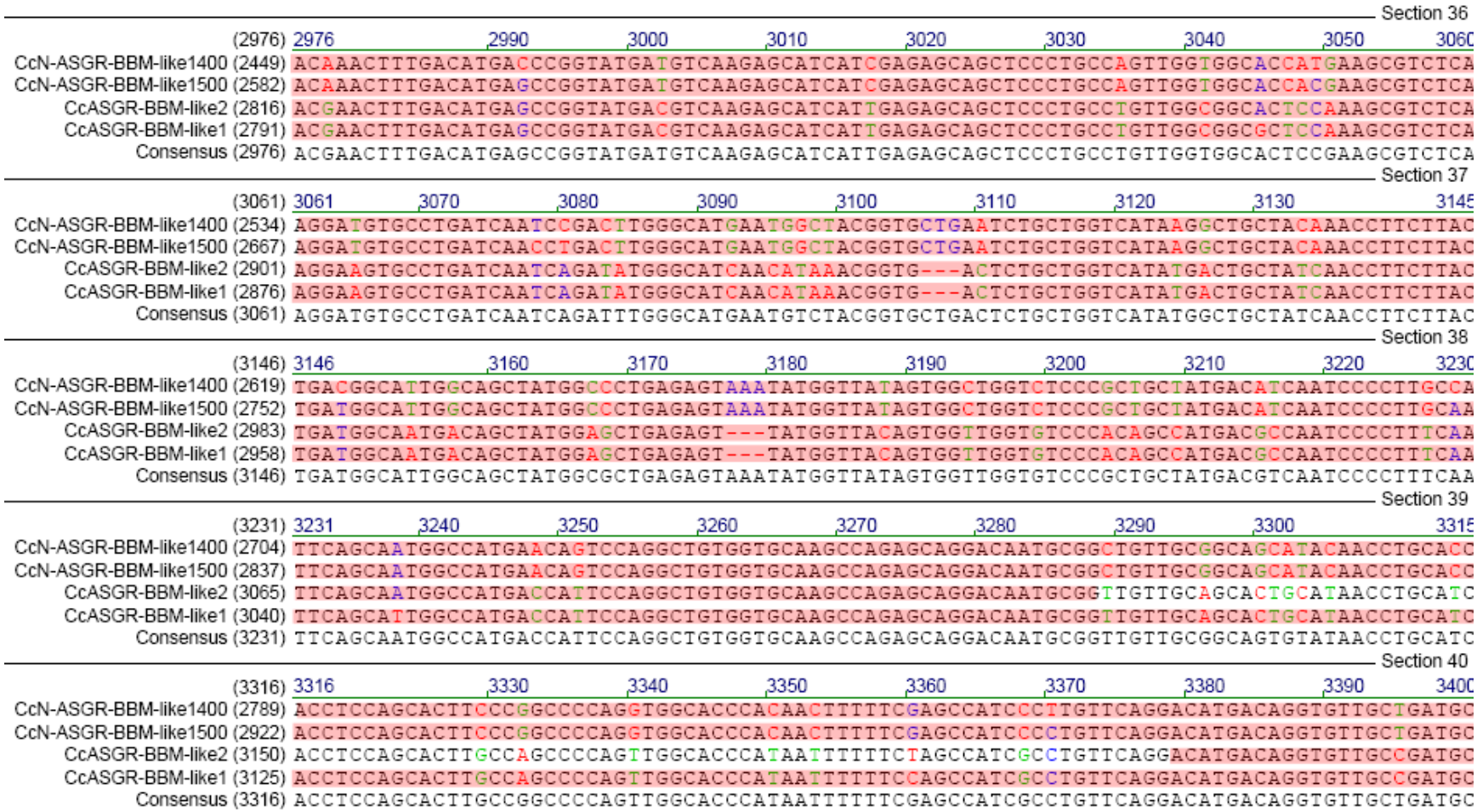

Section 11
(851) 851 860 870 880 890 900 910 920 935
CcN-ASGR-BBM-like1400 (773) AAAAATCAAACTCCTCACA----------
CcN-ASGR-BBM-like1500 (773) AAAAATCAAACTCCTCACA----------
CcASGR-BBM-like2 (828) AAAATGTAAGCACCTCACAGTTCTAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATATGTCT
CcASGR-BBM-like1 (829) AAAATGTAAGCATCTCACAGTTCT----------TT-----TATATATATATATATATATATATATATATATAAGTCT
Consensus (851) AAAATTTAAGCTCCTCACAGTTCT T TATATATATATATATATATATATATA GTCT
Section 12
(936) 936 950 960 970 980 990 1000 1010 1020
CcN-ASGR-BBM-like1400 (792) ----TGT----------AAGCTAGGC---------TGTCCAC----------TGCTC
CcN-ASGR-BBM-like1500 (792) ----TCTTGATGCACGGG------GAAGCTAGGC---------TGTCCAC----------TGCTC
CcASGR-BBM-like2 (913) CTCATGTTGATTTGTGGATGTACTGAAGCCCCGCGCGCACACATGCACACACCGCACGCTCACACGCCCTAAATCCCCGGTGCAA
CcASGR-BBM-like1 (888) CTCATGTTCATTTGTGGATGTACTGAAGCCCCGCGCGCACACATGCACACACCGCACGCTCACACGCCCTAAATCCCCGGTGCAA
Consensus (936) CTCATGTTGATTTGTGGATGTACTGAAGCTCGGCGCGCACACATGTCCACACCGCACGCTCACACGCCCTAAATCCCCGGTGCTC
Section 13
(1021) 1021 1030 1040 1050 1060 1070 1080 1090 1105
CcN-ASGR-BBM-like1400 (816) CTCCA---TT----------TGATTT----------CTGTCA----------TATA-------TT
CcN-ASGR-BBM-like1500 (828) CTCCA---TT----------TGATTT----------CTGTCA----------TATA-------TT
CcASGR-BBM-like2 (998) CACCAGGGTTGTCCCCGATGGTGATCGAACCCTGGCGGGTGGCCTAACCACCGTCAGCTCCCACCACGGAGCTATCAGCTCGTTT
CcASGR-BBM-like1 (973) CACCAGGGTTGTCCCCGATGGGGATCGAACCCTGGCGGGTGGCCTAACCACCGTCAGCTCCCACCACCGAGCTATCAGCTCGTTT
Consensus (1021) CTCCAGGGTTGTCCCCGATGGTGATTTAACCCTGGCGGGTGGCCTAACCACTGTCAGCTCCCACCAC GAGCTATCAGCTCGTTT
Section 14
(1106) 1106 1120 1130 1140 1150 1160 1170 1180 1190
CcN-ASGR-BBM-like1400 (841) GCGAAT-------GTAATTTCAGAATATTTTTATAAG-CTGGATTGTTCCATCTATCATC----------C
CcN-ASGR-BBM-like1500 (853) GCGAAT-------GTAATTCCAGAATATTTTTATAAG-CTGGATTGTTCCATCTATCATC----------C
CcASGR-BBM-like2 (1083) GCCCATATTTCGTGTGGTACCTCGATATTTTTATATTTCTAGATTGCTGTATCTATCTTCTAGACTTATATAAGTGTTGCGCCAC
CcASGR-BBM-like1 (1058) GCCCATATTTCGTGTGGTACCTAGATATTTTTATATTTCTAGATTGCTGTATCTATCTTCTAGACTTATATAAGTGTTGCGCCAC
Consensus (1106) GCGCATATTTCGTGTGGTTCCTGGATATTTTTATATTTCTGGATTGTTGTATCTATCTTCTAGACTTATATAAGTGTTGCGCCAC
Section 15
(1191) 1191 1200 1210 1220 1230 1240 1250 1260 1275
CcN-ASGR-BBM-like1400 (894) TC-TAGTTTTTGCCACCTGTAATAATCGAGTAGAACCACCTCTTCCTCTTTCATAGTATTGTATCAACTAAATCAT--TCGTGAA
CcN-ASGR-BBM-like1500 (906) TC-TAGTTTTTGCCACCTGTAATAATCGAGTAGAACCACCTCTTCCTCTTTCATAGTATTGTATCAATTAAATCAT--TCGTGAA
CcASGR-BBM-like2 (1168) TCATACTTTTTACCGCCTGTAAT---CGAGTAGAACTGCTTCCTCT--TTTGATTATATTGTATCAGTTAAATGATCTTGTTGTT
CcASGR-BBM-like1 (1143) TCATACTTTTTACCGCCTGTAAT---CGAGTAGAACTGCTTCCTCT--TTTGATTATATTGTATCAGTTAAATGATCTTGTTGTT
Consensus (1191) TCATAGTTTTTGCCGCCTGTAATAATCGAGTAGAACTGCTTCTTCTTCTTTGATTGTATTGTATCAGTTAAATGATCTTGTTGTT


Section 16
(1276) 1276 1290 1300 1310 1320 1330 1340 1350 1360
CcN-ASGR-BBM-like1400 (976) AATGTGTTTATCA--------TCATCATTGCATGAAGTCA----------CTAACTGGCTAAATTTACAAGTG
CcN-ASGR-BBM-like1500 (988) AATGTGTTTATCA--------TCATCATTGCATGAAGTCA----------TTGACTGGCTAAATTTACAAGTG
CcASGR-BBM-like2 (1248) GATGTGTTTACCACTTTACCATCACCATTGCATGAAATCACTTCAAGACATGTATTCATGATTTGGCTGGCTAAATTTGCTAGTG
CcASGR-BBM-like1 (1223) GATGTGTTTACCACTTTACCATCACCATTGCATGAAATCACTTCAAGACATGTATTCATGATTTGGCTGGCTAAATTTGCTAGTG
Consensus (1276) GATGTGTTTATCACTTTACCATCATCATTGCATGAAGTCACTTCAAGACATGTATTCATGATTTGGCTGGCTAAATTTGCTAGTG
Section 17
(1361) 1361 1370 1380 1390 1400 1410 1420 1430 1445
CcN-ASGR-BBM-like1400 (1031) CCTGAT------GGTAAA----TCTTTGTCATTTGTGCTTCTTGTTATTCTTTCCTATCTCACCCCTTTGGCCTTATTCCTTAAT
CcN-ASGR-BBM-like1500 (1043) GCTGAT------GGTAAA----TCTTTGTCATTTGTGCTTCTTGTTATTCTTTCCTATCTCACCCCTTTGGCCTTATTCCTTAAT
CcASGR-BBM-like2 (1333) GCACATACATGTGGTAAAAAAATATTTTTAGTTTGTGCT---TGCTATTCTTTTCGGTC--ATCCCTTCGTGCTTGT----TTAT
CcASGR-BBM-like1 (1308) GCACATACATGTGGTAAAAAAATATTTTTAGTTTGTGCT---TGCTATTCTTTTCGGTC--ATCCCTTCGTGCTTGT----TTAT
Consensus (1361) GCTGATACATGTGGTAAAAAAATCTTTTTCGTTTGTGCTTCTTGTTATTCTTTTCTGTCTCATCCCTTTGTGCTTGTTCCTTTAT
Section 18
(1446) 1446 1460 1470 1480 1490 1500 1510 1520 1530
CcN-ASGR-BBM-like1400 (1106) CCAGAACACCCAATCTGCTTCACATTTATGTCTAATGTTGTCGTCATCTTTCATGTTTGCAGATATTTGTACTAAACTATTGGTT
CcN-ASGR-BBM-like1500 (1118) CCAGAACACCCAATCTGCTTCACATTTATGTCTAATGTTGTCGTCATCTTTCATGTTTGCAGATATTTGTACTAAACTATTGGCT
CcASGR-BBM-like2 (1409) CCAGAACACCCAATCTGCTTCACATAGTTTTTGAATGCTATCATCATATTTCTTTTTTGGAGATATTGTTACTAAAAGTTTGGCT
CcASGR-BBM-like1 (1384) CCAGAACACCCAATCTGCTTCACATAGTTTTTGAATGCTATCATCATATTTCTTTTTTGGAGATATTGTTACTAAAAGTTTGGCT
Consensus (1446) CCAGAACACCCAATCTGCTTCACATTTTTTTTTAATGTTGTCGTCATCTTTCTTTTTTGGAGATATTTTTACTAAACTTTTGGCT
Section 19
(1531) 1531 1540 1550 1560 1570 1580 1590 1600 1615
CcN-ASGR-BBM-like1400 (1191) TTGTTCCCAATAGGCATCGATGGACAGGAAGATACGAAGCACATCTTTGGGACAATAGCTGCAGAAGGGAAGGTCAGACTCGCAA
CcN-ASGR-BBM-like1500 (1203) TTGTTCTCAATAGGCATCGATGGACAGGAAGATGCGAAGCACATCTTTGGGACAATAGCTGCAGAAGGGAAGGTCAGACTCGCAA
CcASGR-BBM-like2 (1494) TTGTCCTCAATAGGCATAGATTGACAGGAAGGTATGAAGCCCATCTTTGGGACAATAGCTGCAGAAGAGAAGGTCAGACTCGGAA
CcASGR-BBM-like1 (1469) TTGTCCTCAATAGGCATAGATGGACAGGAAGGTATGAAGCCCATCTTTGGGACAATAGCTGCAGAAGAGAAGGTCAAACTCGGAA
Consensus (1531) TTGTTCTCAATAGGCATCGATGGACAGGAAGGTATGAAGCCCATCTTTGGGACAATAGCTGCAGAAGGGAAGGTCAGACTCGGAA
Section 20
(1616) 1616 1630 1640 1650 1660 1670 1680 1690 1700
CcN-ASGR-BBM-like1400 (1276) AGGAAGACAAGGTAATCATTATAATAGAAATATTTA----TGCATTTCTA---TGTAGCATTTTATTATTGAATGGG--GGTTTT
CcN-ASGR-BBM-like1500 (1288) AGGAAGACAAGGTAATCATTATAATAGAAACATTTA----TGCATTTCTA---TGTAGCATTTTATTATTGAATGGG--GGCTTT
CcASGR-BBM-like2 (1579) AGGTAGACAAGGTAATGATTATAATATAGATATTTAAATTTGTAATTATAAGCTGCATCATATTATTATTTATTAGATCGGCTTT
CcASGR-BBM-like1 (1554) AGGTAGACAAGGTAATGATTATAATATAGATATTTAAATTTGTAATTATAAGCTGCATCATATTATTATTTATTAGATCGGCTTT
Consensus (1616) AGGTAGACAAGGTAATGATTATAATATAGATATTTAAATTTGTATTTCTAAGCTGTATCATTTTATTATTTATTGGGTCGGCTTT


Section 21
(1701) 1701 1710 1720 1730 1740 1750 1760 1770 1785
CcN-ASGR-BBM-like1400 (1352) GCAATTTTCTGTGCTAACCAAATGTTTTGTTTCTGT-----GGTACCTGCAATCCCTTGACTCCATTAAATCAGTGTATCTTGGT
CcN-ASGR-BBM-like1500 (1364) GCAATTTTCTGTGCTAACCAAATGTTTTGTTTCTGT-----GGTACCTGCAATCCCTTGACTCCATTAATTCAGTGTATCTTGGT
CcASGR-BBM-like2 (1664) AAAATTTCACTAGCTAATTTAGTGTTTTTCTTTTCTTCATCGATACCTGCAATCGCTTCATTCCATTGATTCAGTGTATCTTGGT
CcASGR-BBM-like1 (1639) AAAATTTCACTAGCTAATTTAGTGTTTTTCTTTTCTTCATCGATACCTGCAATCGCTTCATTCCATTGATTCAGTGTATCTTGGT
Consensus (1701) GCAATTTTCTTTGCTAATTTAGTGTTTTTTTTTTGTTCATCGGTACCTGCAATCGCTTGATTCCATTGATTCAGTGTATCTTGGT
Section 22
(1786) 1786 1800 1810 1820 1830 1840 1850 1860 1870
CcN-ASGR-BBM-like1400 (1432) AAGTACTAATAA---ACAAT-----ATTGGTTTACAATTACTATTAATTTTACATCTAATTTATATGTAGC--GTTGTTTATGCA
CcN-ASGR-BBM-like1500 (1444) AAGTACTAATAA---ACAAT-----ATTGGTTTACAATTACTATTAATTTTACATCTAATTTTCATGTAGC--GTTGTTTATGCA
CcASGR-BBM-like2 (1749) AAGTAATACTTGTTTACAATTGCAAAATGGTAT---ATCTCTTGTTGTTTCTCATGTCAAATATATTAAATATGTGGTTGATGCA
CcASGR-BBM-like1 (1724) AAGTAATACTTGTTTACAATTGCAAAATGGTAT---ATCTCTTGTTGTTTCTCATGTCAAGTATATTAAATATGTGGTTGATGCA
Consensus (1786) AAGTACTACTTGTTTACAATTGCAAATTGGTTTACAATTTCTTTTTGTTTTTCATGTCATTTATATTTAGTATGTTGTTTATGCA
Section 23
(1871) 1871 1880 1890 1900 1910 1920 1930 1940 1955
CcN-ASGR-BBM-like1400 (1507) TTGAAGGTGGATATGATAAAGAAGAGAAAGCGGCCAGAGCTTATGATCTTGCTGCTCTCAAGTACTGGGGCGCTACAACAACTAC
CcN-ASGR-BBM-like1500 (1519) TTGAAGGTGGATATGATAAAGAAGAGAAAGCGGCCAGAGCTTATGATCTTGCTGCTCTCAAGTACTGGGGCGCTACAACAACTAC
CcASGR-BBM-like2 (1831) TTGAAGGTGGATATGATAAAGAAGAAAAAGCAGCTAGAGCTTAGGATTTAGCTGCTCTCAAGTACCGGGGCACCACAACTACTAC
CcASGR-BBM-like1 (1806) TTGAAGGTGGATATGATAAAGAAGAAAAAGCAGCTAGAGCTTATGATTTAGCTGCTCTCAAGTACCGGGGCACCACAACTACTAC
Consensus (1871) TTGAAGGTGGATATGATAAAGAAGAGAAAGCGGCTAGAGCTTATGATTTTGCTGCTCTCAAGTACTGGGGCGCTACAACTACTAC
Section 24
(1956) 1956 1970 1980 1990 2000 2010 2020 2030 2040
CcN-ASGR-BBM-like1400 (1592) AAATTTTCCGGTACTAAATATTTTTTCTTTGTA--TTCTACA-ATCTGTTTATTTATTAC--ATTT-TCACTCCATATCATGTAT
CcN-ASGR-BBM-like1500 (1604) AAATTTTCCGGTACTAAATATTTTTTCTTTGTA--TTCTACA-ATCTGTTTATTTATTAC--ATTT-TCACTCCATATCATGTAT
CcASGR-BBM-like2 (1916) AAATTTTCCGGTATTACTTATTGTTAATATGTTGGTTCTCCAGAATTGATATTTTACTTCTAATATATAACTGCGTAT-ATGAAT
CcASGR-BBM-like1 (1891) AAATTTTCCGGTATTACTTATTGTTAATATGTTGGTTCTCCAGAATTAATATTTTACTTCTAATATATAACTGCGTAT-ATGAAT
Consensus (1956) AAATTTTCCGGTATTACTTATTTTTTCTTTGTTGGTTCTCCAGATTTGTTTTTTTATTTCTAATTTATCACTGCGTATCATGTAT
Section 25
(2041) 2041 2050 2060 2070 2080 2090 2100 2110 2125
CcN-ASGR-BBM-like1400 (1671) GAGTTTTATAAGATTTTTACATCAATGTTCAGATGGGCAACTATGAAAAAGAGTTGGAAGAGATGAAGCATATGTCCCGACAGGA
CcN-ASGR-BBM-like1500 (1683) GAGTTTTATAAGATTTTTAAATCAATGTTCAGATGGGCAACTACGAAAAAGAGTTGGAAGAGATGAAGCATATGTCCCGACAGGA
CcASGR-BBM-like2 (2000) GAATGTTGTAAGATTTTGCATTTTATGTTCAGATGAGCAACTATGAAAAGGAGTTAGAAGAGATGAAGCATATGTCACGACAAGA
CcASGR-BBM-like1 (1975) GAATGTTGTAAGATTTTGCATTTTATGTTCAGATGAGCAACTATGAAAAGGAGTTAGAAGAGATGAAGCATATGTCACGACAAGA
Consensus (2041) GAGTTTTGTAAGATTTTTCATTTTATGTTCAGATGGGCAACTATGAAAAGGAGTTGGAAGAGATGAAGCATATGTCCCGACAGGA


Section 26
(2126) 2126 2140 2150 2160 2170 2180 2190 2200 2210
CcN-ASGR-BBM-like1400 (1756) ATATGTTGCATCCCTTAGAAGG--CATGGGTGTTCAAAATTTTGTAGCTTTATGGAGGTTGAAATGAGACATTATTTCAAATGGA
CcN-ASGR-BBM-like1500 (1768) ATATGTTGCATCCCTTAGAAGG--CATGGGTGTTCAAAATTTTGTAGCTTTATGGAGGTTGAAATGAGACATTATTTCAAACGGA
CcASGR-BBM-like2 (2085) ATATGTTGCATCACTTAGAAGGTACATGTGTTGTCAAAACTTTGTACCTTCATGGAAATTGAACTTATATATT-TCACAAATGGA
CcASGR-BBM-like1 (2060) ATATGTTGCATCCCTTAGAAGGTACATGTGTTGTCAAAACTTTGTACCTTCATGGAAACTGAACTTATATATT-TCACAAATGGA
Consensus (2126) ATATGTTGCATCCCTTAGAAGGTACATGTGTTTTCAAAATTTTGTAGCTTTATGGAGGTTGAACTTATATATTATTTCAAATGGA
Section 27
(2211) 2211 2220 2230 2240 2250 2260 2270 2280 2295
CcN-ASGR-BBM-like1400 (1839) TTGACATAAAATATATGTTGTTG-TACAGGAAAAGCAGCGGGTTTTCTCGTGGTGCATCTATATACCGAGGGGTTACAAGGTACA
CcN-ASGR-BBM-like1500 (1851) TTGACATAAAATATATGTTGTTG-TACAGGAAAAGCAGCGGGTTTTCTCGTGGTGCATCTATATACCGAGGGGTTACAAGGTACA
CcASGR-BBM-like2 (2169) TTGACATAGAACATATATTTGTGATACAGGAAAAGCAGTGGTTTTTCTCGTGGTGCATCAATTTACCGAGGGGTTACCAGGTACA
CcASGR-BBM-like1 (2144) TTGACATAGAACATATATTTGTGATACAGGAAAAGCAGTGGTTTTTCTCGTGGTGCATCAATTTACCGAGGGGTTACCAGGTACA
Consensus (2211) TTGACATAGAATATATGTTTTTGATACAGGAAAAGCAGTGGTTTTTCTCGTGGTGCATCTATTTACCGAGGGGTTACCAGGTACA
Section 28
(2296) 2296 2310 2320 2330 2340 2350 2360 2370 2380
CcN-ASGR-BBM-like1400 (1923) AAATATTATTTTTTCTTTGTCA-TCTAGTTT-AGTTAGCTAGTGTGTTGTAGTTTCTATCTATATTGGAGTTTATGTGACATTGC
CcN-ASGR-BBM-like1500 (1935) AAATATTATTTTTTCTTTGTCA-TCTAGTTT-AGTTAGCTAGTGTGTTGTAGTTTCTATCTATATGGGAGTTTATGTGACATTGC
CcASGR-BBM-like2 (2254) AAATATTCCTTTTCCTTATTATCTCTGGTTTTAGTTAGCAAGTGCATTG---TTTCTAT------GGGAATTTGTGT-----TGC
CcASGR-BBM-like1 (2229) AAATATTCCTTTTCCTTATTATCTCTGGTTTTAGTTAGCAAGTGCATTG---TTTCTAT------GGGAATTTGTGT-----TGC
Consensus (2296) AAATATTCTTTTTTCTTTTTCTCTCTGGTTTTAGTTAGCTAGTGTGTTGTAGTTTCTATCTATATGGGAGTTTGTGTGACATTGC
Section 29
(2381) 2381 2390 2400 2410 2420 2430 2440 2450 2465
CcN-ASGR-BBM-like1400 (2006) ATTAATGTCATCGGAATTTCTCAAATGAACCTCG-TCATACATA----------
CcN-ASGR-BBM-like1500 (2018) ATTAATGTCATCGGAATTGCTCAAATGAACCTCG-TCATACATACAATTGCTGCTTCTTCTTATATGACATGGCAGTACTCCTGC
CcASGR-BBM-like2 (2325) ATGTAG---ATGGGAATTTGTG--TTGCATGTAGATCATAAATAG----------T--
CcASGR-BBM-like1 (2300) ATGTAG---ATGGGAATTTGTG--TTGCATGTAGATCATAAATAG----------
Consensus (2381) ATTTATGTCATGGGAATTTGTGAATTGCATGTCGATCATACATAG T
Section 30
(2466) 2466 2480 2490 2500 2510 2520 2530 2540 2550
CcN-ASGR-BBM-like1400 (2049) ----------TCTATTCATCCATAGTTGTGCTATTCCTTTTAC----------
CcN-ASGR-BBM-like1500 (2102) CCTTATTTTCAAAAAAAAAAAACCTCGTCATACATATCTATTCATCCATAGTTGTGCTATTCCTTTTACCAAAAAAATAGGCAT
CcASGR-BBM-like2 (2366) ---TGC-----AAC---TATTAATCTCATCGT-----TCTATTGCTGAATAGTTGTGGTACTCCTTT-AC----------
CcASGR-BBM-like1 (2340) --TTGC----AAC----TATTAATCTCATCGT-----TCTATTGCTGAATAGTTGTGGTACTCCTTT-AC----------
Consensus (2466) TTGC AAA TATTAATCTCATCGT TCTATTGCTGCATAGTTGTGGTATTCCTTTTAC


Section 31
(2551) 2551 2560 2570 2580 2590 2600 2610 2620 2635
CcN-ASGR-BBM-like1400 (2082) ----------CAGAGTTG----------ACTCTGATTATCAAAAA-A-TATCTTTGTTACAAAGGTGACGTTTATTT
CcN-ASGR-BBM-like1500 (2187) ACATATCTATTCATCCATAGTTGTGCTATTCCTTTTACTCTGATTATCAAAAA-AATATCTTTGTTACAAAGGTGACGTTGATTT
CcASGR-BBM-like2 (2419) ----------CACAGTTG----------ACTATGATATTCTATTATATTATTTTTCTTGCAAAGTTGATATTTAATT
CcASGR-BBM-like1 (2394) ----------CACAGTTG----------ACTATGATATTCTATTATATTATTTTTCTTGCAAAGTTGATATTTAATT
Consensus (2551) CACAGTTG ACTCTGATTTTCTATTATATTATTTTTGTTGCAAAGTTGATGTTTATTT
Section 32
(2636) 2636 2650 2660 2670 2680 2690 2700 2710 2720
CcN-ASGR-BBM-like1400 (2137) ----------CTAACTTT---GCATACATGTAAAATAGGCACCATCAGCATGGAAGGTGGCAAGCAAGAATAGGAAGAGTTGCAG
CcN-ASGR-BBM-like1500 (2271) ----------CTAACTTT---GCATACATGTAAAATAGGCACCATCAGCATGGAAGGTGGCAAGCAAGAATAGGAAGAGTTGCAG
CcASGR-BBM-like2 (2476) GCTTGTCTAGCTAACTTTCAAGCAATCATGTAAAACAGGCACCATCAGCATGGAAGGTGGCAAGCAAGAATAGGAAGTGTGGCAG
CcASGR-BBM-like1 (2451) GCTTGTCTAGCTAACTTTCAAGCAACCATGTAAAACAGGCACCATCAGCATGGAAGGTGGCAAGCAAGAATAGGAAGTGTGGCAG
Consensus (2636) GCTTGTCTAGCTAACTTTCAAGCATACATGTAAAATAGGCACCATCAGCATGGAAGGTGGCAAGCAAGAATAGGAAGTGTTGCAG
Section 33
(2721) 2721 2730 2740 2750 2760 2770 2780 2790 2805
CcN-ASGR-BBM-like1400 (2209) GAAACAAGGATCTGTATTTGGGCACATTCAGTAAGTCACAGCCAAATATTTT-ATTG---CACCGCCTTTTTATTT-TCAAGGGA
CcN-ASGR-BBM-like1500 (2343) GAAACAAGGATCTGTATTTGGGCACATTCAGTAAGTCACAGCCAAATATTTT-ATTG---CACCGCCTTTTT-TTT-TCAAGGGA
CcASGR-BBM-like2 (2561) GAAACAAGGATCTTTATTTGGGCACATTCAGTAAGTCACATTTTAATATTTTTAATGAAGCACTGATTTTTTTTTTGTCAAGCAA
CcASGR-BBM-like1 (2536) GAAACAAGGATCTTTATTTGGGCACATTCAGTAAGTCACATTTTAATATTTTTAATGAAGCACTGATTTTTTTTTTGTCAAGCAA
Consensus (2721) GAAACAAGGATCTTTATTTGGGCACATTCAGTAAGTCACATTTTAATATTTTTATTGAAGCACTGCTTTTTTTTTTGTCAAGGGA
Section 34
(2806) 2806 2820 2830 2840 2850 2860 2870 2880 2890
CcN-ASGR-BBM-like1400 (2289) AATGGAAGCAAGGTTGAAAAGCATAAAGCTAATTCTGGAACAACTTTTTTATT----TCCCTCTTGAATATAATATCATGTGGCT
CcN-ASGR-BBM-like1500 (2422) AATGGAAGCAAGGTTGAAAAGCATAAAGCTAATTCTGGAACAACTTTTTTATT----TCCCTCTTGAATATAATATCATGTGGCT
CcASGR-BBM-like2 (2646) AATGGAAGCAAGACAGAAAAACATAAACCTACTGCTGGAGCACCTTTTTTCATTATTTTGTCTCTTGAATATAATAGTATGTGGCT
CcASGR-BBM-like1 (2621) AATGGAAGCAAGACAGAAAAACATAAACCTACTGCTGGAGCACCTTTTTTCATTATTTTGTCTCTTGAATATAATAGTATGTGGCT
Consensus (2806) AATGGAAGCAAGGTTGAAAAGCATAAAGCTACTTCTGGAGCACCTTTTTTTATTATTTTGTCTCTTGAATATAATATTATGTGGCT
Section 35
(2891) 2891 2900 2910 2920 2930 2940 2950 2960 2975
CcN-ASGR-BBM-like1400 (2370) GACTT------GTGTAGGTACTCAGGAGGAAGCTGCGGAGGCTTATGACATTGCCGCAATCAAATTTCGAGGCCTTAATGCCGTC
CcN-ASGR-BBM-like1500 (2503) GACTT------GTGTAGGTACTCAGGAGGAAGCTGCGGAGGCTTATGACATTGCCGCAATCAAATTTCGAGGCCTTAATGCCGTC
CcASGR-BBM-like2 (2731) GACCTCTCCCTGTGTAGGTACCCAGGAGGAAGCTGCAGAGGCTTACGACATTGCTGCCATCAAATTCCGAGGCCTCAATGCTGTC
CcASGR-BBM-like1 (2706) GACCTCTCCCTGTGTAGGTACCCAGGAGGAAGCTGCAGAGGCTTACGACATTGCTGCCATCAAATTCCGAGGCCTCAATGCTGTC
Consensus (2891) GACTTCTCCCTGTGTAGGTACTCAGGAGGAAGCTGCGGAGGCTTATGACATTGCTGCCATCAAATTTCGAGGCCTTAATGCTGTC


Section 36
(2976) 2976 2990 3000 3010 3020 3030 3040 3050 3060
CcN-ASGR-BBM-like1400 (2449) ACAAACTTTGACATGACCCGGTATGATGTCAAGAGCATCATCGAGAGCAGCTCCCTGCCAGTTGGTGGCACCATGAAGCGTCTCA
CcN-ASGR-BBM-like1500 (2582) ACAAACTTTGACATGAGCCGGTATGATGTCAAGAGCATCATCGAGAGCAGCTCCCTGCCAGTTGGTGGCACCACGAAGCGTCTCA
CcASGR-BBM-like2 (2816) ACGAACTTTGACATGAGCCGGTATGACGTCAAGAGCATCATTGAGAGCAGCTCCCTGCCTGTTGGCGGCACTCCAAAGCGTCTCA
CcASGR-BBM-like1 (2791) ACGAACTTTGACATGAGCCGGTATGACGTCAAGAGCATCATTGAGAGCAGCTCCCTGCCTGTTGGCGGCGCTCCAAAGCGTCTCA
Consensus (2976) ACGAACTTTGACATGAGCCGGTATGATGTCAAGAGCATCATTGAGAGCAGCTCCCTGCCTGTTGGTGGCACTCCGAAGCGTCTCA
Section 37
(3061) 3061 3070 3080 3090 3100 3110 3120 3130 3145
CcN-ASGR-BBM-like1400 (2534) AGGATGTGCCTGATCAATCCGACTTGGGCATGAATGGCTACGGTGCTGAATCTGCTGGTCATAAGGCTGCTACAAACCTTCTTAC
CcN-ASGR-BBM-like1500 (2667) AGGATGTGCCTGATCAACCTGACTTGGGCATGAATGGCTACGGTGCTGAATCTGCTGGTCATAAGGCTGCTACAAACCTTCTTAC
CcASGR-BBM-like2 (2901) AGGAAGTGCCTGATCAATCAGATATGGGCATCAACATAAACGGTG---ACTCTGCTGGTCATATGACTGCTATCAACCTTCTTAC
CcASGR-BBM-like1 (2876) AGGAAGTGCCTGATCAATCAGATATGGGCATCAACATAAACGGTG---ACTCTGCTGGTCATATGACTGCTATCAACCTTCTTAC
Consensus (3061) AGGATGTGCCTGATCAATCAGATTTGGGCATGAATGTCTACGGTGCTGACTCTGCTGGTCATATGGCTGCTATCAACCTTCTTAC
Section 38
(3146) 3146 3160 3170 3180 3190 3200 3210 3220 3230
CcN-ASGR-BBM-like1400 (2619) TGACGGCATTGGCAGCTATGGCCCTGAGAGTAAATATGGTTATAGTGGCTGGTCTCCCGCTGCTATGACATCAATCCCCTTGCCA
CcN-ASGR-BBM-like1500 (2752) TGATGGCATTGGCAGCTATGGCCCTGAGAGTAAATATGGTTATAGTGGCTGGTCTCCCGCTGCTATGACATCAATCCCCTTGCAA
CcASGR-BBM-like2 (2983) TGATGGCAATGACAGCTATGGAGCTGAGAGT---TATGGTTACAGTGGTTGGTGTCCCACAGCCATGACGCCAATCCCCTTTCAA
CcASGR-BBM-like1 (2958) TGATGGCAATGACAGCTATGGAGCTGAGAGT---TATGGTTACAGTGGTTGGTGTCCCACAGCCATGACGCCAATCCCCTTTCAA
Consensus (3146) TGATGGCATTGGCAGCTATGGCGCTGAGAGTAAATATGGTTATAGTGGTTGGTGTCCCGCTGCTATGACGTCAATCCCCTTTCAA
Section 39
(3231) 3231 3240 3250 3260 3270 3280 3290 3300 3315
CcN-ASGR-BBM-like1400 (2704) TTCAGCAATGGCCATGAACAGTCCAGGCTGTGGTGCAAGCCAGAGCAGGACAATGCGGCTGTTGCGGCAGCATACAACCTGCACC
CcN-ASGR-BBM-like1500 (2837) TTCAGCAATGGCCATGAACAGTCCAGGCTGTGGTGCAAGCCAGAGCAGGACAATGCGGCTGTTGCGGCAGCATACAACCTGCACC
CcASGR-BBM-like2 (3065) TTCAGCAATGGCCATGACCATTCCAGGCTGTGGTGCAAGCCAGAGCAGGACAATGCGGTTGTTGCAGCACTGCATAACCTGCATC
CcASGR-BBM-like1 (3040) TTCAGCATTGGCCATGACCATTCCAGGCTGTGGTGCAAGCCAGAGCAGGACAATGCGGTTGTTGCAGCACTGCATAACCTGCATC
Consensus (3231) TTCAGCAATGGCCATGACCATTCCAGGCTGTGGTGCAAGCCAGAGCAGGACAATGCGGTTGTTGCGGCAGTGTATAACCTGCATC
Section 40
(3316) 3316 3330 3340 3350 3360 3370 3380 3390 3400
CcN-ASGR-BBM-like1400 (2789) ACCTCCAGCACTTCCCGGCCCCAGGTGGCACCCACAACTTTTTCGAGCCATCCCTTGTTCAGGACATGACAGGTGTTGCTGATGC
CcN-ASGR-BBM-like1500 (2922) ACCTCCAGCACTTCCCGGCCCCAGGTGGCACCCACAACTTTTTCGAGCCATCCCCTGTTCAGGACATGACAGGTGTTGCTGATGC
CcASGR-BBM-like2 (3150) ACCTCCAGCACTTGCCAGCCCCAGTTGGCACCCATAATTTTTTCTAGCCATCGCCTGTTCAGGACATGACAGGTGTTGCCGATGC
CcASGR-BBM-like1 (3125) ACCTCCAGCACTTGCCAGCCCCAGTTGGCACCCATAATTTTTTCCAGCCATCGCCTGTTCAGGACATGACAGGTGTTGCCGATGC
Consensus (3316) ACCTCCAGCACTTGCCGGCCCCAGTTGGCACCCATAATTTTTTCGAGCCATCGCCTGTTCAGGACATGACAGGTGTTGCTGATGC

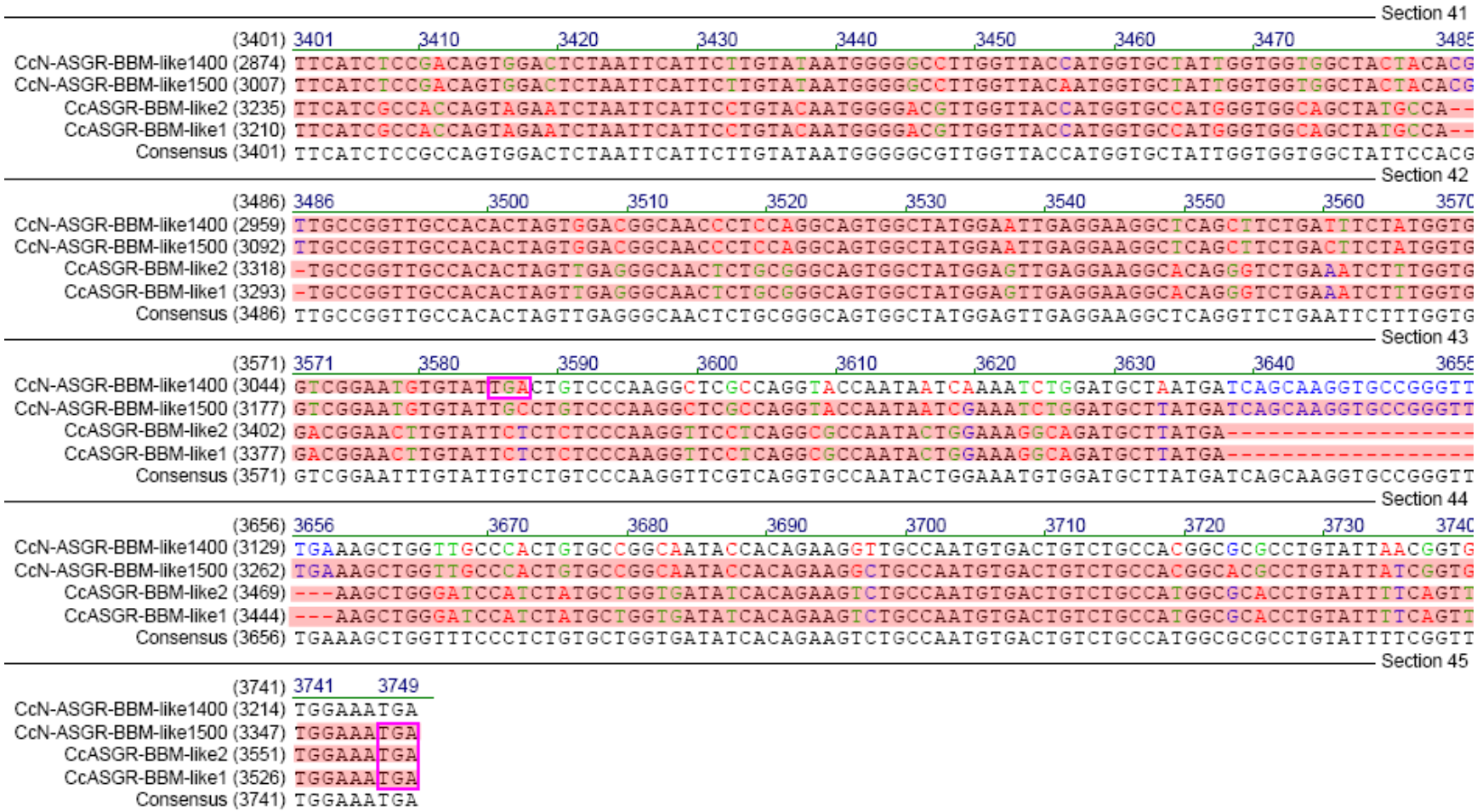


Figure 4.3 *CcASGR-BBM-like* and *CcN-ASGR-BBM-like* genomic sequence alignment. Red shading highlights the rice gene prediction program predicted exons. Pink rectangles highlight the stop codons of each gene.

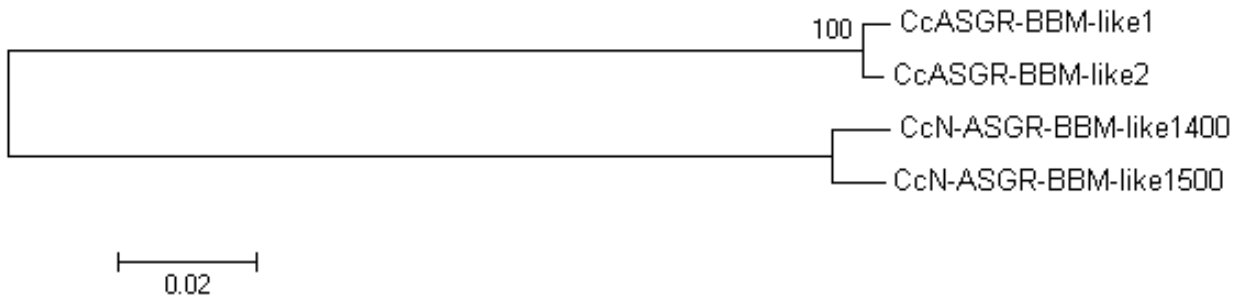


Figure 4.4 Phylogenetic tree was constructed with *CcBBM-like* genomic sequences by using the neighbor-joining method and the Kimura 2-parameter model with 1,000 bootstrap replicates.

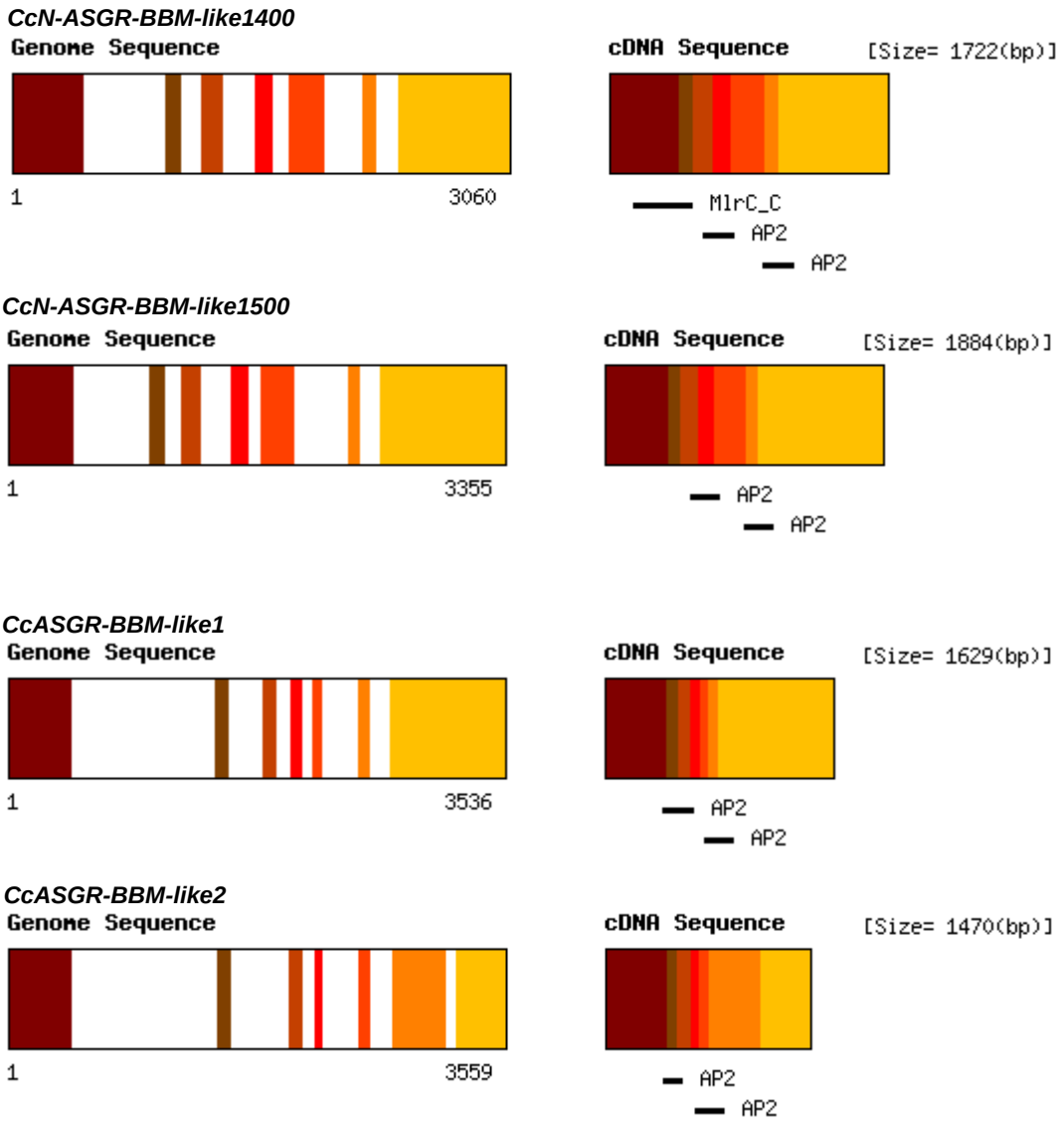


Figure 4.5 Predicted splice sites of four *BBM-like* genes. MlrC_C represents the C-terminus (approximately 200 residues) of the product of a bacterial gene cluster that is involved in the degradation of the cyanobacterial toxin microcystin LR.

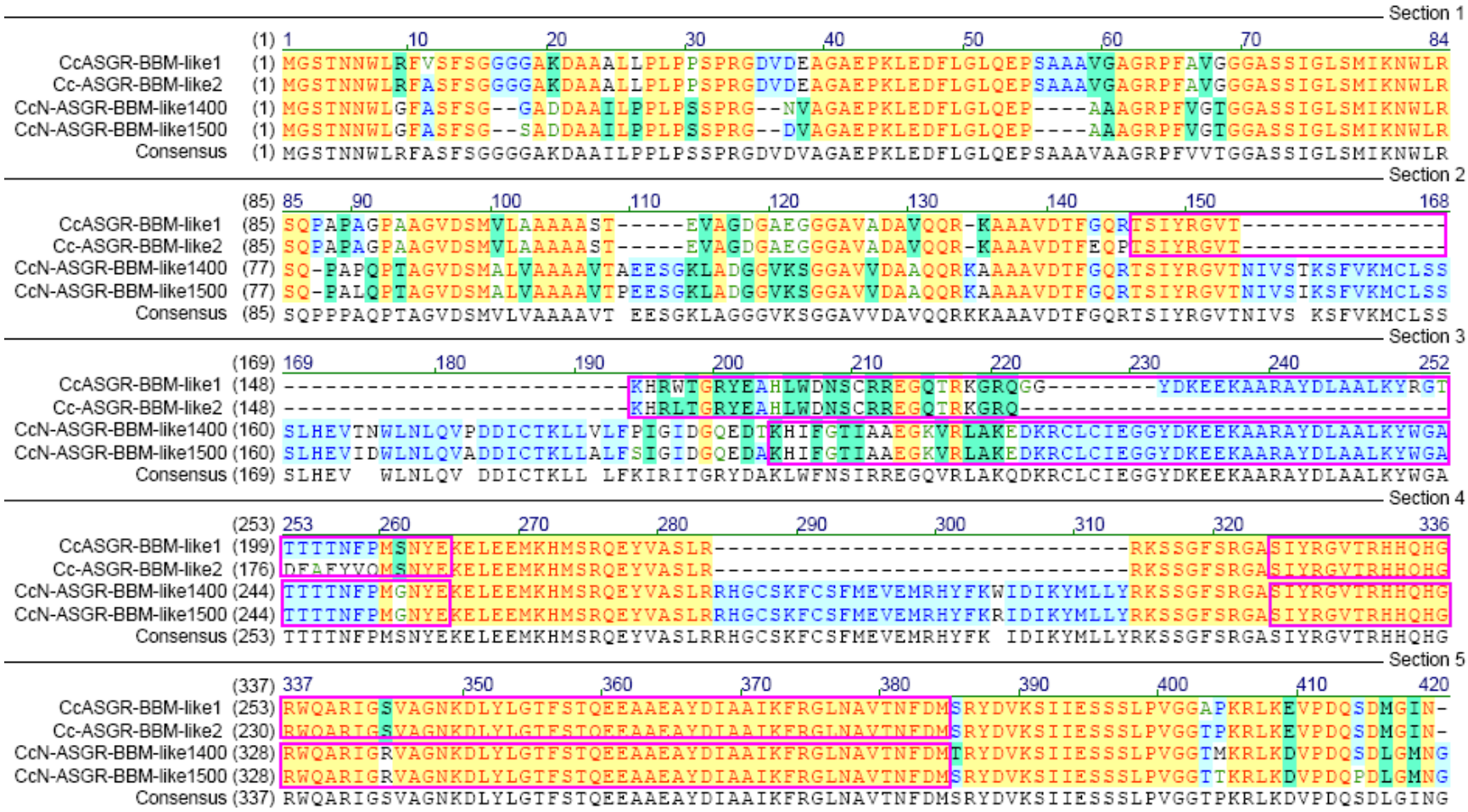

Section 1
CcASGR-BBM-like1 (1) MGSTNNWLRFVSFSGGGGAKDAAALLPLPPSPRGDVDEAGAEPKLEDFLGLQEPSAAAVGAGRPFAVGGGASSIGLSMIKNWLR
Cc-ASGR-BBM-like2 (1) MGSTNNWLRFASFSGGGGAKDAAALLPLPPSPRGDVDEAGAEPKLEDFLGLQEPSAAAVGAGRPFAVGGGASSIGLSMIKNWLR
CcN-ASGR-BBM-like1400 (1) MGSTNNWLGFASFSG--GADDAAILPPLPSSPRG--NVAGAEPKLEDFLGLQEP----AAAGRPFVGTGGASSIGLSMIKNWLR
CcN-ASGR-BBM-like1500 (1) MGSTNNWLGFASFSG--SADDAAILPPLPSSPRG--DVAGAEPKLEDFLGLQEP----AAAGRPFVGTGGASSIGLSMIKNWLR
Consensus (1) MGSTNNWLRFASFSGGGGAKDAAILPPLPSSPRGDVDVAGAEPKLEDFLGLQEPSAAAVAAGRPFVVTGGASSIGLSMIKNWLR
Section 2
CcASGR-BBM-like1 (85) SQPAPAGPAAGVDSMVLAAAAAST-----EVAGDGAEGGGAVADAVQQR-KAAAVDTFGQRTSIYRGVT---------------
Cc-ASGR-BBM-like2 (85) SQPAPAGPAAGVDSMVLAAAAAST-----EVAGDGAEGGGAVADAVQQR-KAAAVDTFEQPTSIYRGVT---------------
CcN-ASGR-BBM-like1400 (77) SQ-PAPQPTAGVDSMALVAAAAVTAEESGKLADGGVKSGGAVVDAAQQRKAAAAVDTFGQRTSIYRGVTNIVSTKSFVKMCLSS
CcN-ASGR-BBM-like1500 (77) SQ-PALQPTAGVDSMALVAAAAVTPEESGKLADGGVKSGGAVVDAAQQRKAAAAVDTFGQRTSIYRGVTNIVSIKSFVKMCLSS
Consensus (85) SQPPPAQPTAGVDSMVLVAAAAVT EESGKLAGGGVKSGGAVVDAVQQRKKAAAVDTFGQRTSIYRGVTNIVS KSFVKMCLSS
Section 3
CcASGR-BBM-like1 (148) ------------------------KHRWTGRYEAHLWDNSCRREGQTRKGRQGG--------YDKEEKAARAYDLAALKYRGT
Cc-ASGR-BBM-like2 (148) ------------------------KHRLTGRYEAHLWDNSCRREGQTRKGRQ----------------------------
CcN-ASGR-BBM-like1400 (160) SLHEVTNWLNLQVPDDICTKLLVLFPIGIDGQEDTKHIFGTIAAEGKVRLAKEDKRCLCIEGGYDKEEKAARAYDLAALKYWGA
CcN-ASGR-BBM-like1500 (160) SLHEVIDWLNLQVADDICTKLLALFSIGIDGQEDAKHIFGTIAAEGKVRLAKEDKRCLCIEGGYDKEEKAARAYDLAALKYWGA
Consensus (169) SLHEV WLNLQV DDICTKLL LFKIRITGRYDAKLWFNSIRREGQVRLAKQDKRCLCIEGGYDKEEKAARAYDLAALKYWGA
Section 4
CcASGR-BBM-like1 (199) TTTTNFPMSNYEKELEEMKHMSRQEYVASLR----------RKSSGFSRGASIYRGVTRHHQHG
Cc-ASGR-BBM-like2 (176) DFAFYVQMSNYEKELEEMKHMSRQEYVASLR----------RKSSGFSRGASIYRGVTRHHQHG
CcN-ASGR-BBM-like1400 (244) TTTTNFPMGNYEKELEEMKHMSRQEYVASLRRHGCSKFCSFMEVEMRHYFKWIDIKYMLLYRKSSGFSRGASIYRGVTRHHQHG
CcN-ASGR-BBM-like1500 (244) TTTTNFPMGNYEKELEEMKHMSRQEYVASLRRHGCSKFCSFMEVEMRHYFKRIDIKYMLLYRKSSGFSRGASIYRGVTRHHQHG
Consensus (253) TTTTNFPMSNYEKELEEMKHMSRQEYVASLRRHGCSKFCSFMEVEMRHYFK IDIKYMLLYRKSSGFSRGASIYRGVTRHHQHG
Section 5
CcASGR-BBM-like1 (253) RWQARIGSVAGNKDLYLGTFSTQEEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSIIESSSLPVGGAPKRLKEVPDQSDMGIN-
Cc-ASGR-BBM-like2 (230) RWQARIGSVAGNKDLYLGTFSTQEEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSIIESSSLPVGGTPKRLKEVPDQSDMGIN-
CcN-ASGR-BBM-like1400 (328) RWQARIGRVAGNKDLYLGTFSTQEEAAEAYDIAAIKFRGLNAVTNFDMTRYDVKSIIESSSLPVGGTMKRLKDVPDQSDLGMNG
CcN-ASGR-BBM-like1500 (328) RWQARIGRVAGNKDLYLGTFSTQEEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSIIESSSLPVGGTTKRLKDVPDQPDLGMNG
Consensus (337) RWQARIGSVAGNKDLYLGTFSTQEEAAEAYDIAAIKFRGLNAVTNFDMSRYDVKSIIESSSLPVGGTPKRLKDVPDQSDLGING

Figure 4.6 Alignment of predicted proteins of the four Cc*BBM-like* genes. Pink rectangles highlight the two AP2 domains of each gene.

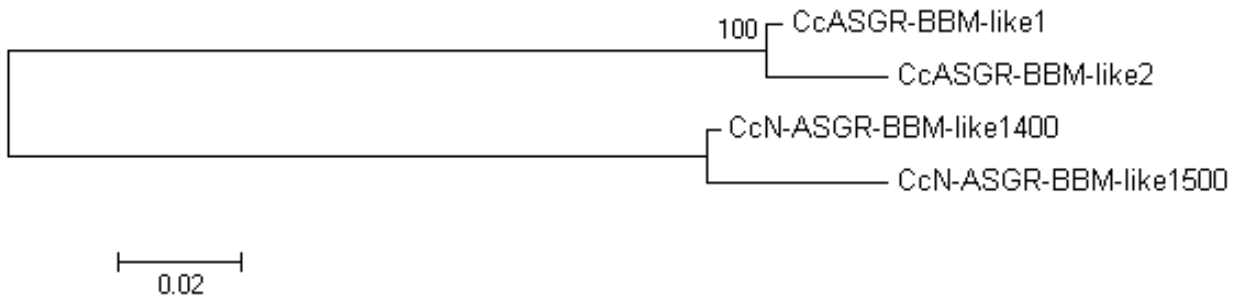


Figure 4.7 Phylogenetic tree was constructed with predicted amino acid sequences of *CcBBM-like* genes by using the neighbor-joining method and the PAM matrix model with 1,000 bootstrap replicates.

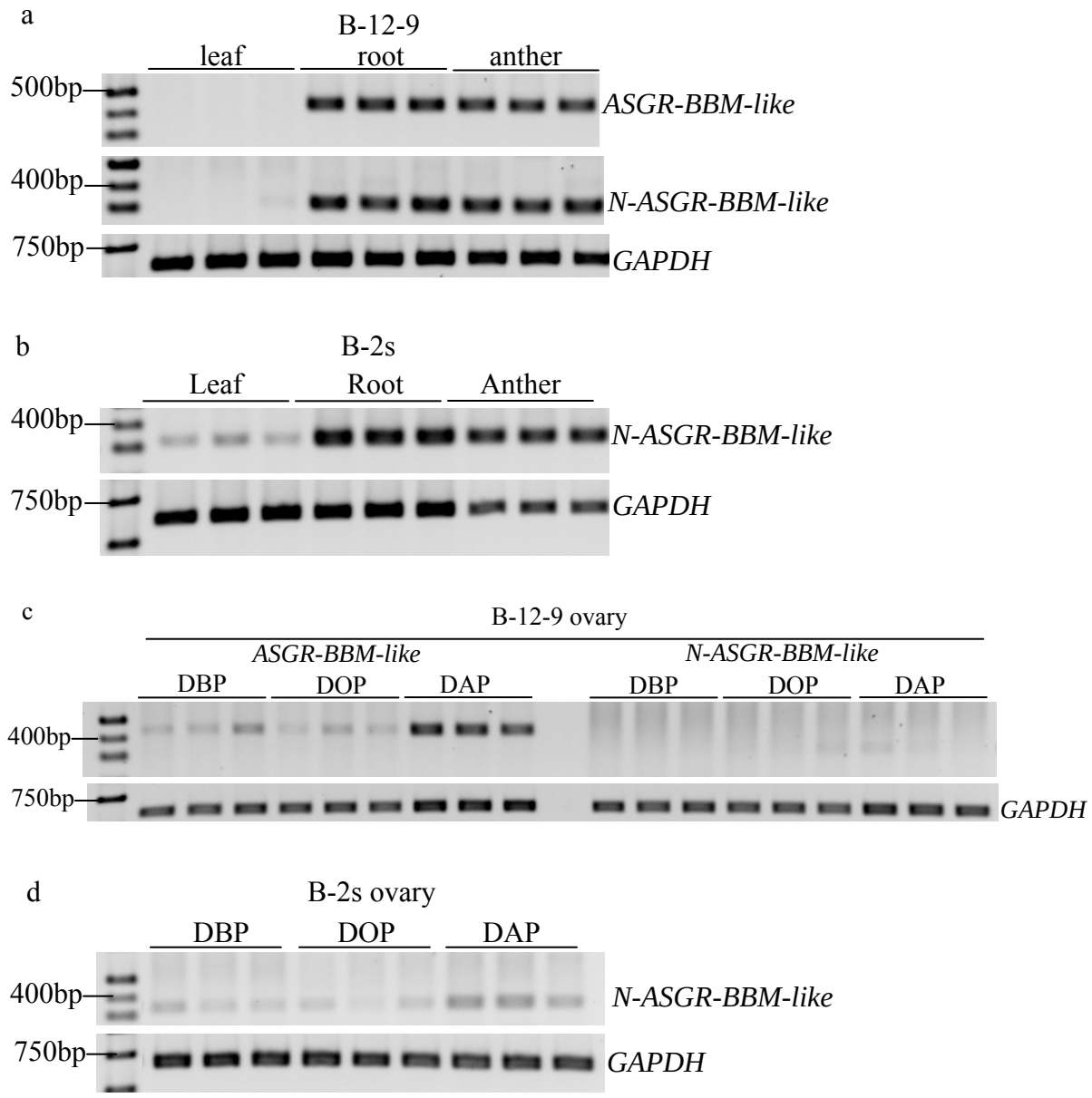


Figure 4.8 Expression patterns of *BBM-like* genes. Twenty-one cycle of RT-PCR for the internal control glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and 37 cycles of RT-PCR for *BBM-like* genes were performed. Each set of RT-PCR reactions was performed in triplicate and each time point/tissue was processed with biological triplicates. Figures show the results of one of the biological triplicates since quantitative differences between biological triplicates were not statistically different. DBP-one day before pollination, DOP-the day of pollination, DAP-one day after pollination.

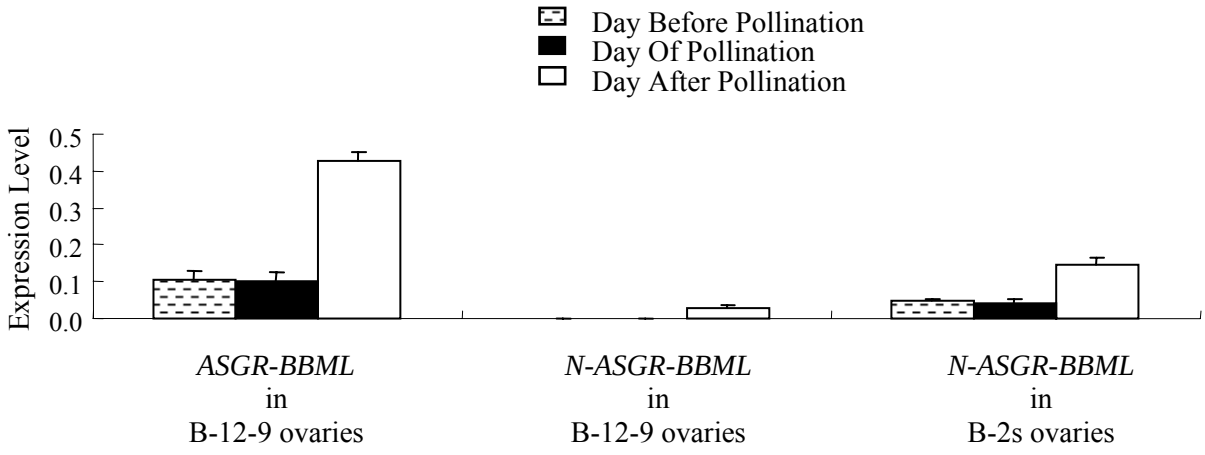


Figure 4.9 Semi-quantitative RT-PCR analysis of expression of *BBM-like* genes in ovaries relative to the expression of *GAPDH* gene. RT-PCR products were separated on agarose gels and the pixels of each band were quantified with ImageQuant software. To evaluate the changes in relative levels of mRNAs, the pixel counts for the target genes were divided against those of the housekeeping *GAPDH* mRNA. Values were expressed in ratio between the target gene and *GAPDH*. The data shown here was an average of the biological triplicates, with which each set of RT-PCR reaction was performed in triplicate.

CHAPTER 5

CONCLUSIONS

Apospory is a form of apomixis where the embryo develops from an unreduced egg in an embryo sac derived through mitosis of a somatic nucellar cell. It has been reported that apospory was inherited as a dominant Mendelian trait and this trait was associated with an approximately 50 Mb, heterochromatic and hemizygous aposporous specific genomic region (ASGR) in *Pennisetum squamulatum* and *Cenchrus ciliaris*. To elucidate the genetic control regulating apospory, two studies conducted were focus on identification and characterization of apospory candidate genes in *P. squamulatum* and *C. ciliaris*.

In the first study, we compared two transcriptomes generated by Roche 454 high-throughput sequencing technology from dissected ovule tissues staged for (aposporous initial) AI formation from two apomictic lines, *P. squamulatum* (Ps26) and its apomictic derivative backcross 8 (BC_8). These two lines were chosen for their common features of apospory and a single shared chromosome, the ASGR-carrier chromosome. Our hypothesis was that candidate genes for regulating AI specification should localize to the ASGR, function in both Ps26 and BC_8 at the same developmental stage, and be identical in sequence. Analysis of the two aposporous ovule transcriptomes resulted in identification of 61 putative ASGR-carrier chromosome candidate expressed genes, of which 46 have been confirmed after screening by PCR with Ps26, IA4X (sexual), N37 (sexual) and a small number of progeny from apomictic BC_8 segregating for mode of reproduction. The 46 confirmed ASGR-carrier chromosome-linked contigs were further screened by PCR with a limited number of apomictic and sexual F_1s for

mapping to the ASGR. To date, only one ASGR-linked contig, Ps26_c9369, was assigned to the ASGR. Mapping of the remaining 15 contigs to the ASGR-carrier chromosome and testing for linkage of the ASGR-carrier chromosome linked candidate genes to the ASGR will require application of other methods of mapping. *P. squamulatum* is an octoploid plant; therefore, mapping can only be effective when single-dose alleles are identified. In contrast to the large number of candidates resulting from other studies comparing transcriptomes between apomictic and sexual genotypes, our strategy of transcriptome comparison between two apomictic lines gave rise to a much smaller group of candidates because we were able to separate ASGR-carrier chromosome linked candidates from unlinked candidates whose expression may change in response to ASGR-carrier chromosome controlled upstream events. This is a great advantage for functional analysis of candidate genes although we may lose the down-regulated candidate genes in apomictic ovules since we were unable to include the sexual genotype in this study.

To obtain longer sequences of the candidate genes to design primers for testing their linkage to the ASGR, a cDNA library containing about 300,000 phage clones was constructed from apomictic BC_8 mature ovary and anther RNA. This source of RNA was chosen because of the observed gene expression patterns of a set of the candidate genes. RT-PCR with RNA extracted from apomictic BC_8 leaf, root, and anther and ovary tissues was conducted for a set of 36 ASGR-carrier chromosome linked genes to analyze their expression patterns. Except one putative MADS-domain containing transcription factor, which showed reproductive tissue specific amplification, and a putative Lon protease, which showed expression in all organs except anther, the remaining 34 were expressed in all four organ types examined. Screening of the cDNA library with probes amplified from 27 contigs resulted in recovery of primary plaques for 24 probes. PCR screening with contig-specific primers followed by secondary screening

identified individual phage clones containing 17 of the desired transcript sequences. Sequencing of the 17 clones returned with good quality sequences for 14 of the candidate genes. The other 3 clones returned with bad quality sequences therefore more sequencing will be required to determine their sequences. Primers will be designed based on the more complete gene sequences obtained from the cDNA clones to test the linkage of the ASGR-carrier chromosome linked expressed genes to the ASGR. Without functional analysis of candidate apospory initiation regulating genes identified from ovule transcriptomes the genetic mechanism regulating apospory initiation remains unclear. However, the ovule transcriptomes provide significant new data for studying early ovule development and potential insight into reproductive pathways.

Two *ASGR-BABY BOOM-like* (*ASGR-BBM-like*) genes have been isolated from ASGR-linked bacterial artificial chromosome clones (BACs) of apomictic buffelgrass B-12-9, whose nucleotide sequences and predicted protein sequences shared high similarity with *BBM* genes from rice, *Medicago*, *Brassica* and *Arabidopsis*. Functional analyses of *BBM* in other species suggest its important role during zygotic embryogenesis. The present study was to investigate the function of *ASGR-BBM-like* genes during parthenogenesis. First we identified the *N-ASGR-BBM-like* genes, which exist in buffelgrass but are not associated with the ASGR, by screening the BAC library of apomictic buffelgrass with the probes containing 5'-end, AP2 domain and 3'-end sequences of previously identified *ASGR-BBM-like*. Southern blot analysis was conducted to confirm their existence and estimate their copy numbers. To obtain the *N-ASGR-BBM-like* sequences, two identified BACs containing two potential *N-ASGR-BBM-like* genes were selected for shotgun library construction and named as C1400 (original BAC address: C091M03) and C1500 (original address: C164L13), respectively. Sequencing of the positive subclones from C1400 and C1500 followed by DNA sequence analysis identified two *N-ASGR-BBM-like* genes:

the *CcASGR-BBM-like1400* gene derived from BAC C1400 and the *CcASGR-BBM-like1500* gene derived from BAC C1500. Similar to the two *ASGR-BBM-like* genes, the two *N-ASGR-BBM-like* genes shared high similarity in nucleotide sequences. Analyzed with the gene prediction program at RiceGAAS, the four *BBM-like* genes from buffelgrass were all predicted to produce translatable transcripts with predicted proteins containing two AP2 domains. When the predicted amino acid sequences of *N-ASGR-BBM-like* genes were compared to *ASGR-BABY BOOM-like* genes, their similarity was decreased compared to that of nucleotide sequences. Predicted splice site changes caused the main differences including the differences in the first AP2 domain. However, the second AP2 domain was highly conserved among the four genes. With only one amino acid change between *N-ASGR-BBM-like* and *ASGR-BBM-like*, the second AP2 domain was identical between the two *N-ASGR-BBM-like* proteins as well as between *ASGR-BBM-like1* and *ASGR-BBM-like2*. Prediction artifact seems likely based on our unpublished data from the *PsASGR-BBM-like1* (EU559277) gene. To confirm the authenticity of the predictions, the cDNA sequences covering the first AP2 domain will have to be obtained.

Semi-quantitative RT-PCR with *CcASGR-BBM-like*/*CcN-ASGR-BBM-like* specific primers showed that *ASGR-BBM-like* had continuous expression in aposporous ovaries before and after pollination with a significant increase after pollination in B-12-9 while there was no detectable transcription of *N-ASGR-BBM-like* until one day after pollination. However, the expression pattern of *N-ASGR-BBM-like* in the three stages (one day before pollination, day of pollination and one day after pollination) of B-2s (sexual) ovaries was somewhat similar to that of *ASGR-BBM-like* in B-12-9 (aposporous) ovaries. Whether the transcription differences were caused by protein interactions or heterochronic regulatory changes remains unknown. With respect to the important role of *BBM* genes in induction of embryonic cell development in

sexually reproducing plants, different expression patterns of *ASGR-BBM-like* and *N-ASGR-BBM-like* in aposporous ovules supported our hypothesis that *ASGR-BBM-like* functioned in the initiation and/or maintenance of parthenogenesis.

Apospory initial specification and parthenogenesis are two critical events for occurrence of apospory. By isolation of ASGR-carrier chromosome linked expressed sequence tags (ESTs) from dissected ovule tissues staged for aposporous initial formation, we provide potential significant insight into the regulation of apospory initiation. And the expression study of the candidate gene suggests that genes regulating sexual embryogenesis may also involve in parthenogenesis with modified expression patterns.