CONTROLLING THE INDUCTION OF PARTHENOGENESIS IN TRANSGENIC RICE VIA POST-TRANSLATIONAL ACTIVATION OF PsASGR-BBML AND DETERMINING THE EFFECT OF EGG CELL PLOIDY LEVEL ON PENETRANCE OF PARTHENOGENESIS

by

GURJOT SINGH SIDHU

(Under the Direction of Joann A. Conner)

ABSTRACT

Doubled-haploid (DH) technology significantly shortens the process of generating homozygous lines and it involves haploid induction followed by chromosome doubling. Parthenogenesis, which involves embryogenesis from unfertilized egg cells, can be used for *in vivo* haploid induction and an AP2 transcription factor, PsASGR-BBML induces parthenogenesis in a natural apomict *Pennisetum squamulatum*. *PsASGR-BBML* transgenes promote parthenogenesis in several crop plants, however their dominant nature inhibits their use for DH technology. We show that a post-translational activation system can be used to regulate PsASGR-BBML at anthesis. Also, in rice transformed with *PsASGR-BBML*, parthenogenesis occurs in reduced egg cells and this may be one cause for the observed low penetrance of the trait and abnormalities in developing ovaries. In natural apomicts, which are mostly polyploid, parthenogenesis occurs in unreduced egg cells. Here, we tried to observe if seed-set and penetrance of parthenogenesis is affected by the ploidy level of egg cells in transgenic rice.

INDEX WORDS: apomixis, *PsASGR-BBML*, parthenogenesis for haploid induction, Glucocorticoid receptor, polyploidy and parthenogenesis, *osd1*

CONTROLLING THE INDUCTION OF PARTHENOGENESIS IN TRANSGENIC RICE VIA POST-TRANSLATIONAL ACTIVATION OF PsASGR-BBML AND DETERMINING THE EFFECT OF EGG CELL PLOIDY LEVEL ON PENETRANCE OF PARTHENOGENESIS

by

GURJOT SINGH SIDHU

B.S., Punjab Agricultural University, India, 2018

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

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

CONTROLLING THE INDUCTION OF PARTHENOGENESIS IN TRANSGENIC RICE VIA POST-TRANSLATIONAL ACTIVATION OF PsASGR-BBML AND DETERMINING THE EFFECT OF EGG CELL PLOIDY LEVEL ON PENETRANCE OF PARTHENOGENESIS

by

GURJOT SINGH SIDHU

Major Professor: Joann A. Conner

Committee: Peggy Ozias-Akins

Katrien Devos

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

May 2021

DEDICATION

To the desire of discovery inherent in *Homo sapiens*.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION & LITERATURE REVIEW

Hybrids are preferred in modern agriculture as they give higher yields due to the phenomenon of heterosis. However, due to genetic recombination during sexual reproduction, hybrid vigor breaks down after the first generation and farmers are forced to buy new seed every year. Moreover, the large number of resources consumed in the maintenance of inbred lines and the tedious procedures of emasculation and pollination result in high costs of hybrid seeds, which limits the adoption of hybrid cultivars in developing countries. By engineering apomictic crop plants, heterosis can be fixed, which will reduce the need for farmers to buy new hybrid seed every year. This can lead to wide adoption of hybrids resulting in a significant increase in crop yields and hence can contribute to ensuring global food security.

The process of sexual reproduction

The life cycle of plants undergoing sexual reproduction alternates between a diploid sporophytic phase and a haploid gametophytic phase. In angiosperms, specialized diploid cells from the sporophytic phase, known as mother cells, undergo meiosis to produce a tetrad of genetically recombined, haploid spores (Drews and Koltunow, 2011). Female spores (megaspores) are produced in the ovules, in a process known as megasporogenesis while male spores (microspores) are produced in anthers, in a process called microsporogenesis. While three of the megaspores from a tetrad degrade, one undergoes three consecutive mitotic divisions to produce a seven-celled and eight-nucleated female gametophyte, known as an embryo sac. The most abundant type of embryo sac is the polygonum-type, which is comprised of one egg cell and two

synergid cells at the micropylar end, a central cell comprising two polar nuclei and three antipodal cells at the chalazal end. In microsporogenesis, all the four microspores from a tetrad survive and divide mitotically to produce a vegetative nucleus and a generative cell. The generative cell undergoes one more mitotic division to produce two sperm cells, which after pollination, travel to the embryo sac through the pollen tube. During double-fertilization, one of the sperm nuclei fertilizes the egg cell and the other fuses with the polar nuclei, leading to the formation of a diploid zygote and a triploid endosperm respectively. The zygote divides mitotically to produce an embryo, which along with endosperm, ultimately leads to the formation of seed.

Introduction to apomixis

Apomixis, defined as asexual reproduction through seed, involves the modification of the sexual process and leads to progeny genetically identical to the mother plant (Asker and Jerling, 1992). Apomixis is found in more than 300 angiosperm plant species from about 40 families although most of these belong to three families – *Asteraceae*, *Rosaceae* and *Poaceae* (Whitton et al., 2008). In nature, apomixis is achieved through either of the two developmental pathways – sporophytic or gametophytic apomixis (Conner and Ozias-Akins, 2017). In sporophytic apomixis, somatic cells (known as embryo initial cells) from the ovule undergo repeated mitotic divisions to produce an adventitious embryo without an intermediate embryo sac. As a viable endosperm is required for the nourishment of these apomictic embryos, sexual reproduction in sporophytic apomicts also remains functional which leads to polyembryony, a common phenomenon observed in *Citrus* (Koltunow et al., 1996). In gametophytic apomixis, the apomictic embryo develops from the egg cell of an unreduced embryo sac and this process is known as parthenogenesis. Based on the origin of unreduced embryo sac, gametophytic apomixis is further divided into two categories - apospory, where the unreduced embryo sac develops from a diploid somatic cell known as an

aposporous initial cell, and diplospory, where the megaspore mother cell forms an unreduced embryo sac by avoiding or modifying meiosis. This avoidance or modification of meiosis, known as apomeiosis, not only produces unreduced egg cells, but also skips chromosomal crossing-over which is the main reason behind genetic reshuffling. The gametes produced through apomeiosis are thus genetically identical to the mother parent. In most cases, viable endosperm formation is also required to nurture the embryo, so many apomictic plant species require pollination followed by the fertilization of the central cell – a phenomenon called pseudogamy. However, in a few apomictic species, endosperm may develop autonomously from the central cell. The view of apomixis has shifted from being considered totally different from sexual reproduction to its genetic alteration (Koltunow et al., 2011). Apomictic plant species can be obligate, which reproduce only through apomixis, or facultative, which undergo both apomixis and sexual reproduction producing a mixture of genotypes in their progenies.

Apomixis and polyploidy

Due to the various apomictic mechanisms found in nature and the diverse phylogenetic positioning of various apomictic species, it is assumed that apomixis has evolved independently many times during angiosperm evolution (Hand and Koltunow, 2014). Although apomixis is found in more than 400 plant genera, most of the natural apomicts are polyploid. Earlier theories on apomixis speculated that it was associated with polyploidy (Carman, 1997), but all polyploids are not apomicts and natural as well as induced diploid apomicts have been identified, which suggests that polyploidy is not an absolute requirement for the expression of apomixis. However, it has been observed that expression of apomictic factors may be dependent on the ploidy level of plants. For example, in *Paspalum notatum*, most diploid races are sexual while tetraploid races are usually aposporous apomicts, but when a diploid sexual plant was chromosomally doubled through

colchicine, it exhibited a high frequency of aposporous embryo sacs (Quarin et al., 2001). This suggest that higher ploidy levels might enhance expression of apomictic factors either through higher dosage effect or epigenetic regulation. Obligate apomixis will lead to an evolutionary dead end, however, the facultative nature of most apomicts provides these species a chance of sexual recombination and evolution (Hojsgaard, 2012).

Genetics of apomixis

Earlier studies demonstrated that a single dominant locus with low recombination controls apomixis (Ozias-Akins et al., 1998). However, further genetic analysis has revealed that in some species various developmental components of apomixis such as apomeiosis, parthenogenesis and fertilization-independent endosperm development are controlled by independent loci (Catanach et al., 2006; Conner et al., 2013). Three independent loci – *Loss of Apomeiosis* (*LOA*), *Loss of Parthenogenesis* (*LOP*) and *Autonomous endosperm* (*AutE*) have been identified in *Hieracium* that control these three components of apomixis (Koltunow et al., 2011; Ogawa et al., 2013). Due to suppressed recombination around the apomixis loci, map-based cloning approaches to identify apomixis related genes has resulted in little success. This loss of recombination also results in apomictic loci in many species being heterochromatic and possessing substantial repeat sequences, which in extreme cases have resulted in hemizygosity at the locus. The degree of recombination and the chromosome structure at the apomixis loci can provide insights about the evolution of apomixis in a given genus. For example, genera in which individual apomixis components can be separated by recombination are considered to be recent apomicts (*Poa*, *Taraxacum*, *Panicum*) while the genera in which recombination is difficult are considered older apomicts (*Pennisetum*, *Brachiaria*, *Paspalum*) (Hand and Koltunow, 2014).

Apomixis in genus *Pennisetum*

The genus *Pennisetum* (reclassified as *Cenchrus*) of the *Poaceae* family contains many aposporous apomictic species including *Pennisetum squamulatum* (2n = 8X = 56) and *Cenchrus ciliaris* (2n = 4X = 36). This genus also contains Pearl millet (*Pennisetum glaucum*, 2n = 2X = 14), a sexual, diploid species widely cultivated as a forage and grain crop in Africa and South-east Asia. One of the earliest attempts to introgress apomixis in a crop plant from its wild relatives by traditional crossing has been made in this genus (Dujardin and Hanna, 1989). For mapping the apospory locus, Ozias-Akins et al. (1998) used *P. squamulatum* as a pollen parent in a cross with an induced tetraploid of sexual *P. glaucum* (2n = 4X = 28). Analysis of RAPD (random amplified polymorphic DNA) markers revealed that apospory in *P. squamulatum* is controlled by a physically large, hemizygous, and non-recombining chromosomal region, which was named the apospory-specific genomic region (ASGR). The ASGR is heterochromatic and located at the telomeric end of a single chromosome in *P. squamulatum* (Akiyama et al., 2004). Sequencing of Bacterial Artificial Chromosome (BAC) clones from *P. squamulatum* identified many putative protein-coding regions in the ASGR, one of which shared high protein similarity to the *Brassica napus BABY BOOM* (*BnBBM*) gene and was named *PsASGR-BABY BOOM-like* (*PsASGR-BBML*) (Conner et al., 2008). *BBM-like* genes are known for their function in promoting cell differentiation and inducing somatic embryogenesis (Jha and Kumar, 2018). Furthermore, the *ASGR-BBML* sequences are conserved between various apomictic species within genus *Pennisetum* (Akiyama et al., 2011). A loss of *CcASGR-BBML* genes, due to recombination, in the ASGR region of *C. ciliaris* is associated with lost parthenogenesis (Conner et al., 2013). These observations suggested that *ASGR-BBML* genes are strong candidates for inducing parthenogenesis. By using the native promoter from *PsASGR-BBML* to drive the GUS reporter gene in transgenic pearl millet lines,

Conner et al. (2015) observed that the expression of GUS is confined to the egg cell and potentially to adjoining synergid cells in unfertilized ovaries and continues up to three days after fertilization in the developing embryos. This observation suggested that *PsASGR-BBML* expression aligns with the requirement for a parthenogenesis gene that it must be expressed before egg cell fertilization. Function of *PsASGR-BBML* was confirmed when a genomic DNA based transgene expressed under the control of its endogenous promoter was able to induce parthenogenesis in sexual pearl-millet (Conner et al., 2015). Furthermore, both genomic and cDNA based *PsASGR-BBML* transgenes expressed under the control of native and an egg cell-specific promoter *AtDD45* were able to induce parthenogenesis in rice and maize (Conner et al., 2017).

***BABY BOOM-like* genes**

Brassica napus BABY BOOM (*BnBBM*) was first identified as a highly expressed transcript in the somatic embryos induced from *Brassica napus* microspore culture (Boutilier et al., 2002). It was also observed that overexpression of *BnBBM* in *Arabidopsis* and *Brassica napus* leads to somatic embryogenesis which suggested its role in promoting cell proliferation and embryo morphogenesis. This was further supported when Passarinho et al. (2008), using a post-translational activation system coupled with microarray analysis, found that the BnBBM transcription factor activates a complex network of developmental pathways associated with cell proliferation and growth. Cell proliferation due to the overexpression of *BBM-like* genes has been exploited to improve the regeneration capacity of explants in tissue culture. Lowe et al. (2016) over-expressed the maize *BBM* and *WUSCHEL2* genes to successfully transform recalcitrant inbred lines of maize and to enhance transformation efficiency in sorghum, sugarcane and rice. Khanday et al. (2019) also observed that ectopic expression of *OsBBM1*, a *BBM-like* gene from rice, induces somatic embryogenesis when expressed under a constitutive promoter.

BBM and BBM-like proteins are members of the APETALA 2/ETHYLENE RESPONSE FACTOR (AP2/ERF) DNA-binding domain family, which includes a large number of genes found in many plant species (Horstman et al., 2014). The AP2/ERF family is divided into two major clades – ERF-like and AP2-like. The AP2-like clade is further divided into eudicotAP2 and AINTEGUMENTA (ANT) lineages, where ANT bifurcates into basalANT and eudicotANT lineages that contain specific motifs euANT1-4 (Kim et al., 2005). Proteins in the eudicotANT lineage are involved in meristem maintenance, somatic embryogenesis, cell-proliferation and embryo differentiation (Horstman et al., 2014). The EudicotANT lineage is divided into many sub-groups, one of which is BBM-like which contains BBM, BBM-like, and ASGR-BBM-like proteins. These proteins share a conserved bbm-1 domain which is unidentified in other sub-groups within the euANT lineage (El Ouakfaoui et al., 2010). PsASGR-BBML contains two AP2 domains and hence is considered a DNA binding transcription factor (Conner et al., 2015).

***BBM-like* genes and synthetic apomixis**

During sexual reproduction in rice, most transcripts expressed in the zygote are from the female genome, except a few, which are expressed from the male genome (Anderson et al., 2017). Interestingly, *OsBBM1* is expressed from the male genome and triggers embryonic development after fertilization. This explains why an egg cell fails to initiate embryogenesis in the absence of fertilization. Building upon this, Khanday et al. (2019) ectopically expressed *OsBBM1* in rice egg cells using an *Arabidopsis* egg cell-specific promoter *AtDD45*, which leads to the parthenogenetic embryo formation. Parthenogenesis conferred by the egg cell-specific expression of *OsBBM1* was combined to the apomeiotic *MiMe* mutant for engineering the apomictic rice. *MiMe* (*Mitosis instead of Meiosis*) genotypes in *Arabidopsis* and rice combine three mutations which turn meiosis into a mitotic-like division, leading to the production of clonal diploid gametes at a very high

frequency (d'Erfurth et al., 2009; Mieulet et al., 2016). As the diploid *MiMe* genotype invariably produces tetraploid progeny upon self-fertilization, its combination with the egg cell expressed *OsBBM1* parthenogenetic system should produce diploid progeny. Khanday et al. (2019) observed maximum penetrance of apomixis to be about 30 % suggesting that incomplete penetrance of parthenogenesis is a limiting factor in achieving the clonal propagation through this system. In this study, *MiMe* plants expressing egg cell-specific *OsBBM1* were allowed to self-pollinate for facilitating the endosperm development which leads to the production of viable seeds containing parthenogenetically derived haploid embryos.

Apart from parthenogenetic systems, paternal genome elimination after fertilization has also been coupled with *MiMe* mutants for engineering apomixis. Modifications in centromeric histone H3 (CENH3), a histone variant which helps centromeres to bind with the spindle fibers leads to its genome elimination when this parent is crossed to a parent with normal CENH3 (Ravi and Chan, 2010). Marimuthu et al. (2011) crossed the diploid *Arabidopsis MiMe* with a CENH3-mediated chromosome elimination line, leading to the production of diploid clonal progeny. Haploid inducer lines such as Stock-6 in maize represent another system for inducing haploids. Although these lines have been used in maize hybrid breeding since the advent of Doubled-Haploid (DH) production technology, only recently the gene underlying this phenomenon has been identified. *MATRILINEAL* (*MTL*)/ *NOT LIKE DAD* (*NLD*)/ *ZmPHOSPHOLIPASE A1* (*ZmPLA1*) encodes a patatin-like phospholipase expressed primarily in pollen (Gilles et al., 2017; Kelliher et al., 2017; Liu et al., 2017) and a 4 bp insertion in its fourth exon is responsible for the haploid inducer capability of Stock-6 lines. Wang et al. (2019) produced apomictic rice lines by coupling the mutation in *OsMATL* (rice homolog of *ZmMTL*) with *MiMe*, although these lines suffered from low seed-setting (4.4 %) and had lower penetrance of apomixis (6.4 %). These studies suggest the

need for a fully penetrant and stable parthenogenetic system for the successful engineering of synthetic apomixis in crop plants.

References

Akiyama, Y., Conner, J.A., Goel, S., Morishige, D.T., Mullet, J.E., Hanna, W.W., Ozias-Akins, P. (2004) High-Resolution Physical Mapping in Pennisetum squamulatum Reveals Extensive Chromosomal Heteromorphism of the Genomic Region Associated with Apomixis. Plant Physiology, **134**(4), 1733-1733.

Akiyama, Y., Goel, S., Conner, J.A., Hanna, W.W., Yamada-Akiyama, H., Ozias-Akins, P. (2011) Evolution of the apomixis transmitting chromosome in Pennisetum. BMC Evolutionary Biology, **11**(1), 289.

Anderson, S.N., Johnson, C.S., Chesnut, J., Jones, D.S., Khanday, I., Woodhouse, M., Li, C., Conrad, L.J., Russell, S.D., Sundaresan, V. (2017) The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. Developmental Cell, **43**(3), 349-358.e344.

Asker, S.E. and Jerling, L. (1992) Apomixis in Plants. CRC Press, London.

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C.-M., van Lammeren, A.A.M., Miki, B.L.A., Custers, J.B.M., van Lookeren Campagne, M.M. (2002) Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth. The Plant Cell, **14**(8), 1737-1737.

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

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

Conner, J.A., Goel, S., Gunawan, G., Cordonnier-Pratt, M.-M., Johnson, V.E., Liang, C., Wang, H., Pratt, L.H., Mullet, J.E., DeBarry, J., Yang, L., Bennetzen, J.L., Klein, P.E., Ozias-Akins, P. (2008) Sequence Analysis of Bacterial Artificial Chromosome Clones from the Apospory-Specific Genomic Region of Pennisetum and Cenchrus. Plant Physiology, **147**(3), 1396-1396.

Conner, J.A., Gunawan, G., Ozias-Akins, P. (2013) Recombination within the apospory specific genomic region leads to the uncoupling of apomixis components in Cenchrus ciliaris. Planta, **238**(1), 51-63.

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

Conner, J.A. and Ozias-Akins, P. (2017) Apomixis: Engineering the Ability to Harness Hybrid Vigor in Crop Plants (A. Schmidt, ed, Springer New York, New York, NY: pp 17-34.

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

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., Mercier, R. (2009) Turning Meiosis into Mitosis. PLOS Biology, **7**(6), e1000124-e1000124.

Drews, G.N. and Koltunow, A.M.G. (2011) The female gametophyte. The arabidopsis book, **9**, e0155-e0155.

Dujardin, M. and Hanna, W.W. (1989) Developing apomictic pearl millet—characterization of a BC3 plant. J. Genet. Breed., **43:145–51**.

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

Gilles, L.M., Khaled, A., Laffaire, J.B., Chaignon, S., Gendrot, G., Laplaige, J., Bergès, H., Beydon, G., Bayle, V., Barret, P., Comadran, J., Martinant, J.P., Rogowsky, P.M., Widiez, T. (2017) Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. The EMBO Journal, **36**(6), 707-717.

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

Hojsgaard, D. (2012) The evolution of apomixis in angiosperms. Plant Biosystems, **146**(3), 681-693.

Horstman, A., Willemsen, V., Boutilier, K., Heidstra, R. (2014) AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. Trends in Plant Science, **19**(3), 146-157.

Jha, P. and Kumar, V. (2018) BABY BOOM (BBM): a candidate transcription factor gene in plant biotechnology. Biotechnology Letters, **40**(11), 1467-1475.

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M.L., Green, J., Chen, Z., McCuiston, J., Wang, W., Liebler, T., Bullock, P., Martin, B. (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature, **542**(7639), 105-109.

Khanday, I., Skinner, D., Yang, B., Mercier, R., Sundaresan, V. (2019) A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. Nature, **565**(7737), 91-95.

Kim, S., Soltis, P.S., Wall, K., Soltis, D.E. (2005) Phylogeny and Domain Evolution in the APETALA2-like Gene Family. Molecular Biology and Evolution, **23**(1), 107-120.

Koltunow, A.M., Hidaka, T., Robinson, S.P. (1996) Polyembryony in Citrus. Accumulation of seed storage proteins in seeds and in embryos cultured in vitro. Plant Physiology, **110**(2), 599-609.

Koltunow, A.M., Johnson, S.D., Rodrigues, J.C.M., Okada, T., Hu, Y., Tsuchiya, T., Wilson, S., Fletcher, P., Ito, K., Suzuki, G., Mukai, Y., Fehrer, J., Bicknell, R.A. (2011) Sexual reproduction is the default mode in apomictic Hieracium subgenus Pilosella, in which two dominant loci function to enable apomixis. The Plant Journal, **66**(5), 890-902.

Liu, C., Li, X., Meng, D., Zhong, Y., Chen, C., Dong, X., Xu, X., Chen, B., Li, W., Li, L., Tian, X., Zhao, H., Song, W., Luo, H., Zhang, Q., Lai, J., Jin, W., Yan, J., Chen, S. (2017) A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize, Cell Press: pp 520-522.

Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M.-J., Scelonge, C., Lenderts, B., Chamberlin, M., Cushatt, J., Wang, L., Ryan, L., Khan, T., Chow-Yiu, J., Hua, W., Yu, M., Banh, J., Bao, Z., Brink, K., Igo, E., Rudrappa, B., Shamseer, P.M., Bruce, W., Newman, L., Shen, B., Zheng, P., Bidney, D., Falco, C., Register, J., Zhao, Z.-Y., Xu, D., Jones, T., Gordon-Kamm, W. (2016) Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. The Plant Cell, **28**(9), 1998-2015.

Marimuthu, M.P.A., Jolivet, S., Ravi, M., Pereira, L., Davda, J.N., Cromer, L., Wang, L., Nogué, F., Chan, S.W.L., Siddiqi, I., Mercier, R. (2011) Synthetic Clonal Reproduction Through Seeds. Science, **331**(6019), 876-876.

Mieulet, D., Jolivet, S., Rivard, M., Cromer, L., Vernet, A., Mayonove, P., Pereira, L., Droc, G., Courtois, B., Guiderdoni, E., Mercier, R. (2016) Turning rice meiosis into mitosis. Cell Research, **26**, 1242-1242.

Ogawa, D., Johnson, S.D., Henderson, S.T., Koltunow, A.M.G. (2013) Genetic separation of autonomous endosperm formation (AutE) from the two other components of apomixis in Hieracium. Plant Reproduction, **26**(2), 113-123.

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

Passarinho, P., Ketelaar, T., Xing, M., van Arkel, J., Maliepaard, C., Hendriks, M.W., 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**(3), 225-237.

Quarin, C.L., Espinoza, F., Martinez, E.J., Pessino, S.C., Bovo, O.A. (2001) A rise of ploidy level induces the expression of apomixis in Paspalum notatum. 243-249.

Ravi, M. and Chan, S.W.L. (2010) Haploid plants produced by centromere-mediated genome elimination. Nature, **464**(7288), 615-618.

Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., Wu, M., Sun, T., Cheng, Z., Mercier, R., Wang, K. (2019) Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. Nature Biotechnology, **37**(3), 283-286.

Whitton, J., Sears, C., Baack, E., Otto, S. (2008) The Dynamic Nature of Apomixis in the Angiosperms. International Journal of Plant Sciences, **169**(1), 169-182.

CHAPTER 2

CONTROLLING THE INDUCTION OF PARTHENOGENESIS IN TRANSGENIC RICE VIA POST-TRANSLATIONAL ACTIVATION OF PsASGR-BBML[1]

[1]Sidhu, G.S., Conner, J., Ozias-Akins, P. (2021). To be submitted to *Plant Reproduction*.

Abstract

Modern plant breeding programs rely heavily on homozygous lines and the traditional process involves inbreeding a heterozygous stock for 5-6 generations. Doubled haploid (DH) technology, a process of regenerating haploid plants from the initial heterozygous stock followed by chromosome doubling, reduces this process to just two generations. *In vitro* methods of haploid induction include androgenesis and gynogenesis, while *in vivo* methods are based either on parthenogenesis, which involves embryogenesis from unfertilized egg cells, or uni-parental genome elimination. *PsASGR-BBML*, an AP2 transcription factor, induces parthenogenesis in a natural apomict *Pennisetum squamulatum*. *PsASGR-BBML* transgenes also promote parthenogenesis in several crop plants, however, their dominant nature inhibits their use in DH technology. Using a Glucocorticoid based post-translational regulation system, we have successfully shown that PsASGR-BBML can be regulated at the flowering stage, which presents a novel opportunity to use parthenogenetic genes in DH technology and to elucidate the molecular mechanism underlying parthenogenetic embryogenesis.

Introduction

Doubled-haploid (DH) technology is used extensively in various crop breeding programs as it can generate homozygous lines from a heterozygous stock in just two generations as compared to 5-6 generations with traditional breeding (Dunwell, 2010). In cross-pollinated crops such as corn and *Brassica* spp., which show a high degree of heterosis, this technology is used to develop inbred lines which are then crossed to get superior hybrids (Chaikam et al., 2019; Ferrie and Möllers, 2011). In self-pollinated crops, such as wheat and rice, this technology can be used to produce homozygous pure-lines from the heterozygous F_1 progeny of the cross between two homozygous parents (Weyen, 2008). In principle, DH technology consists of inducing plants from the haploid gametophyte followed by chromosome doubling using mitotic inhibitors such as colchicine to get doubled-haploids. The methods to induce haploids are mainly divided into two types – *in vitro* and *in vivo*. *In vitro* methods involve tissue-culture techniques to regenerate haploid cells from microspores or unfertilized egg cells which differentiate to produce embryos and ultimately haploid plants. *In vivo* methods are mainly field-based and produce haploids through uni-parental genome elimination or spontaneous embryogenesis from unfertilized egg cells.

In vitro haploid induction

The induction of haploid plantlets from microspores is called androgenesis and was first reported by Guha and Maheshwari (1964) in *Datura innoxia*. Androgenesis redirects the gametophytic development of microspores to sporophytic development resulting in haploid plant regeneration instead of these developing into a pollen grain. Traditionally, androgenesis involved isolating anthers followed by culturing in sterile medium supplemented with growth regulators to initiate embryogenesis, in a technique called anther culture (Veilleux, 2010). The young embryos are then transferred to a different medium with high cytokinin to auxin ratio in order to encourage

regeneration. Despite the fact that anther culture is a popular technique for haploid induction especially in crops such as rice (Mishra and Rao, 2016), it has several limitations – most important being its genotype specificity within a species (Forster et al., 2007). Some important agricultural species such as woody plants and leguminous crops are recalcitrant to this technique (Murovec and Bohanec, 2012). The occasional regeneration of diploid plants from the somatic tissue of the anther walls constitutes another drawback, and to address that, isolated microspore culture has been proposed as an improved alternative where haploid microspores are isolated from the anthers before media culture. The culturing of un-pollinated floral buds or isolated ovaries on tissue-culture media in order to induce haploids is called gynogenesis and this is technique widely used for DH production in onion and sugar beet (Bohanec, 2009). Gynogenesis is a laborious technique and is usually preferred only in those species where androgenesis is not successful.

***In vivo* haploid induction**

In vivo methods mainly operate through two different mechanisms – uni-parental genome elimination or parthenogenesis. In uni-parental genome elimination, the genome of one parent (inducer line) gets eliminated from the diploid zygote resulting from a cross between two parents while in parthenogenesis, an unfertilized egg cell initiates embryogenesis to produce a haploid plant. An inducer line, Stock-6, and its derivatives have been widely used for haploid induction in maize since the advent of DH technology. In an effort to identify the genomic regions which are responsible for haploid induction in these inducer lines, Prigge et al. (2012) identified two major QTLs - *qhir1* on chromosome one and *qhir8* on chromosome nine, explaining 66 % and 20 % of the genotypic variance for Haploid Induction Rate (HIR), respectively. These QTLs were fine-mapped and a recent study has found that *qhir1* carries a 4 bp insertion in the fourth exon of the *MATRILINEAL* (*MTL*)/ *NOT LIKE DAD* (*NLD*) / *ZmPHOSPHOLIPASE A1* (*ZmPLA1*) gene,

which encodes a patatin-like phospholipase expressed primarily in pollen (Gilles et al., 2017; Kelliher et al., 2017; Liu et al., 2017). It was not clear whether the Stock-6 based haploid induction is achieved by uni-parental genome elimination or through parthenogenesis, until the report by Kelliher et al. (2019), in which they used an inducer line carrying the CRISPR-Cas9 based genome-editing machinery and found that resulting haploid plants had corresponding edits, suggesting that a transient hybrid state precedes uni-parental chromosome elimination. The fact that the *MTL* gene is conserved across cereals, potentially provides a new opportunity to generate haploid inducer lines in various cereal crops by generating and testing novel mutations in this gene (Kelliher et al., 2017). As a proof of concept, mutations in rice *MATRILINEAL* gene, *OsMATL* (LOC_Os03g27610) have resulted in haploid inducer lines with an HIR of about 2 - 6 % (Yao et al., 2018). Similarly in wheat, an allohexaploid crop, triple-knockout of *ZmMTL* orthologous genes - *TaMTL-4A*, *TaMTL-4B*, and *TaMTL-4D* resulted in an HIR of 11.8 - 31.6 % (Liu et al., 2019). *qhir8*, the second major QTL responsible for haploid induction in Stock-6, has been fine mapped to a DUF679 membrane protein – *ZmDMP* (Zhong et al., 2019). Although knock-out of wild-type *ZmDMP* resulted in a very low HIR of about 0.15 %, its combination with *mtl*/*nld*/*zmpla1* mutation increased HIR by two to six-fold. This synergistic effect suggests that at least two distinct pathways play a role in Stock-6 based haploid induction (Jacquier et al., 2020). Unlike *MTL*, which is conserved only across monocots, *DMP-like* genes exist both in monocots and dicots and *Arabidopsis* mutants with loss-of-function of two *DMP-like* genes – *AtDMP8* and *AtDMP9* have shown an average HIR of 1 – 3.2 % (Zhong et al., 2020). Paternal haploids can also be induced in maize using lines with mutation in *ig1* (*indeterminate gametophyte*) gene (Evans, 2007), although low HIR of about 1 – 2 % hinders their use in commercial DH production (Kermicle, 1994).

Other two popular mechanisms of uni-parental genome elimination are inter-specific crosses and wide hybridization. In bulbosum method, which has been used extensively in barley breeding to produce doubled-haploids, the genome of wild *Hordeum bulbosum* gets eliminated when crossed to cultivated *H. vulgare* (Devaux and Kasha, 2008). Similarly, maize chromosomes get eliminated when crossed to wheat and this wide-hybridization technique is widely used to produce wheat doubled-haploids (Weyen, 2008). In such crosses, the rapidly dividing endosperm also suffers chromosomal elimination (mainly in monocots) and aborts early in seed development (Ishii et al., 2016). This demands that the developing haploid embryo must be rescued through *in vitro* culture which can be time-consuming (Forster et al., 2007).

More recently, uni-parental genome elimination in *Arabidopsis* has been achieved by modifying the centromeric histone H3 (CENH3), a histone variant which helps centromeres to bind with the spindle fibers (Ravi and Chan, 2010). With any modifications of CENH3, especially in the hypervariable histone fold domain, centromeres lose their ability to bind with spindle fibers and hence get lost during mitotic/meiotic cell divisions. When a parent with normal CENH3 is crossed to a parent with modified CENH3, this leads to genome elimination of the latter parent from zygotes, resulting in haploids from the former parent. It should be noted that the preferential spindle binding occurs only when both the centromeres – with normal CENH3 and with modified CENH3 are present in a cell, but not when competition is absent. This is the reason that parents with modified CENH3 are usually self-fertile. Unlike the *MTL* based haploid induction, which works through a mutated pollen-specific protein, this technology can be used for both androgenetic/paternal (modified CENH3 ♀ crossed to WT CENH3 ♂) and gynogenetic/maternal (WT CENH3 ♀ crossed to modified CENH3 ♂) haploid induction. In androgenetic haploid induction, the maternal genome with modified CENH-3 gets eliminated from the diploid zygote

resulting in haploid paternal genome surrounded by maternal cytoplasm. This phenomenon, knows as cytoplasm swapping, can be used for single-step establishment of cytoplasmic male sterility (CMS), a tool frequently used in hybrid seed production to avoid emasculation during crossing of inbred lines (Jacquier et al., 2020). As the genome from the parent with the modified CENH-3 gets eliminated, haploid plants generated through this technique are transgene free, although testing various CENH-3 modifications can be time-consuming. The function of CENH3 is conserved across a wide-variety of plant species (Britt and Kuppu, 2016), nevertheless, low haploid induction rates have become a bottleneck for its commercial exploitation. For example, in maize, complementing loss-of-function *cenh3*/*cenh3* mutants with a *tailswap-CENH3* resulted in average gynogenetic HIR of only 0.86 %, reaching a maximum of 3.6 % (Kelliher et al., 2016). In a recent study in maize, Wang et al. (2021) observed an androgenetic HIR of about 5 % in *CENH3*/*cenh3* heterozygous plants, which also avoided the vigor loss observed in *cenh3*/*cenh3* loss-of-function mutants. Efforts to engineer CENH3 based haploid induction in bread wheat, which is already an allohexaploid crop (with three subgenomes – A, B, and D), has been made more complex by the presence of two *CENH3* genes (α and *β*), making a total of 12 *CENH3* alleles. While retaining the six *CENH3β* alleles in wild-type state, a specific type of mutation in *CENH3α-A* (termed restored frameshift - *RFS*) coupled with knockout alleles for both *CENH3α-B* and *CENH3α-D*, resulted in HIR of about 7 - 8 % (Lv et al., 2020). Much like the difference of haploid induction rates between loss-of-function mutants complemented with *tailswap-CENH3* (Kelliher et al., 2016) and *CENH3*/*cenh3* heterozygous plants (Wang et al., 2021) in maize, Lv et al. (2020) also observed higher HIR when a *RFS* mutation was in a heterozygous state (*RFS*/*WT* – 8 %) as compared to when in homozygous state (*RFS*/*RFS* – 1.8 %). As all four methods described above (CENH3-mediated induction, haploid inducer lines, inter-specific hybridization and wide hybridization)

operate through double-fertilization followed by uni-parental genome elimination, these methods can also be used for one-step genome editing during haploid induction (Kelliher et al., 2019).

Parthenogenesis

During parthenogenesis, the egg cell initiates embryogenesis without being fertilized by a sperm cell. Natural apomicts produce chromosomally unreduced and genetically non-recombined egg cells through apomeiosis, which develop parthenogenetically to produce a progeny identical to the maternal parent. Introgression of parthenogenesis in sexually reproducing plants will produce haploid progeny as embryogenesis will be initiated from meiotically reduced egg cells. Parthenogenesis has not yet been used to produce haploid plants for commercial breeding programs primarily because of limited knowledge about the molecular mechanisms involved in this complex process.

Although direct introgression of apomixis (or its components) into crop plants by conventional breeding has not been successful (Ozias-Akins and van Dijk, 2007), the genetic transformation of crops plants with heterologous genes that control parthenogenesis in natural apomicts has shown some promising results. For example, *PsASGR-BBML*, the gene responsible for parthenogenesis in natural apomict *Pennisetum squamulatum*, can induce parthenogenesis in sexual pearl-millet when expressed under the control of its native promoter (Conner et al., 2015). Furthermore, both genomic and cDNA based *PsASGR-BBML* transgenes expressed under the control of native and an egg cell-specific promoter *AtDD45* induced parthenogenesis in rice and maize (Conner et al., 2017). *pAtDD45* is an egg cell-specific promoter, which was originally identified in *Arabidopsis* (Steffen et al., 2007) and later shown to be functional in rice egg cells as well (Ohnishi et al., 2014). *AtDD45*, also known as *EC1.2* (LOC_At2g21740), belongs to the *Arabidopsis EC1-like* gene family whose members are specifically expressed in the female

reproductive tissues (Sprunck et al., 2012). The dominant nature of *PsASGR-BBML*, which makes it a good candidate for the synthesis of apomixis in crop plants, limits its use for DH technology as these lines will again reduce to haploids. For the successful deployment of dominant parthenogenetic genes in commercial DH production pipelines, their expression needs to be regulated such that these are turned on in parental lines from which haploids are desired but off in the resulting DH progeny.

Gene regulation systems

There are two prominent ways of regulating the expression of a transgene – transcriptional regulation and post-translational regulation. The transcriptional regulation systems aim to control, through chemical induction, a transcriptional activator (TA) which, when activated will regulate the expression of gene(s) of interest (GOI). For this system to work, expression of the TA should bound the desired spatial and temporal range of GOI expression. A chemical ligand responsive TA is usually expressed under a strong constitutive promoter which upon activation with the chemical ligand will regulate the GOI expressed under a TA inducible promoter. In this way, the TA protein will be constitutively transcribed but will remain inactive until the time when GOI expression is desired, at which point, the chemical ligand can be applied which will activate the TA leading to the transcription of its downstream GOI. The GOI will need to be transcribed and then translated, which results in a time lag between the ligand application and the GOI's response, making this system less preferable when a rapid GOI response is required. An egg cell transcriptional activation in pearl millet was previously attempted in the lab without success. In a post-translational regulation system, the GOI protein is responsive to the ligand application. When and where the GOI expression is desired, the chemical ligand can be applied at least to that temporal and spatial range, which will activate the already translated GOI protein and hence regulate its downstream targets. As this system involves

no additional GOI protein synthesis, it produces a rapid response as compared to the transcriptional activation system. The Glucocorticoid Receptor system (Figure 2.1) has been widely used in plants both for transcriptional and post-translational regulation (Yamaguchi et al., 2015). Post-translational regulation system using the Glucocorticoid Receptor (GR) involves a fusion construct where the GOI is fused to the ligand binding domain (LBD) from the rat GR. Instead of the usual relocation of a TA to the nucleus after translation, the GOI:GR fusion protein forms a cytoplasmic complex with heat shock protein 90 (HSP-90), which is retained in the cytoplasm until the application of Dexamethasone (DEX), a synthetic steroid hormone. DEX binds to the GR domain and disrupts its interaction with HSP-90, allowing the GOI:GR fusion protein to relocate to the nucleus, where the GOI induces the expression of its downstream targets.

There have been several studies in different plant species which used a GR dependent post-translational activation system to induce transcriptional responses of various transcription factors such as in *Arabidopsis* – BnBBM (Passarinho et al., 2008), LEAFY (William et al., 2004), WUSCHEL (Leibfried et al., 2005) and in rice – OsMADS26 (Lee et al., 2008), WOX11 (Zhao et al., 2009) and WOX3 (Dai et al., 2007). *OsMADS26* is a *MADS-box* gene from rice whose over-expression induces stress response related phenotypes. Lee et al. (2008) were able to observe similar phenotypes in *ubi:OsMADS26* rice plants and DEX treated *ubi::OsMADS26:GR* plants (*ubi* – a constitute promoter), both at the seedling stage (growing on MS media supplemented with DEX) as well as in fully grown plants (80 d old plants – DEX solution added to soil). A GR dependent post-translational regulation system can also be used for the identification of transcriptional targets of a given Transcription Factor (TF) by comparing the expression profile of the control plants with that of plants treated with DEX to induce the TF:GR (Yamaguchi et al.,

2015). Moreover, Cyclohexamide (CYC) – a protein synthesis inhibitor, can be used to differentiate between the direct and indirect transcriptional targets of the TF.

For the successful regulation of parthenogenesis using a post-translational activation system, there are two main requirements – i) the expression of PsASGR-BBML:GR fusion protein in the egg cells and ii) the availability of DEX to the egg cells before fertilization for activating the PsASGR-BBML. The first condition can be met by using an egg cell-specific promoter such as *AtDD45*, which was originally identified in *Arabidopsis* (Steffen et al., 2007) and later shown to be functional in rice egg cells (Ohnishi et al., 2014). *AtDD45*, also known as *EC1.2* (LOC_At2g21740), belongs to the *Arabidopsis EC1-like* gene family whose members are specifically expressed in the female reproductive tissues (Sprunck et al., 2012). Conner et al. (2017) have shown that in *AtDD45::PsASGR-BBML* rice lines, PsASGR-BBML is expressed in unfertilized egg cells and can induce parthenogenesis. Building upon that, this study also utilized the *AtDD45* promoter to drive the egg cell-specific expression of PsASGR-BBML:GR fusion protein. For meeting the second condition, watering the soil-grown plants with DEX solution seems to be a simple and easy approach, although it might be hard for DEX to translocate all the way from roots to egg cells through the vascular system. While using a transcriptional regulation system, Samalova et al. (2005) used a DEX solution to water 8-week-old transgenic tobacco plants transformed with a *pOp6::GUS* reporter and *35S::LhG4-GR* activator construct. *LhG4-GR* encodes a fusion protein consisting of a ligand-binding domain of the rat GR gene fused to the N-terminus of a transcriptional activator, LhGR, which binds to the *pOp6* promoter and induces the *GUS* reporter gene. GUS activity was detected within 24 hours of the DEX treatment and was predominantly confined to the vascular tissues. Similarly, Craft et al. (2005) watered four-week old *Arabidopsis* plants transformed with the same activator construct but a different reporter construct

– *pH-TOP::GUS*. Within eight hours of DEX treatment, GUS activity was detected in rosette leaves which continued to increase up to 128 hours. After 24 hours of DEX treatment the staining pattern of these plants was comparable to that of *CaMV35S::GUS* plants, indicating that most tissues can receive DEX via the vascular system. In a similar study, Ouwerkerk et al. (2001) transformed rice plants with *OsGos2::GVG* activator and *pINDEX1::GUS* reporter constructs where *OsGos2* is a rice promoter driving constitutive expression while GVG is a fusion protein consisting of yeast Gal4 binding domain (G), the *Herpes simplex* VP16 activation domain (V) and the rat glucocorticoid receptor (G). They were able to observe GUS activity in flowers of the plants that were maintained hydroponically in 100 µM DEX. Although all these three studies are based on the transcriptional activation, they provide good evidence that DEX can be transported to various plant tissues by watering the soil-grown plants with DEX solution.

In addition to watering the plants with a DEX solution, we also employed *in planta* and *in vitro* floral DEX treatment. *In planta* treatment involves the use of a pipette tip to place the DEX solution inside the floret while in the *in vitro* treatment, florets are cultured on MS media supplemented with DEX. Brand et al. (2006) transformed *Arabidopsis* plants with a construct consisting of the *p35S::XVE* activator and *AtEASE::GUS* reporter units where XVE is a fusion protein consisting of the DNA-binding domain of the bacterial repressor LexA (X), VP16 activation domain (V) and the regulatory region of the human estrogen receptor (ER; E). The regulatory region of the human ER is similar to the LBD of rat GR, such that it forms a cytoplasmic complex which can be dissociated upon the application of the steroid hormone, 17-β-estradiol. The GUS reporter gene was driven by the *AtEASE* promoter which shows egg-apparatus specific expression in *Arabidopsis*. After the application of 17-β-estradiol to the exterior of flower buds,

GUS expression was observed in the egg apparatus, demonstrating the success of *in planta* floral treatments.

***PsASGR-BBML* and somatic embryogenesis**

Although *PsASGR-BBML* has been shown to induce parthenogenesis in various sexual species, its potential for inducing somatic embryogenesis hasn't been assessed yet. Using a constitutive promoter, *pZmUBIQUITIN1*, Khanday et al. (2019) have shown that *OsBBM1* has potential for inducing somatic embryogenesis. In a recent study, they were able to regulate *OsBBM1* induced somatic embryogenesis using a GR based post-translational regulation system, indicating that addition of the GR domain to *OsBBM1* doesn't suppress its potential for somatic embryogenesis (Khanday et al., 2020). In this study, we used a constitutive promoter from rice, *pOsCc1*, to drive the *PsASGR-BBML:GR* fusion construct to see if the application of DEX on vegetative tissues can promote somatic embryogenesis. *OsCc1* is the rice *Cytochrome c* gene (LOC_Os05g34770) which is constitutively expressed in rice, especially in non-photosynthetic tissues and developing embryos (Jang et al., 2002). The *pOsCc1* promoter has previously been utilized to constitutively overexpress *OsGS*, which encodes for glutathione synthetase and helps in detoxifying reactive oxygen species (ROS) by producing high levels of an oxidant – glutathione (Park et al., 2017). The *pOsCc1* promoter was also used to constitutively overexpress *AP37*, a member of the AP2 gene family, which improved rice grain yield under drought conditions (Oh et al., 2009). Apart from somatic embryogenesis, *pOsCc1::PsASGR-BBML:GR* lines can also be used to test the parthenogenetic potential of PsASGR-BBML when it is being driven by a constitutive promoter as compared to an egg cell-specific promoter, *pAtDD45*.

Using a glucocorticoid based post-translational regulation system, we have successfully shown that *PsASGR-BBML-GR* driven by an egg cell-specific promoter can be regulated at the

flowering stage to induce parthenogenesis. This technique presents a new opportunity to use parthenogenetic genes in DH production technology and to elucidate the molecular mechanism underlying parthenogenetic embryogenesis.

Materials and methods

Genetic material

Twenty-two transgenic rice lines (Nipponbare) were generated after transformation with *Agrobacterium tumefaciens* possessing *AtDD45::PsASGR-BBML:GR* transformation vector based on binary vector *pCAMBIA-1300* at the CALS Plant Transformation Facility, Cornell University (https://sips.cals.cornell.edu/research/plant-transformation-facility/). All lines had at least two plants. For the constitutive promoter, twenty-five transgenic rice lines were transformed with *OsCc1::PsASGR-BBML:GR* at the Iowa State University Transformation Facility (https://www.biotech.iastate.edu/ptf/) and each line had one to four plants. The two constructs used in this study - *AtDD45::PsASGR-BBML:GR* and *OsCc1::PsASGR-BBML:GR* are represented in Figure 2.2.

Growth conditions and fertilizers

Rice plants were grown in the greenhouse facility at NESPAL, Tifton, GA maintained at 24 – 29°C. Growing mix was composed of a 1:1 mixture of 'PROMIX BX-MYCORRHIZE soil' (Premier Tech Horticulture, Quakertown, PA) and 'Oil-Dri® Premium Absorbent' (Oil-Dri Corporation, Chicago, IL) supplemented with 3 g each of 'Everris Nursery Mix 18-5-12' and 'Encap Fast Acting Iron' (Encap, LLC, Green Bay, WI) were placed in a large plastic tub and were bottom-watered with low pH water (pH 5.2–5.8). Two-week old rice seedlings were fertilized with a 50 mL per pot solution of 3.96 g L^{-1} 'Miracle-Gro® Water Soluble All Purpose Plant Food' and

7.8 mL L^{-1} 'Ferti-Lome Chelated Liquid Iron and Other Micro Nutrients.' Mature rice plants were applied with same fertilizer mix as per requirement.

DNA extraction and genotyping

About 100 mg of young leaf tissue was ground in 200 µL of 2X CTAB buffer (100 mM Tris-Cl (pH = 8), 20 mM EDTA, 1.4 M NaCl, 2 % CTAB, and 0.2 % β-mercaptoethanol) in 1.5 mL eppendorf tubes with an electric mortar-pestle. After grinding, an additional 300 µL CTAB buffer was added, vortexed and incubated at 65°C in a water bath for at least 15 minutes. 500 µL of 24:1 Chloroform:Iso-amyl alcohol was added, mixed and followed by centrifugation for 10 minutes at 14,000 rpm. 350 µL of supernatant was transferred to a new eppendorf tube and an equal volume of iso-propanol was added and mixed followed by centrifugation for 10 minutes at 14,000 rpm. The supernatant was discarded, and the DNA pellet was washed with 70 % ethanol, air dried for 10 minutes and re-suspended in 75 µL 0.5X Tris-EDTA pH 8.0 buffer along with 0.1 µL RNase A (working conc. 10 mg mL^{-1}) per sample. Genotyping for *PsASGR-BBML* was performed using primers 1792/1801 [(Conner et al., 2017), sequences presented in Figure 2.2] in a 25 µL PCR reaction consisting of 2 µL DNA, 1X PrimeSTAR GXL Buffer, 200 µM dNTP, 0.2 µM primers, and 0.625 U PrimeSTAR GXL DNA Polymerase (Takara Bio Inc., CA, USA). PCR conditions included an initial denaturation step for 1 min 30 s at 98°C followed by 35 cycles of 98°C for 10 s, 60°C for 30s, and 68°C for 4 min and 30 s except for the last elongation step at 68°C for 7 min.

Transgene expression analysis

For examining *PsASGR-BBML:GR* expression in unfertilized egg cells from *AtDD45::PsASGR-BBML:GR* lines, total RNA was isolated from rice ovaries one day before anthesis using the Qiagen RNeasy Plant Mini Kit. 1 µg of isolated RNA was used for expression

analysis by performing RT-PCR using SuperScript® III First-Strand Synthesis System from Invitrogen. Similar analysis was done in the *OsCc1::PsASGR-BBML:GR* lines, although this will only indicate *PsASGR-BBML:GR* expression in pre-anthesis ovaries as a whole and not specifically in unfertilized egg cells. *PsASGR-BBML:GR* expression in young leaves of *OsCc1::PsASGR-BBML:GR* lines was also determined through a similar procedure.

Flow cytometry

Five samples containing a mixture of five seeds from 21 T_0 lines were analyzed through Bulk Seed Flow Cytometry (BSFC) according to methods described in Conner et al. (2017). Line 13B could not be evaluated due to a lack of seeds. Ploidy level of T_1 plants was determined by flow cytometry on leaf tissues bulked from at most five plants separately from each line.

Dexamethasone (DEX) treatments

A 100 mM stock solution was prepared by dissolving 1 g of Dexamethasone (Sigma-Aldrich) in 25.48 mL of dimethyl sulfoxide (DMSO) and was stored in aliquots at -20° C. *PsASGR-BBML:GR* transgene positive plants were treated with 20 µM, 50 µM or 100 µM DEX solution for 10 days beginning at the booting stage for the 1st panicle of the plant and ending at the time when the first flush of flowering was complete (all florets passed anthesis). DEX working solutions were prepared by diluting the stock solution in low pH water (pH 5.2–5.8) which was then used to bottom-water plants in plastic tubs as needed (Figure 2.3). Transgene positive plants treated with 0.05 % and 0.1 % DMSO and transgene negative plants treated with all three concentrations of DEX served as negative controls. Wild type plants were also treated with all three concentrations of DEX and 0.05 % and 0.1 % DMSO to assess any effect of these treatments on plant development and seed set.

***In planta* floral DEX treatment**

This involved the use of a pipette tip to place the DEX solution inside the floret (Figure 2.3).

1. **Emasculated florets**

All florets from one panicle of a given T_1 plant were emasculated and then treated with 5 µL of 100 µM DEX for three consecutive days. Florets from another panicle from the same plant were emasculated and treated with 0.1 % DMSO in a similar way. Ovaries were collected and fixed overnight in FAA (Formaldehyde Alcohol Acetic Acid – 47.5 % ethanol, 3.7 % formaldehyde, 5 % acetic acid) on the fifth day after emasculation. Fixed ovaries were dehydrated successively in 70 %, 85 % and 100 % ethanol each for two hours and then transferred to 100 % ethanol overnight. Dehydrated ovaries were cleared successively in 2:1 ethanol:methyl salicylate (MS), 1:2 ethanol:MS, pure MS each for two hours and then transferred to pure MS overnight. Cleared ovaries were then observed under microscope.

2. **Un-emasculated florets**

Floret tops from one panicle of a given T_1 plant were cut without disturbing the anthers and then treated with 5 µL of 100 µM DEX for three consecutive days. Floret tops from another panicle from same plant were cut in a similar way followed by 0.1 % DMSO treatment. Upon maturity these seeds were collected and then analyzed with BSFC.

***In vitro* floral DEX treatment**

This involved placing the florets on MS (Murashige & Skoog) media (4.44 g/L of Murashige & Skoog Basal Medium with Vitamins from PhytoTech Labs, 4 g/L Gelzan and 2 mL/L PPM™ from Plant Cell Technology) along with 2 % sucrose (20 g/L). The method is shown in Figure 2.3., and both emasculated and non-emasculated florets were analyzed.

Results

Genotyping of T_0 plants

As shown in the Table 2.1, 49 out of 53 T_0 plants were positive for *AtDD45::PsASGR-BBML:GR* when genotyped with primers 1792/1801. All lines had at-least one transgene positive plant except line 16, where both plants were transgene negative.

cDNA sequencing to verify the coding sequence

In order to verify the coding sequence of the *PsASGR-BBML:GR* transcript being expressed by *AtDD45::PsASGR-BBML:GR* lines, *PsASGR-BBML:GR* cDNA from two random T_0 plants – 7B and 13B were sequenced and compared with the transformation vector sequence. No sequence or splicing variants of the transcript were identified.

Genotyping and expression analysis in the T_1 generation

Flow cytometry on T_1 seeds was done in order to identify any haploid signal, which would indicate unwanted activity of PsASGR-BBML:GR without any application of the inducer steroid. A majority of the lines showed the expected diploid signal. Line 13 B could not be analyzed due to low seed set, line 14B didn't flower and lines 3A, 3B and 12B displayed a triploid genome size. Based on *PsASGR-BBML:GR* expression analysis in pre-anthesis T_0 ovaries, 10 lines were selected and at least 10 T_1 seeds were sown from each to obtain T_1 progeny.

The number of germinated T_1 seeds and genotyping results for T_1 progeny are presented in Table 2.2. Out of the 83 T_1 plants germinated, 66 were transgene positive and flow cytometry on leaf tissue from these T_1 plants showed diploid signal, confirming that PsASGR-BBML:GR is not active without application of the inducer steroid.

For the *PsASGR-BBML:GR* expression analysis in unfertilized ovaries from *AtDD45::PsASGR-BBML:GR* T_1 plants, pre-anthesis ovaries were collected, followed by RNA

extraction and RT-PCR. Figure 2.4 shows *PsASGR-BBML:GR* expression with primer pair 1792/1801. All ovary samples tested from transgene positive parents showed transgene expression, while 4C-13 which was from a transgene negative plant, didn't show any expression in T_2 ovaries.

DEX through watering treatment in the T_1 generation

Three transgene positive T_1 plants from each T_0 line were treated with 100 µM, 50 µM or 20 µM DEX through watering at the boot stage, along with wild-type plants (n = 6) treated with all three concentrations of DEX, transgene positive plants treated with only 0.1 % DMSO (equivalent of DMSO concentration in 100 µM DEX solution, n = 2) and 0.05 % DMSO (equivalent of DMSO concentration in 50 µM DEX solution, n = 2) and some untreated transgene positive plants (n = 5) serving as negative controls. Bulk seed flow cytometry was performed on mature T_2 seeds and the results, along with seed set of these T_1 plants are presented in the Table 2.3. Seven out of 10 transgene positive plants showed haploid signal in BSFC indicating that 100 µM 'DEX through watering' treatment can induce parthenogenesis. Similarly, for 50 µM and 20 µM 'DEX through watering' treatments respectively, four and two plants out of 10 showed haploid signal. None of the wild-type plants treated with any concentration of DEX, transgene positive plants treated with two different concentrations of DMSO or untreated transgene positive plants showed haploid signal in BSFC, indicating that both the *AtDD45::PsASGR-BBML:GR* transgene and inducer steroid are required for successful induction of parthenogenesis. Wild-type plants treated with all three concentrations of DEX and with two concentrations of DMSO had normal seed set (average seed-set from two WT plants: 100 µM – 90.6 %, 50 µM – 68.9 %, 20 µM – 93.4 %, 0.1 % DMSO – 95.3 % and 0.05 % DMSO – 95.9 %; calculated as the number of mature seed over total florets analyzed), which indicated that DEX or DMSO treatment through watering was not greatly affecting seed set. Most transgene positive lines treated with DEX or DMSO had a

moderate level of seed set (average seed-set from 10 transgene positive plants: 100 µM – 33.2 %, 50 µM – 20.9 %, 20 µM – 36.7 %, 0.1 % DMSO – 42.7 % and 0.05 % DMSO – 25.7 %) which indicated that the seed set of plants didn't vary with different concentrations of DEX or DMSO treatments and the overall reduction in seed set observed in these lines as compared to WT plants is possibly because of their transgenic origin. Five T_1 plants, 6B-1, 6B-3, 7B-2, 13B-1 and 13B-2 had no seed set, which might indicate some sterility inducing mutation segregating in these lines, which may be due to their transgenic origin.

Genotyping and DEX through watering treatment in theT_2 generation

Out of 10 T_1 plants analyzed with 100 µM DEX through watering, four lines (4C, 10B, 15B and 22B) were selected as the best lines as these showed BSFC haploid signals in T_2 seeds in 100 µM as well as 50 µM DEX through watering treatments. 10 seeds each from four T_1 plants from these lines 4C-1, 15B-4 (100 µM DEX through watering treated) and 10B-3, 22B-2 (50 µM DEX through watering treated) were sown to produce the T_2 generation. T_2 plants were genotyped as well as analyzed with flow cytometry and the results are presented in Table 2.4. None of the four T_2 progenies had all plants transgene positive showing that these four T_1 plants were heterozygous for the transgene. Out of 32 T_2 plants, two were haploid which is low as compared to BSFC results on T_2 seeds from these T_1 plants. Nine T_2 plants (seven transgene positive and two transgene negative) were treated with 100 µM DEX at the boot stage followed by Individual Seed Flow Cytometry (ISFC) on mature seeds. As shown in the Table 2.5, all the transgene positive plants treated with DEX showed haploid signal in ISFC, while transgene positive plants untreated with DEX or transgene negative plants treated with DEX didn't show haploid signal. Most of the lines had good seed set although it couldn't be correlated to transgene or DEX treatment.

Identifying homozygous lines

As the T_2 progenies obtained from all four T_1 plants (4C-1, 15B-4, 10B-3 and 22B-2) had both transgene positive and transgene negative plants, these T_1 plants were possibly heterozygous for the transgene, and the transgene positive plants in their T_2 progenies were either homozygous or heterozygous for *AtDD45::PsASGR-BBML:GR*. The true rate of parthenogenesis and seed set can only be determined by DEX treatment on homozygous plants. In order to find such plants, about 30 T_2 seeds were germinated from four different T_1 plants (4C – 4C-5/7/11/14, 10B – 10B-1/5/11/14, 15B – 15B-2/3/12/13 and 22B – 22B-1/4/11/13). The T_2 progenies where all the plants were transgene positive, were considered homozygous for *AtDD45::PsASGR-BBML:GR* along with their parent T_1 plant. As shown in the Table 2.6, the 10B line had three homozygous T_2 progenies resulting from homozygous T_1 plants 10B-5, 10B-11 and 10B-14 while 4C, 15B and 22B lines each had one homozygous T_2 progeny resulting from homozygous T_1 plants 4C-14, 15B-3 and 22B-13 respectively.

Five of the T_1 plants (4C-11, 10B-11, 15B-12, 15B-13 and 22B-13) were also used for *in planta* floral DEX treatment. Seeds from the untreated panicles from these plants were collected to produce T_2 progenies. The identification of haploid T_2 seedlings would suggest that DEX was able to travel from treated panicles to untreated panicles through the plant vascular system. None of the T_2 progenies from untreated panicles had haploid plants, suggesting that DEX doesn't translocate within the plant when applied directly to floral tissue or the translocation happened after the appropriate reproductive stage to trigger parthenogenesis had passed.

Six of the T_1 plants were treated with DEX through watering (100 µM – 10B-14, 50 µM – 4C-5, 15B-2 and 20 µM – 4C-7, 10B-1 and 22B-1) of which only 10B-14 was identified as homozygous. Its T_2 progeny consisted of 13 haploids, nine haploid twins, and 1 haploid+diploid

twin out of 25 plants analyzed. T_2 progenies from 10B-1 and 15B-2 each had one haploid plant and two haploid+diploid twins. 15B-3, another homozygous T_1 plant which was treated with 0.1 % DMSO through watering had no haploid plants in its T_2 progeny. T_2 progenies from 15B line (from T_1 plants – 15B-3, 15B-12 and 15B-13) had albino/variegated plants in about 1:4 ratio to normal plants.

***In planta* floral DEX treatment in T_1 lines**

1. **Un-emasculated florets**

As presented in Table 2.7, seed set was very low for these panicles which resulted due to the mere delivery of liquid in the florets as it was low both in DEX as well as DMSO treated panicles. Although seed set was low, DEX treated panicles showed haploid signal in BSFC while no haploid signal was observed in 0.1 % DMSO treated or untreated panicles.

2. **Emasculated florets**

Results are presented in Figure 2.5 and Figure 2.6. for two T_1 lines – 10B-11 and 15B-12. Fixed and cleared sample from line 10B-11 had 28 ovaries out of which 17 showed clear parthenogenesis i.e. appearance of a parthenogenetic embryo along with the presence of intact polar nuclei. For line 15B-12, which had 15 ovaries in total, 3 showed clear parthenogenesis. We also observed some ovary samples using a confocal microscope and one parthenogenetic ovary from plant 15B-12 is shown in Figure 2.7.

***In vitro* floral DEX treatment in T_1 lines**

As shown in the results in Table 2.8, we observed good seed set in the un-emasculated *in vitro* floral DEX treatment when a portion of the inflorescence was placed on media in magenta boxes. Seeds from four out of five lines maintained on DEX positive media showed haploid signal in BSFC while one line maintained on DEX negative media didn't show any haploid signal.

Emasculated florets fixed and cleared after culturing on DEX positive media were also observed under the microscope, but the ovary structure was degenerated and we unable to get any clear indication of parthenogenesis.

***OsCc1::PsASGR-BBML:GR* T_0 lines**

Genotyping results of *OsCc1::PsASGR-BBML:GR* T_0 lines are shown in Table 2.9. Line 9B had a rearranged version of the transgene with a truncated GR domain, and the cDNA sequence showed that it contains a pre-mature stop codon such that the GR region is not expressed. Some *OsCc1::PsASGR-BBML:GR* plants were treated with 100 µM DEX starting at a young plant stage while others were treated at the boot stage. Young plants treated with DEX didn't show signs of somatic embryogenesis and preliminary BSFC on T_1 seeds from transgene positive T_0 plants treated with DEX at the boot stage did not show any haploid signal. Germination experiments of T_1 seeds on media supplemented with DEX, in order to see somatic embryogenesis, are currently being performed.

Discussion

DH technology has proven to be a valuable asset to modern hybrid breeding programs due to its ability to generate inbred lines in a considerably shorter period of time as compared to traditional methods. Although methods for genome doubling of haploid plants are well established (at least in cereals), stable haploid induction techniques with high efficiency are less common. Four methods of *in vivo* haploid induction have been developed which include – CENH3-mediated induction, haploid inducer lines, inter-specific hybridization, and wide hybridization

CENH-3 mediated haploid induction has been engineered in maize [*cenh3*/*cenh3* complemented with *AcGREEN-tailswap-CENH3*, maximum gynogenetic HIR – 3.6 %, maximum androgenetic HIR – 2.4 % (Kelliher et al., 2016); *CENH3*/*cenh3*, maximum gynogenetic HIR –

0.5 %, maximum androgenetic HIR – 5.2 % (Wang et al., 2021)], wheat [maximum androgenetic HIR – 8 % (Lv et al., 2020)], rice [maximum HIR – 1 % (Op Den Camp et al., 2017)] and tomato [maximum HIR – 2.3 % (Op Den Camp et al., 2017)], but low haploid induction rates and aneuploidy observed in offspring have limited the commercial exploitation of this technology. Moreover, this technique requires crossing, which will be more difficult in autogamous species such as rice and wheat (Jacquier et al., 2020).

Since the advent of DH technology, the Stock-6 based haploid inducer lines have been used for commercial DH production in maize (Chaikam et al., 2019). Since the first report in 1959 (Coe, 1959), 60 years of intensive breeding has increased HIR of stock-6 from 3 % to about 15 % (Uliana Trentin et al., 2020), and this has become a threshold which any new haploid induction technology needs to surpass in order to be adopted in a commercial setting in maize breeding. Recent identification of the gene responsible for haploid induction, *MTL*/*NLD*/*ZmPLA* (Gilles et al., 2017; Kelliher et al., 2017; Liu et al., 2017) has provided researchers with an opportunity to tweak the underlying genetic mechanisms in order to achieve higher HIR. *MTL* (or similar genes) can be utilized for haploid induction in two ways – i) crossing *mtl* carrying haploid inducer lines as males with the parental lines from which haploids are desired or ii) mutating endogenous *MTL* in parental lines. Inducer lines have been generated by mutating *MTL* in maize [*de novo MTL* mutation in a non-inducer line – *mtl*/*mtl*, maximum HIR – 12.5 %, average HIR – 6.7 % (Kelliher et al., 2017)] and rice [*mtl*/*mtl*, HIR – about 3.2 % (Yao et al., 2018)]. When inducing mutations in endogenous *MTL* (or similar genes) from parental lines, two approaches can be used to avoid the resulting DH lines reducing back to haploids – i) using *mtl* mutation in a heterozygous state or ii) regulating the *mtl* mutation. In the first approach, haploids could be obtained from seeds where wild-type *MTL* carrying egg cells were fertilized with *mtl* pollen (25 % of all seeds in case of complete penetrance

of *mtl*). These haploid plants will be free of *mtl* mutation and can be doubled to get stable DH plants. Based on pollen tube growth, pollen viability of *Osmatl*-edited rice plants has been found to be similar to that of wild-type plants (Yao et al., 2018). A second way to avoid DH reduction is to regulate the endogenous *MTL* in the parental lines, which can be achieved by attaching regulatory constructs with *MTL* through gene editing. *MTL* can be turned off in parental lines to induce haploids, and then turned on in resulting DH lines for preventing these to reduce back to haploids. In self-pollinated *mtl*/*mtl* rice plants, Yao et al. (2018) observed an HIR of about 6 % with about 20 % seed-set while Wang et al. (2019) observed an HIR of 3.63 % with about 11.5 % seed-set. Haploids can also be induced by crossing distantly related species such as inter-specific cross (bulbosum method in barley) and wide-hybridization (wheat X maize crosses). Molecular analysis has shown that these methods are based on uni-parental genome elimination (Kelliher et al., 2019; Sanei et al., 2011).

With the implementation of parthenogenetic genes, a new haploid induction technology is possible which involves direct embryogenesis from unfertilized egg cells rather than uni-parental genome elimination after crossing. For parthenogenetic haploid induction in crops plants, proof-of-concept has been shown by using a parthenogenetic gene, *PsASGR-BBML*, cloned from a natural apomict (Conner et al., 2015; Conner et al., 2017) and by egg-cell specific expression of an endogenous *BBM-like* gene (Khanday et al., 2019). As parthenogenesis operates through egg cells, parental lines from which haploids are desired must express the parthenogenetic genes in unfertilized egg cells, which will require transformation or recurrent back-crossing. Recent interest in using the spontaneous haploid genome doubling to further shorten the DH production cycle (Boerman et al., 2020) can make transformation approaches, such as parthenogenesis, more compelling as compared to crossing approaches, such as *MTL* or CENH-3 mediated haploid

induction. Nevertheless, the dominant nature of *PsASGR-BBML* limits its use for DH technology as these lines will again reduce to haploids. For successful deployment of dominant parthenogenetic genes in commercial DH production pipelines, their expression needs to be regulated such that these genes are turned on in parental lines from which haploids are desired but kept off in the resulting DH progeny.

For regulating the GOI expression using a ligand dependent post-translational regulation system, both the expression of GOI:GR fusion protein as well as the availability of ligand must occur in the desired spatial and temporal frame. Using the *AtDD45* promoter, an egg cell-specific promoter from *Arabidopsis* (Steffen et al., 2007) which has been shown to be functional in rice egg cells as well (Ohnishi et al., 2014), we ensured that the PsASGR-BBML:GR fusion protein was expressed in rice egg cells. This was also confirmed with the *PsASGR-BBML:GR* expression analysis though RT-PCR on RNA extracted from pre-anthesis rice ovaries in *AtDD45::PsASGR-BBML:GR* lines. Although there are many studies with leaf tissues as the desired location of GOI expression (Dai et al., 2007; Lee et al., 2008; Zhao et al., 2009) studies on floral tissues are relatively fewer (Ouwerkerk et al., 2001). To our knowledge, there are no reports of post-translational regulation of gametophytic tissue specific expression by watering the plants with a chemical ligand solution. In this study, *AtDD45::PsASGR-BBML:GR* plants were watered with DEX solution starting at the boot stage, and the successful induction of parthenogenesis has provided strong evidence that DEX can translocate from roots to egg cells through the vascular tissues of the plant. DEX treatment through watering is an easy method and its simplicity indicates that it has high potential for use in large scale commercial DH development programs. We have shown that DEX can also be applied through *in planta* and *in vitro* floral treatments. Although these treatments are practically more difficult than watering the plants with DEX solution, the

proximity of application site to the egg cell might have provided an edge, had the translocation of DEX through the plant been more difficult.

A self-pollinated line, homozygous for *AtDD45::PsASGR-BBML:GR* (10B-14 T_1 plant), treated with 100 µM DEX at the flowering stage, showed an HIR of 92 % and seed set of 17.2 %. Although analysis in more homozygous lines will provide a true estimate of HIR and seed-set, this data shows the potential of using a post-translational regulation system for utilizing parthenogenetic genes in DH production. We have observed no adverse effect of 100 µM DEX on rice plants, but the treatment can be optimized for concentration and timing along with testing additional DEX-like steroids, as recent studies have reported more efficient steroids in rice (Samalova et al., 2019).

Dominant gene(s) responsible for spontaneous haploid genome doubling (SHGD), if identified, would also need to be regulated in the same way as dominant parthenogenetic genes (turned on in parental lines producing haploids and off in DH lines). Successful post-translational regulation of PsASGR-BBML achieved in this study suggests that parthenogenetic and SHGD genes can be delivered using a single construct and can be co-regulated for developing a single step DH production system. Precision genome editing tools such as CRISPR-Cas9 (Wada et al., 2020) can be used to insert the *PsASGR-BBML:GR* construct at the desired location in the genome in a bi-allelic fashion which is expected to show a higher rate of parthenogenesis as all the ovaries will possess the transgene.

Understanding the molecular pathways involved in PsASGR-BBML induced parthenogenesis will provide more options for increasing the efficiency of the system both for DH production and for synthesis of apomixis in crop plants. Although apomixis has been synthetically engineered using different approaches, our understanding of underlying molecular mechanisms is

still limited (Khanday and Sundaresan, 2021). The successful post-translational regulation of PsASGR-BBML reported in this study can be used for identifying downstream target genes by using differential expression analysis in DEX treated vs untreated *AtDD45::PsASGR-GR* plants (Yamaguchi et al., 2015). Moreover, direct transcriptional targets of PsASGR-BBML can be identified by the simultaneous application of cycloheximide, an inhibitor of protein biosynthesis. As *BBM-like* genes are known to be involved in auto-activation [*BnBBM1* (Horstman et al., 2015), *OsBBM1* (Khanday et al., 2019)], PsASGR-BBML might be inducing parthenogenesis by directly activating the egg cell *OsBBM1*, which in sexual reproduction is activated by the paternally expressed *OsBBM1* to initiate zygotic embryogenesis (Khanday et al., 2019). Chromatin Immunoprecipitation (ChIP) coupled with quantitative PCR can also be used for investigating the transcriptional targets of PsASGR-BBML. Protein pull-down by using a GR-specific antibody will also determine if PsASGR-BBML is part of a larger protein complex. A recent study by Khanday et al. (2020) used a combination of Glucocorticoid based post-transcriptional regulation and ChIP-seq to determine that *OsBBM1* induces somatic embryogenesis by directly upregulating auxin-biosynthesis genes. In this study, we could not identify *PsASGR-BBML* induced somatic embryogenesis in mature plants; however, additional experiments will be performed to verify this preliminary result.

Although apomixis hasn't been utilized in practical crop breeding yet, when it happens it will be highly desirable to use the same parthenogenetic factor both for inducing haploids from parental stock as well as for preserving the hybrid vigor across generations. In order to achieve this, fully penetrant and completely regulated apomeiotic and parthenogenetic systems are required with the following regulation approach – to induce haploids from heterozygous parental lines, apomeiosis – off, parthenogenesis – on; to maintain DH lines, apomeiosis – on, parthenogenesis –

on; crossing two DH lines to produce hybrid seed or to develop new heterozygous parental stock, apomeiosis – off, parthenogenesis – off; to achieve clonal reproduction in hybrids, apomeiosis – on, parthenogenesis – on. Apomeiosis and parthenogenesis should be kept on by default so that no regulation is necessary by the farmer during the hybrid cultivation and can be kept on or turned off as per requirement at any step of the breeding process. Synthetic apomeiotic systems such as *MiMe* (d'Erfurth et al., 2009; Mieulet et al., 2016), although highly penetrant, are based on multiple gene knockouts and will be difficult to regulate as compared to a single apomeiosis factor, if identified, from natural apomicts. Unless autonomous endosperm development can be achieved, aposporous apomeiotic systems (producing panicum type embryo sacs) would be desirable for maintaining the 2:1 maternal to paternal genome ratio in endosperm as these will produce 2n central cells independent of whether apomeiosis is turned on (one nucleus – 2n) or off (two nuclei – n+n) as compared to diplosporous apomeiotic systems which will produce 2n central cells when apomeiosis is off (two nuclei – n+n) and 4n when it is on (two nuclei – 2n+2n). By successfully regulating the *PsASGR-BBML* induced parthenogenesis, this study shows that apomictic factors can be successfully regulated to achieve such kinds of systems.

References

Boerman, N.A., Frei, U.K., Lübberstedt, T. (2020) Impact of Spontaneous Haploid Genome Doubling in Maize Breeding. Plants, **9**(3), 369.

Bohanec, B. (2009) Doubled Haploids via Gynogenesis. In Advances in Haploid Production in Higher Plants (A. Touraev, B.P. Forster & S.M. Jain, eds), Springer Netherlands, Dordrecht: pp 35-46.

Brand, L., Hörler, M., Nüesch, E., Vassalli, S., Barrell, P., Yang, W., Jefferson, R.A., Grossniklaus, U., Curtis, M.D. (2006) A versatile and reliable two-component system for tissue-specific gene induction in Arabidopsis. Plant Physiol, **141**(4), 1194-1204.

Britt, A.B. and Kuppu, S. (2016) Cenh3: An emerging player in haploid induction technology, Frontiers Media S.A.

Chaikam, V., Molenaar, W., Melchinger, A.E., Boddupalli, P.M. (2019) Doubled haploid technology for line development in maize: technical advances and prospects, Springer Verlag: pp 3227-3243.

Coe, E.H. (1959) A Line of Maize with High Haploid Frequency. The American Naturalist, **93**(873), 381-382.

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

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

Craft, J., Samalova, M., Baroux, C., Townley, H., Martinez, A., Jepson, I., Tsiantis, M., Moore, I. (2005) New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. The Plant Journal, **41**(6), 899-918.

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., Mercier, R. (2009) Turning Meiosis into Mitosis. PLOS Biology, **7**(6), e1000124-e1000124.

Dai, M., Hu, Y., Zhao, Y., Liu, H., Zhou, D.-X. (2007) A WUSCHEL-LIKE HOMEOBOX Gene Represses a YABBY Gene Expression Required for Rice Leaf Development. Plant Physiology, **144**(1), 380-380.

Devaux, P. and Kasha, K.J. (2008) Overview of Barley Doubled Haploid Production, Springer Netherlands: pp 47-63.

Dunwell, J.M. (2010) Haploids in flowering plants: Origins and exploitation. Plant Biotechnology Journal, **8**(4), 377-424.

Evans, M.M.S. (2007) The *indeterminate gametophyte1* Gene of Maize Encodes a LOB Domain Protein Required for Embryo Sac and Leaf Development. The Plant Cell, **19**(1), 46-62.

Ferrie, A.M.R. and Möllers, C. (2011) Haploids and doubled haploids in Brassica spp. for genetic and genomic research, Springer: pp 375-386.

Forster, B.P., Heberle-Bors, E., Kasha, K.J., Touraev, A. (2007) The resurgence of haploids in higher plants, Elsevier Current Trends: pp 368-375.

Gilles, L.M., Khaled, A., Laffaire, J.B., Chaignon, S., Gendrot, G., Laplaige, J., Bergès, H., Beydon, G., Bayle, V., Barret, P., Comadran, J., Martinant, J.P., Rogowsky, P.M., Widiez, T. (2017) Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. The EMBO Journal, **36**(6), 707-717.

Guha, S. and Maheshwari, S.C. (1964) In vitro Production of Embryos from Anthers of Datura. Nature, **204**(4957), 497-497.

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

Ishii, T., Karimi-Ashtiyani, R., Houben, A. (2016) Haploidization via Chromosome Elimination: Means and Mechanisms. Annual Review of Plant Biology, **67**(1), 421-438.

Jacquier, N.M.A., Gilles, L.M., Pyott, D.E., Martinant, J.-P., Rogowsky, P.M., Widiez, T. (2020) Puzzling out plant reproduction by haploid induction for innovations in plant breeding. Nature Plants, **6**(6), 610-619.

Jang, I.-C., Choi, W.-B., Lee, K.-H., Song, S.I., Nahm, B.H., Kim, J.-K. (2002) High-Level and Ubiquitous Expression of the Rice Cytochrome c Gene OsCc1 and Its Promoter Activity in Transgenic Plants Provides a Useful Promoter for Transgenesis of Monocots. Plant Physiology, **129**(4), 1473-1473.

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M.L., Green, J., Chen, Z., McCuiston, J., Wang, W., Liebler, T., Bullock, P., Martin, B. (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature, **542**(7639), 105-109.

Kelliher, T., Starr, D., Su, X., Tang, G., Chen, Z., Carter, J., Wittich, P.E., Dong, S., Green, J., Burch, E., McCuiston, J., Gu, W., Sun, Y., Strebe, T., Roberts, J., Bate, N.J., Que, Q. (2019) One-step genome editing of elite crop germplasm during haploid induction, Nature Publishing Group: pp 287-292.

Kelliher, T., Starr, D., Wang, W., McCuiston, J., Zhong, H., Nuccio, M.L., Martin, B. (2016) Maternal Haploids Are Preferentially Induced by CENH3-tailswap Transgenic Complementation in Maize. Front. Plant Sci, **7**, 414-414.

Kermicle, J.L. (1994) Indeterminate Gametophyte (ig): Biology and Use. In The Maize Handbook (M. Freeling & V. Walbot, eds), Springer New York, New York, NY: pp 388-393.

Khanday, I., Santos-Medellín, C., Sundaresan, V. (2020) Rice embryogenic trigger BABY BOOM1 promotes somatic embryogenesis by upregulation of auxin biosynthesis genes. bioRxiv, 2020.2008.2024.265025.

Khanday, I., Skinner, D., Yang, B., Mercier, R., Sundaresan, V. (2019) A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. Nature, **565**(7737), 91-95.

Khanday, I. and Sundaresan, V. (2021) Plant zygote development: recent insights and applications to clonal seeds. Current Opinion in Plant Biology, **59**, 101993.

Lee, S., Woo, Y.-M., Ryu, S.-I., Shin, Y.-D., Kim, W.T., Park, K.Y., Lee, I.-J., An, G. (2008) Further characterization of a rice AGL12 group MADS-box gene, OsMADS26. Plant Physiology, **147**(1), 156-168.

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

Liu, C., Li, X., Meng, D., Zhong, Y., Chen, C., Dong, X., Xu, X., Chen, B., Li, W., Li, L., Tian, X., Zhao, H., Song, W., Luo, H., Zhang, Q., Lai, J., Jin, W., Yan, J., Chen, S. (2017) A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize, Cell Press: pp 520-522.

Liu, H., Wang, K., Jia, Z., Gong, Q., Lin, Z., Du, L., Pei, X., Ye, X. (2019) Efficient induction of haploid plants in wheat by editing of TaMTL using an optimized Agrobacterium-mediated CRISPR system. Journal of Experimental Botany, **71**(4), 1337-1349.

Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., Wang, Y., Zhou, H., Carlin, R., Rich, R., Lu, T., Que, Q., Wang, W.C., Zhang, X., Kelliher, T. (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. Nature Biotechnology, **38**(12), 1397-1401.

Mieulet, D., Jolivet, S., Rivard, M., Cromer, L., Vernet, A., Mayonove, P., Pereira, L., Droc, G., Courtois, B., Guiderdoni, E., Mercier, R. (2016) Turning rice meiosis into mitosis. Cell Research, **26**, 1242-1242.

Mishra, R. and Rao, G.J.N. (2016) In-vitro Androgenesis in Rice: Advantages, Constraints and Future Prospects. Rice Science, **23**(2), 57-68.

Murovec, J. and Bohanec, B. (2012) Haploids and Doubled Haploids in Plant Breeding, InTech.

Oh, S.-J., Kim, Y.S., Kwon, C.-W., Park, H.K., Jeong, J.S., Kim, J.-K. (2009) Overexpression of the Transcription Factor AP37 in Rice Improves Grain Yield under Drought Conditions. Plant Physiology, **150**(3), 1368-1368.

Ohnishi, Y., Hoshino, R., Okamoto, T. (2014) Dynamics of Male and Female Chromatin during Karyogamy in Rice Zygotes. Plant Physiology, **165**(4), 1533-1543.

Op Den Camp, R., Van Dijk, P., A, G. (2017) Method for the production of haploid and subsequent doubled haploid plants.

Ouwerkerk, P.B., de Kam, R.J., Hoge, H.J., Meijer, A.H. (2001) Glucocorticoid-inducible gene expression in rice. Planta, **213**(3), 370-378.

Ozias-Akins, P. and van Dijk, P.J. (2007) Mendelian Genetics of Apomixis in Plants. Annual Review of Genetics, **41**(1), 509-537.

Park, S.-I., Kim, Y.-S., Kim, J.-J., Mok, J.-E., Kim, Y.-H., Park, H.-M., Kim, I.-S., Yoon, H.-S. (2017) Improved stress tolerance and productivity in transgenic rice plants constitutively expressing the Oryza sativa glutathione synthetase OsGS under paddy field conditions. Journal of Plant Physiology, **215**, 39-47.

Passarinho, P., Ketelaar, T., Xing, M., van Arkel, J., Maliepaard, C., Hendriks, M.W., 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**(3), 225-237.

Prigge, V., Xu, X., Li, L., Babu, R., Chen, S., Atlin, G.N., Melchinger, A.E. (2012) New Insights into the Genetics of in Vivo Induction of Maternal Haploids, the Backbone of Doubled Haploid Technology in Maize. Genetics, **190**(2), 781-793.

Ravi, M. and Chan, S.W.L. (2010) Haploid plants produced by centromere-mediated genome elimination. Nature, **464**(7288), 615-618.

Samalova, M., Brzobohaty, B., Moore, I. (2005) pOp6/LhGR: A stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. Plant Journal, **41**(6), 919-935.

Samalova, M., Kirchhelle, C., Moore, I. (2019) Universal Methods for Transgene Induction Using the Dexamethasone-Inducible Transcription Activation System pOp6/LhGR in Arabidopsis and Other Plant Species. Current Protocols in Plant Biology, **4**(2), e20089.

Sanei, M., Pickering, R., Kumke, K., Nasuda, S., Houben, A. (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proceedings of the National Academy of Sciences, **108**(33), E498-E505.

Sprunck, S., Rademacher, S., Vogler, F., Gheyselinck, J., Grossniklaus, U., Dresselhaus, T. (2012) Egg Cell–Secreted EC1 Triggers Sperm Cell Activation During Double Fertilization. Science, **338**(6110), 1093-1093.

Steffen, J.G., Kang, I.-H., Macfarlane, J., Drews, G.N. (2007) Identification of genes expressed in the Arabidopsis female gametophyte. **51**(2), 281-292.

Uliana Trentin, H., Frei, U.K., Lübberstedt, T. (2020) Breeding Maize Maternal Haploid Inducers. Plants, **9**(5), 614.

Veilleux, R.E. (2010) Haploid Plant Production, Pollen, Anther, and Ovule Culture, John Wiley & Sons, Inc., Hoboken, NJ, USA: pp 1-6.

Wada, N., Ueta, R., Osakabe, Y., Osakabe, K. (2020) Precision genome editing in plants: state-of-the-art in CRISPR/Cas9-based genome engineering. BMC Plant Biology, **20**(1), 234.

Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., Wu, M., Sun, T., Cheng, Z., Mercier, R., Wang, K. (2019) Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. Nature Biotechnology, **37**(3), 283-286.

Wang, N., Gent, J.I., Dawe, R.K. (2021) Haploid induction by a maize cenh3 null mutant. Science Advances, **7**(4), eabe2299.

Weyen, J. (2008) Barley and Wheat Doubled Haploids in Breeding, Springer Netherlands: pp 179-187.

William, D.A., Su, Y., Smith, M.R., Lu, M., Baldwin, D.A., Wagner, D. (2004) Genomic identification of direct target genes of LEAFY. Proc Natl Acad Sci USA, **101**(6), 1775-1780.

Yamaguchi, N., Winter, C.M., Wellmer, F., Wagner, D. (2015) Identification of direct targets of plant transcription factors using the GR fusion technique. Methods in molecular biology (Clifton, N.J.), **1284**, 123-138.

Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., Liu, J., Sahoo, G., Kelliher, T. (2018) OsMATL mutation induces haploid seed formation in indica rice. Nature Plants, **4**(8), 530-533.

Zhao, Y., Hu, Y., Dai, M., Huang, L., Zhou, D.-X. (2009) The WUSCHEL-related homeobox gene WOX11 is required to activate shoot-borne crown root development in rice. The Plant Cell, **21**(3), 736-748.

Zhong, Y., Chen, B., Li, M., Wang, D., Jiao, Y., Qi, X., Wang, M., Liu, Z., Chen, C., Wang, Y., Chen, M., Li, J., Xiao, Z., Cheng, D., Liu, W., Boutilier, K., Liu, C., Chen, S. (2020) A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous Arabidopsis. Nature Plants, **6**(5), 466-472.

Zhong, Y., Liu, C., Qi, X., Jiao, Y., Wang, D., Wang, Y., Liu, Z., Chen, C., Chen, B., Tian, X., Li, J., Chen, M., Dong, X., Xu, X., Li, L., Li, W., Liu, W., Jin, W., Lai, J., Chen, S. (2019) Mutation of ZmDMP enhances haploid induction in maize. Nature Plants, **5**(6), 575-580.

Tables

Table 2.1 Genotyping data from *AtDD45::PsASGR-BBML:GR* T_0 lines.

Twenty-two *AtDD45::PsASGR-BBML:GR* T_0 lines, with at least two plants in each line, were genotyped for *PsASGR-BBML:GR* using primer pair 1792/1801. '+' transgene positive plant, '–' transgene negative plant. Died – plant died before reaching the flowering stage. n/a – a third plantlet in the line was not recovered during the transformation process.

Line	A	B	C
1	+	+	n/a
2	+	+	+
3	+	+	n/a
4	+	+	+
5	+	+	+
6	+	+	+
7	+	+	+
8	+	+	+
9	+	+	+
10	-	+	-
11	+	+	+
12	+	+	+
13	+	+	n/a
14	Died	+	n/a
15	+	+	n/a
16	-	-	n/a
17	+	+	+
18	+	+	n/a
19	+	+	n/a
20	+	+	n/a
21	Died	+	n/a

Table 2.2 *AtDD45::PsASGR-BBML:GR* T_1 lines – germination and genotyping data.

Ten seed from ten T_0 lines (column one) were sown, and the number of germinated seedlings is shown in column 2. The third column represents the number of *AtDD45::PsASGR-BBML:GR* transgene positive plants. Genotyping was done by the 1792/1801 primer pair.

T_0 Line	T_1 seeds germinated out of total sown	Transgene positive out of germinated
4C	12/15	7/12
6B	6/10	4/6
7B	7/10	5/7
8B	7/10	7/7
9B	8/10	7/8
10B	9/15	8/9
11B	10/10	7/10
13B	8/10	7/8
15B	8/15	7/8
22B	8/15	7/8

Table 2.3 BSFC & seed set data from DEX through watering treatment of transgene positive *AtDD45::PsASGR-BBML:GR* T_1 plants.

Bulk Seed Flow Cytometry results on T_2 seed from 'DEX through watering' treated transgene positive T_1 plants. Transgene positive T_1 plants treated with 0.05 % DMSO (9B-4, 13B-5, WT-7 and WT-8), 0.1 % DMSO (8B-4, 15B-3, WT-9 and WT-10), and untreated plants were used as negative controls. Wild type plants are denoted by WT and were used as negative controls in 100 µM (WT-1, WT-2), 50 µM (WT-3, WT-4) and 20 µM (WT-5, WT-6) DEX through watering treatments.

100 µM			50 µM			20 µM			DMSO			Untreated		
T_1 plant	BSFC	Seed set	T_1 plant	BSFC	Seed set	T_1 plant	BSFC	Seed set	T_1 plant	BSFC	Seed set	T_1 plant	BSFC	Seed set
4C-1	3/4	32.7 %	4C-5	4/4	34.9 %	4C-7	0/5	51.9 %						
6B-2	4/4	26.8 %	6B-3	-	0.0 %	6B-1	-	0.0 %				6B-6	0/3	11.0 %
7B-5	0/4	37.7 %	7B-2	-	0.0 %	7B-1	0/3	19.6 %						
8B-3	0/4	23.6 %	8B-2	0/4	18.8 %	8B-1	0/5	51.1 %	8B-4	0/5	39.5 %			
9B-3	4/4	45.0 %	9B-2	0/3	13.9 %	9B-1	1/5	53.5 %	9B-4	0/5	7.6 %	9B-6	0/4	77.4 %
10B-4	4/4	27.5 %	10B-3	3/4	28.1 %	10B-1	3/5	43.5 %				10B-5	0/4	54.4 %
11B-3	0/4	43.8 %	11B-2	0/4	50.5 %	11B-1	0/5	93.1 %						
13B-3	3/4	45.5 %	13B-2	-	0.0 %	13B-1	-	0.0 %	13B-5	0/5	43.8 %	13B-6	0/4	74.5 %
15B-4	3/4	30.7 %	15B-2	4/4	36.9 %	15B-1	0/5	41.8 %	15B-3	0/4	45.9 %			
22B-3	4/4	18.7 %	22B-2	4/4	25.7 %	22B-1	0/1	12.0 %				22B-4	0/4	42.7 %
WT-1	0/4	92.7 %	WT-3	0/4	78.6 %	WT-5	0/4	91.7 %	WT-7	0/4	96.8 %			
WT-2	0/4	88.5 %	WT-4	0/4	59.1 %	WT-6	0/4	95.0 %	WT-8	0/4	95.0 %			
									WT-9	0/4	98.8 %			
									WT-10	0/4	91.8 %			

Table 2.4 *AtDD45::PsASGR-BBML:GR* T_2 progeny genotyping results.

10 seeds from four T_1 plants – 4C-1, 15B-4 (100 µM DEX through watering treated) and 10B-3, 22B-2 (50 µM DEX through watering treated) were sown to get T_2 generation. The genotyping results are presented in the table below. Plants are labelled A through I and grey color indicates haploid plants.

T_1 plant (DEX through watering treatment)				
Plant label.	4C-1 (100 µM)	10B-3 (50 µM)	15B-4 (100 µM)	22B-2 (50 µM)
A	+	–	–	+
B	–	–	–	+
C	+	+	+	–
D	–	+	+	+
E	–	+	–	+
F	+	+	+	–
G	–	+		+
H	+	+		+
I		+		+
Transgene Positive Plants	4/8	7/9	3/6	7/9

Table 2.5 Individual Seed Flow Cytometry (ISFC) on T_3 seeds.

ISFC on T_3 seeds from DEX treated or untreated T_2 plants. DEX treated T_2 plants are marked with the grey color while DEX untreated are unmarked. '+' sign represents that the given T_2 plant was positive for the *AtDD45::PsASGR-BBML:GR* plants while '–' sign represents that it was negative. For DEX treatment 100 μM DEX was applied through watering starting at booting stage and terminated when flowering was over.

T_2 plant (genotype)		Haploid ISFC seed signal out of total seeds analyzed	Seed set
DEX	No DEX		
4C-1A			
4C-1A +		14/44	46.69 %
4C-1B –		0/25	48.01 %
4C-1F +		0/25	67.56 %
4C-1H +		0/25	58.80 %
10B-3			
10B-3E +		19/45	44.82 %
10B-3H +		10/45	42.45 %
10B-3C +		0/25	51.11 %
10B-3D +		0/25	66.40 %
15B-4			
15B-4E –		0/25	24.82 %
15B-4C –		0/25	10.08 %
15B-4F +		0/25	38.86 %
22B-2			
22B-2A +		5/12	6.31 %
22B-2G +		8/35	23.87 %
22B-2H +		3/8	3.36 %
22B-2I +		10/27	17.94 %
22B-2B +		0/25	21.05 %
22B-2D +		0/25	16.00 %

Table 2.6 Identifying homozygous plants - genotyping and flow cytometry results from T_2 progenies.

About 30 T_2 seeds were germinated from four different T_1 plants from each of the four best lines. Treatment applied to each T1 plant is shown in the second column. Grey color in the last column marks homozygous plants. Number in parenthesis in third column shows the number of albino/variegated plants observed in the T_2 progenies from three 15B plants. Last column, T – *AtDD45::PsASGR-BBML:GR*.

T_1 plant	Treatment	T_2 plants/30 germinated (albino)	Diploid	Twin (D+H)	Haploid	Twin (H+H)	Transgene +/-	Genotype of T_1 plant
4C-5	50 µM DEX through watering	27	25	-	2	-	9/16	Tt
4C-7	20 µM DEX through watering	30	30	-	-	-	13/7	Tt
4C-11	*In planta*/*in vitro* floral DEX	27	24	-	-	-	22/5	Tt
4C-14	*In vitro* floral DEX	28	22	-	-	-	28/0	TT
10B-1	20 µM DEX through watering	27	24	2	1	-	21/6	Tt
10B-5	-	29	29	-	-	-	29/0	TT
10B-11	*In planta*/*in vitro* floral DEX	28	28	-	-	-	28/0	TT
10B-14	100 µM DEX through watering	25	2	1	13	9	23/0	TT
15B-2	50 µM DEX through watering	28	25	2	1	-	22/6	Tt
15B-3	0.1 % DMSO through watering	28 (8)	18	-	-	-	25/0	TT
15B-12	*In planta*/*in vitro* floral DEX	28 (9)	17	-	-	-	20/5	Tt
15B-13	*In planta* floral DEX	27 (10)	17	-	-	-	18/7	Tt
22B-1	20 µM DEX through watering	25	25	-	-	-	17/8	Tt
22B-4	-	29	29	-	-	-	22/7	Tt
22B-11	-	29	24	-	-	-	20/9	Tt
22B-13	*In planta*/*in vitro* floral DEX	24	24	-	-	-	24/0	TT

Table 2.7 BSFC results from *in planta* floral DEX treatment on un-emasculated florets from *AtDD45::PsASGR-BBML:GR* T_1 lines.

Bulk Seed Flow Cytometry results on seeds from *in planta* floral DEX treated un-emasculated panicles is shown in the table.

Treatment →	100 µM DEX		0.1 % DMSO		Untreated	
Plant	Seed set	BSFC	Seed set	BSFC	Seed set	BSFC
4C-11	14/75	0/3	10/59	0/2	89.4 %	0/4
4C-13	14/37	0/3			85.7 %	
9B-13	2/80	1/1			77.8 %	
10B-11	4/70	1/1	11/45	0/3	90.7 %	0/4
10B-13	9/60	2/2			83.5 %	0/4
15B-12	2/50	1/1			38.2 %	0/4
22B-13	4/49	0/1			64.9 %	

Table 2.8 Bulk Seed Flow Cytometry (BSFC) results from *in vitro* floral DEX treatment on un-emasculated florets from *AtDD45::PsASGR-BBML:GR* T_1 lines.

Treatment →	100 μM DEX		0.1 % DMSO	
Plant	Enlarged	BSFC	Enlarged	BSFC
10B-12	32/48	5/6	15/20	0/3
15B-13	24/41	5/5		
15B-14	35/51	6/7		
22B-11	12/36	2/3		
22B-12	25/51	0/5		

Table 2.9 *OsCc1::PsASGR-BBML:GR* lines – genotyping, DEX through watering and *in vitro* floral DEX treatment.

Table 2.9.1 shows genotyping results of *OsCc1::PsASGR-BBML:GR* lines. '+' and '-' denote transgene positive and transgene negative plants, n/a - no plant. Table 2.9.2 shows 'DEX through watering treatment' and Table 2.9.3 shows '*in vitro* 100 µM DEX treatment'.

Table 2.9.1				
Line	A	B	C	D
2	n/a	+	n/a	n/a
3	+	+	n/a	n/a
4	-	-	-	n/a
5	+	+	n/a	n/a
6	+	+	n/a	n/a
7	+	+	n/a	n/a
8	+	+	n/a	n/a
9	+	+*	n/a	n/a
10	+	+	+	n/a
11	+	+	n/a	n/a
12	+	+	n/a	n/a
13	+	+	n/a	n/a
14	+	+	n/a	n/a
15	+	+	+	n/a
16	+	+	+	+
17	+	+	n/a	n/a
18	+	+	n/a	n/a
19	+	+	n/a	n/a
20	+	+	n/a	n/a
21	+	+	n/a	n/a
22	+	+	+	n/a
23	+	n/a	n/a	n/a
24	+	+	n/a	n/a
25	+	+	n/a	n/a
26	+	+	n/a	n/a
*	Rearranged transgene			

Table 2.9.2				
Line	A	B	C	D
2	n/a	+	n/a	n/a
3	+	+	n/a	n/a
4	-	-	-	n/a
5	+	+	n/a	n/a
6	+	+	n/a	n/a
7	+	+	n/a	n/a
8	+	+	n/a	n/a
9	+	+*	n/a	n/a
10	+	+	+	n/a
11	+	+	n/a	n/a
12	+	+	n/a	n/a
13	+	+	n/a	n/a
14	+	+	n/a	n/a
15	+	+	+	n/a
16	+	+	+	+
17	+	+	n/a	n/a
18	+	+	n/a	n/a
19	+	+	n/a	n/a
20	+	+	n/a	n/a
21	+	+	n/a	n/a
22	+	+	+	n/a
23	+	n/a	n/a	n/a
24	+	+	n/a	n/a
25	+	+	n/a	n/a
26	+	+	n/a	n/a
	0.1 % DMSO at booting			
	100 µM DEX at booting			
	100 µM DEX young plant			

Table 2.9.3				
Line	A	B	C	D
2	n/a	+	n/a	n/a
3	+	+	n/a	n/a
4	-	-	-	n/a
5	+	+	n/a	n/a
6	+	+	n/a	n/a
7	+	+	n/a	n/a
8	+	+	n/a	n/a
9	+	+*	n/a	n/a
10	+	+	+	n/a
11	+	+	n/a	n/a
12	+	+	n/a	n/a
13	+	+	n/a	n/a
14	+	+	n/a	n/a
15	+	+	+	n/a
16	+	+	+	+
17	+	+	n/a	n/a
18	+	+	n/a	n/a
19	+	+	n/a	n/a
20	+	+	n/a	n/a
21	+	+	n/a	n/a
22	+	+	+	n/a
23	+	n/a	n/a	n/a
24	+	+	n/a	n/a
25	+	+	n/a	n/a
26	+	+	n/a	n/a
	In vitro floral DEX			

Figures

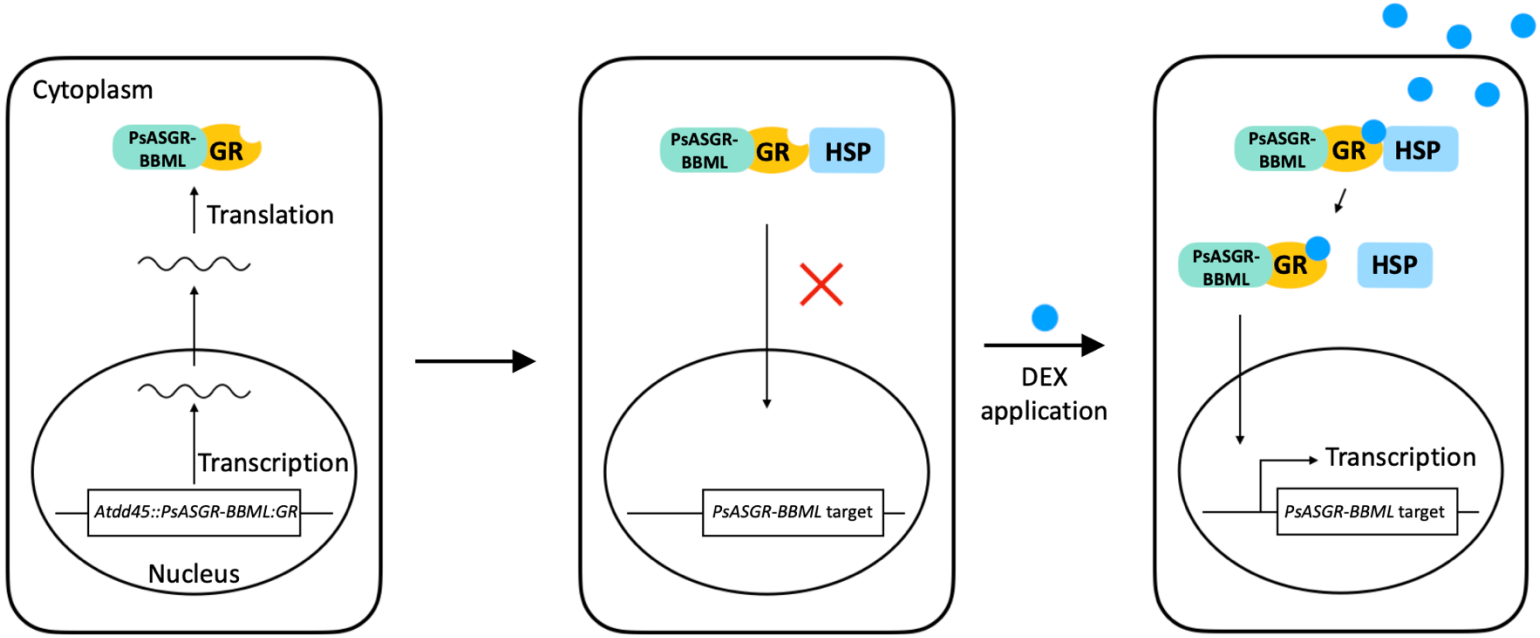


Figure 2.1 The Glucocorticoid Receptor (GR) – DEX system for post-translational regulation.

This system involves a fusion construct where *PsASGR-BBML* is fused to the ligand binding domain (LBD) from the rat GR. Instead of the usual relocation to the nucleus after translation, the PsASGR-BBML:GR fusion protein forms a cytoplasmic complex with heat shock protein 90 (HSP-90), which is retained in the cytoplasm until the application of Dexamethasone (DEX), a synthetic steroid hormone. DEX binds to the GR domain and disrupts its interaction with HSP-90, allowing the PsASGR-BBML:GR fusion protein to relocate to the nucleus and to induce the expression of its downstream targets. (Modified from Yamaguchi et al., 2015)

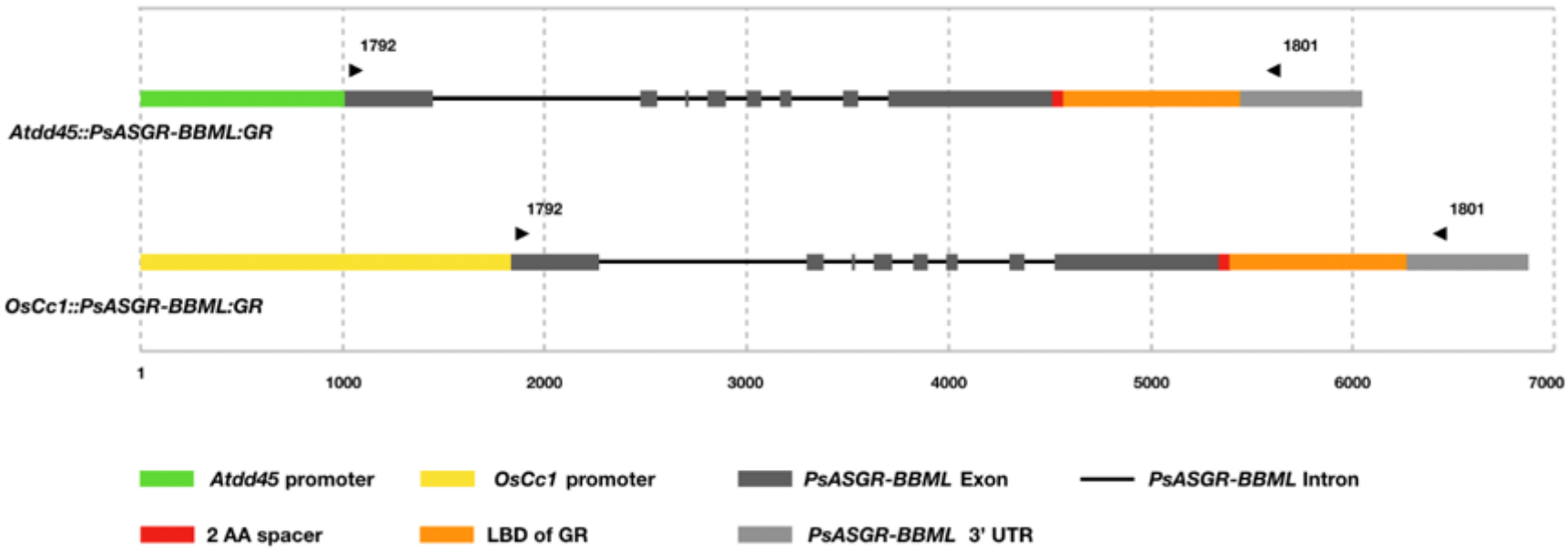


Figure 2.2 Visual representation of *AtDD45::PsASGR-BBML:GR* and *OsCc1::PsASGR-BBML:GR* transgene cassettes.

Green rectangle denotes *dd45* promoter from *Arabidopsis* and yellow rectangle denotes *OsCc1* promoter from rice. Gray rectangles denote the *PsASGR-BBML* exons with introns denoted as black lines. Red rectangle denotes two amino acid spacer between last (8^{th}) exon of *PsASGR-BBML* and Ligand binding domain (LBD) of GR which is represented by orange rectangle. Light grey rectangle denotes the native *PsASGR-BBML* 3' UTR. Primers sequences (5' to 3') 1792 – TTCCACCAACAACTGGCTGCGCT and 1801 – TTCCACCAACAACTGGCTGCGCT.

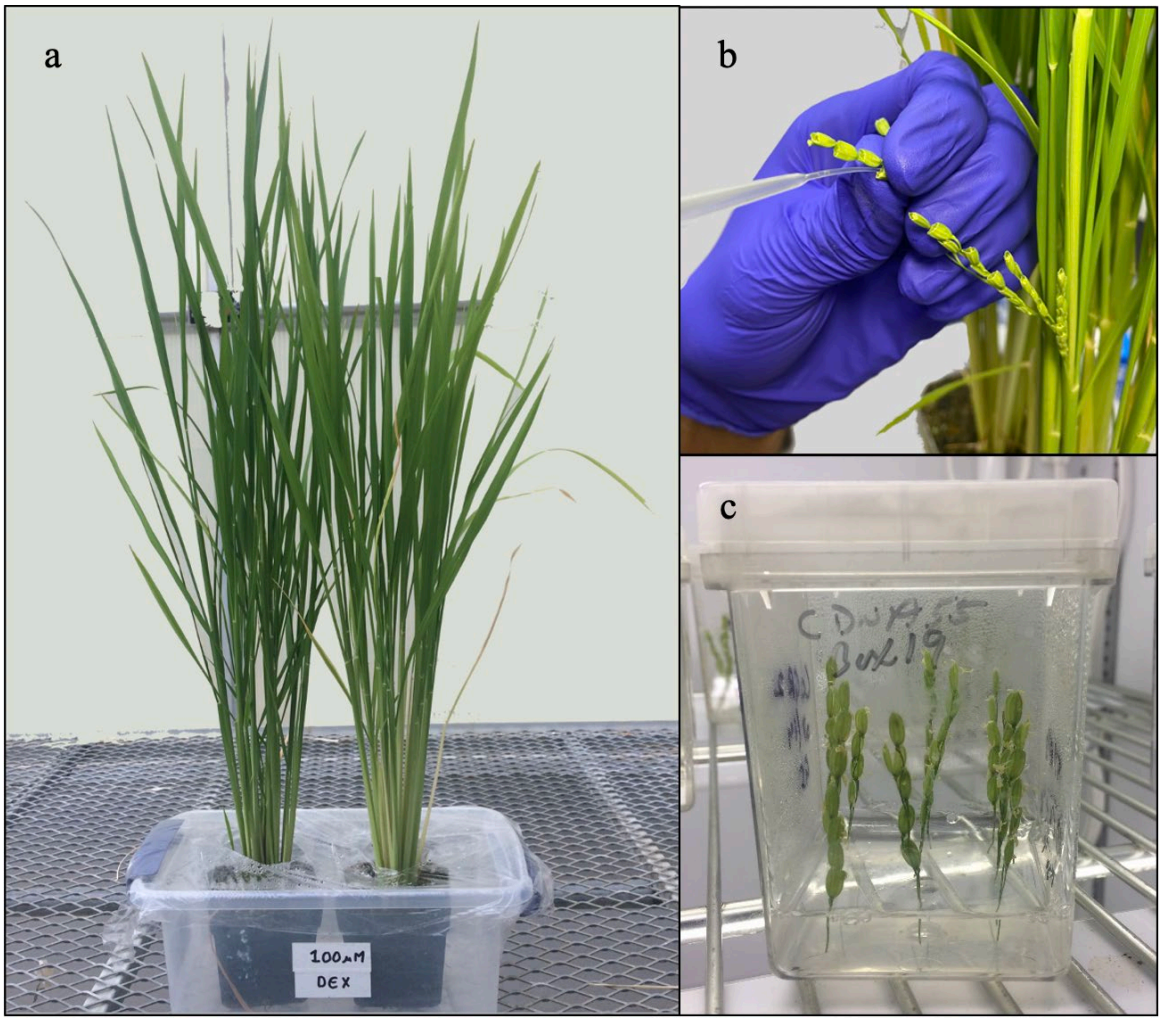


Figure 2.3 Various types of DEX treatments.

Three types of DEX treatments provided to *AtDD45::PsASGR-BBML:GR* T1 plants

DEX through watering (a), in planta floral DEX treatment (b) and in vitro floral DEX treatment (c).

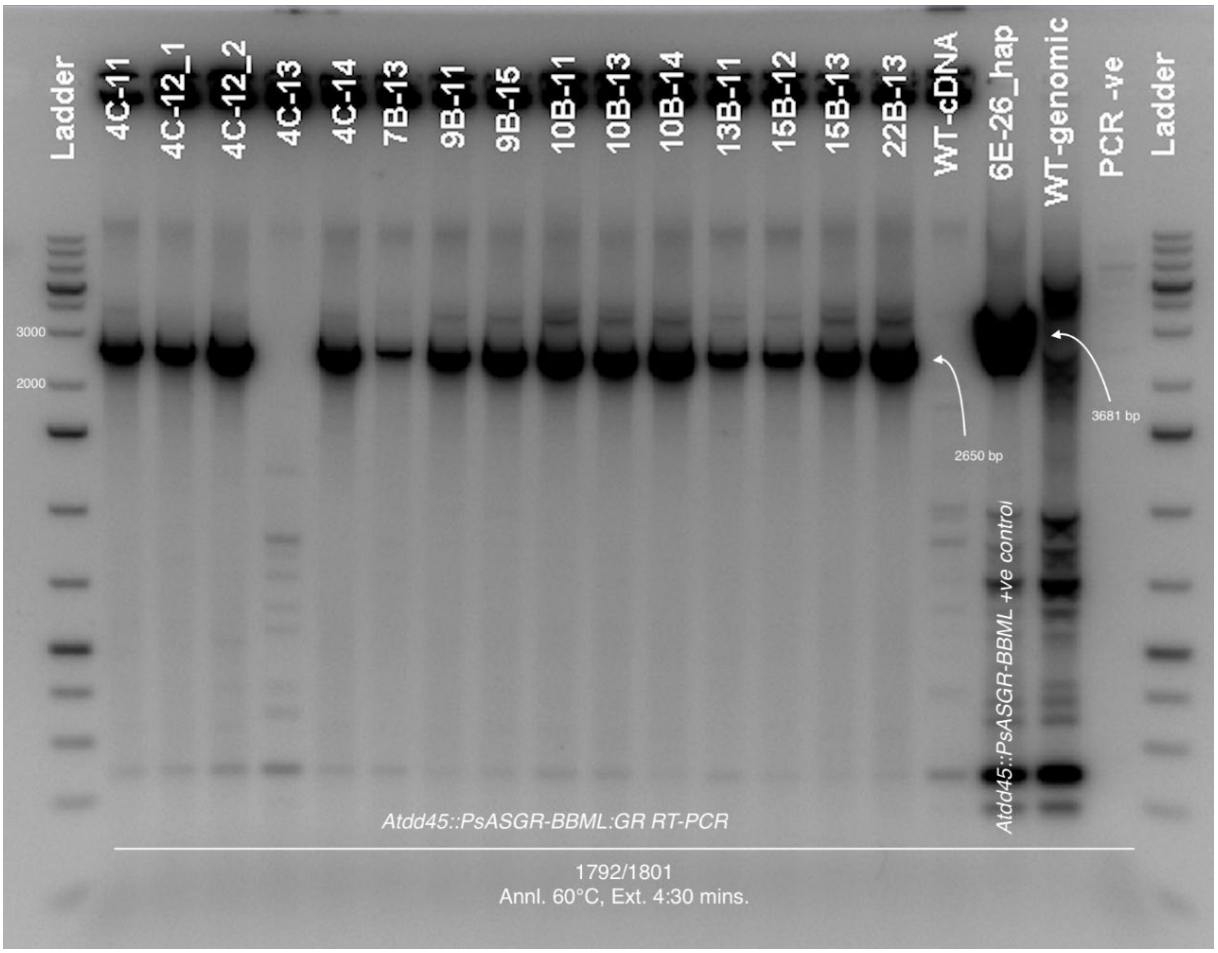


Figure 2.4 PsASGR-BBML:GR expression analysis in pre-anthesis ovaries from *AtDD45::PsASGR-BBML:GR* T_1 plants.

Pre-anthesis ovaries were collected from *AtDD45::PsASGR-BBML:GR* T_1 plants, followed by RNA extraction and RT-PCR. Figure shows the expression of *PsASGR-BBML:GR* with primer pair 1792/1801. All samples were from transgenic positive parents except 4C-13 which was from transgene negative plant. *PsASGR-BBML:GR* cDNA amplified with 1792/1801 gives a band of size 2650 bp. Genomic DNA from R448 T_3 haploid plant 6E-26 (PsASGR-BBML transgene amplicon with 1792/1801 size – 3681 bp) was used as positive control. RNA extracted from pre-anthesis ovaries of a wild-type plant was used as negative control.

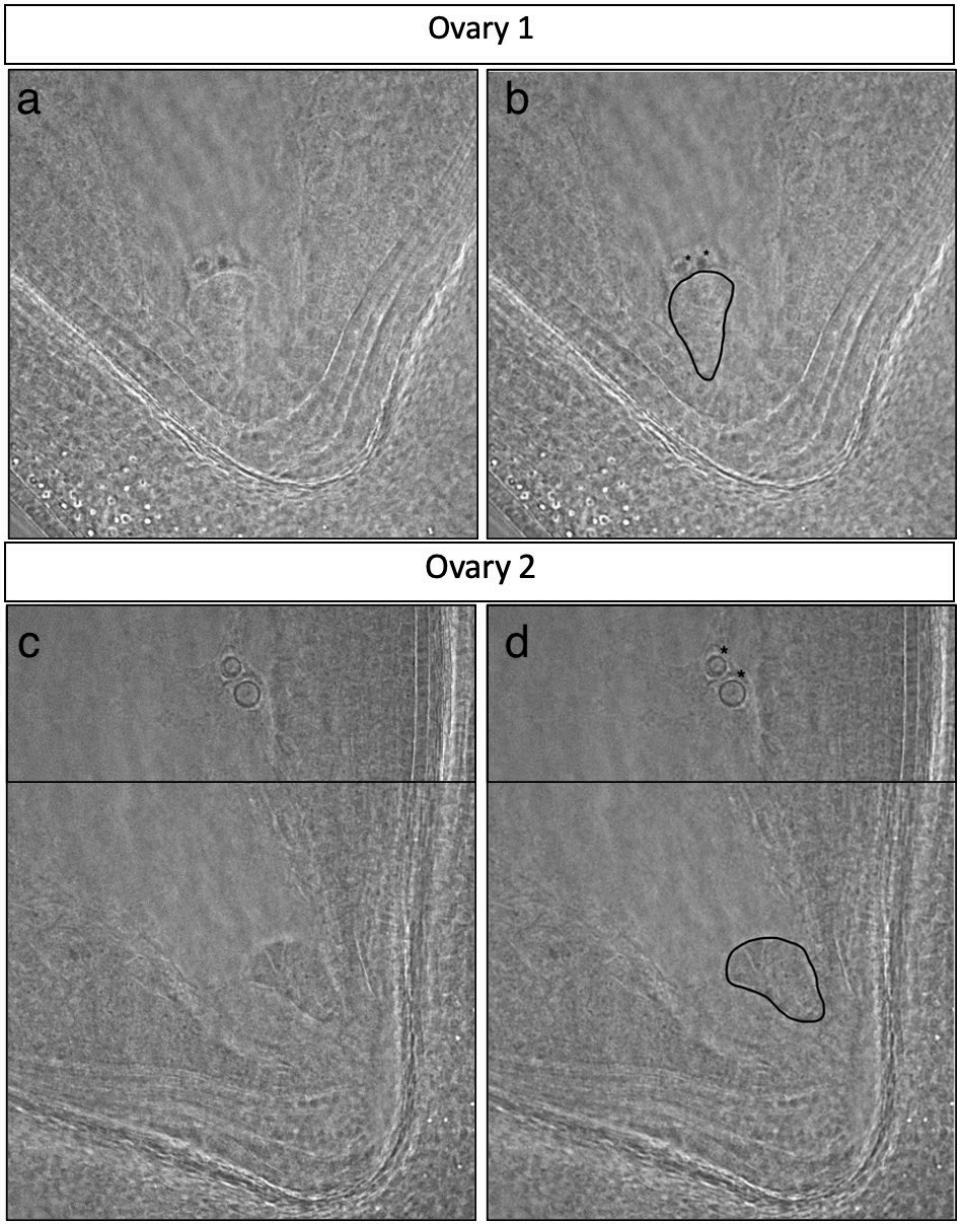


Figure 2.5 Parthenogenesis in *in planta* DEX treated emasculated florets from T_1 10B-11.

Emasculated ovaries from T_1 plant 10B-11 showed parthenogenetic embryo formation (panel a and c) in two different ovaries. The embryo is outlined for distinction (panel b and d). The presence of intact polar nuclei (marked with asterisks in b and d) shows that ovaries are unfertilized. Polar nuclei in ovary 2 image were in different plane than the parthenogenetic embryo, so two images were merged to produce one as shown in c and d.

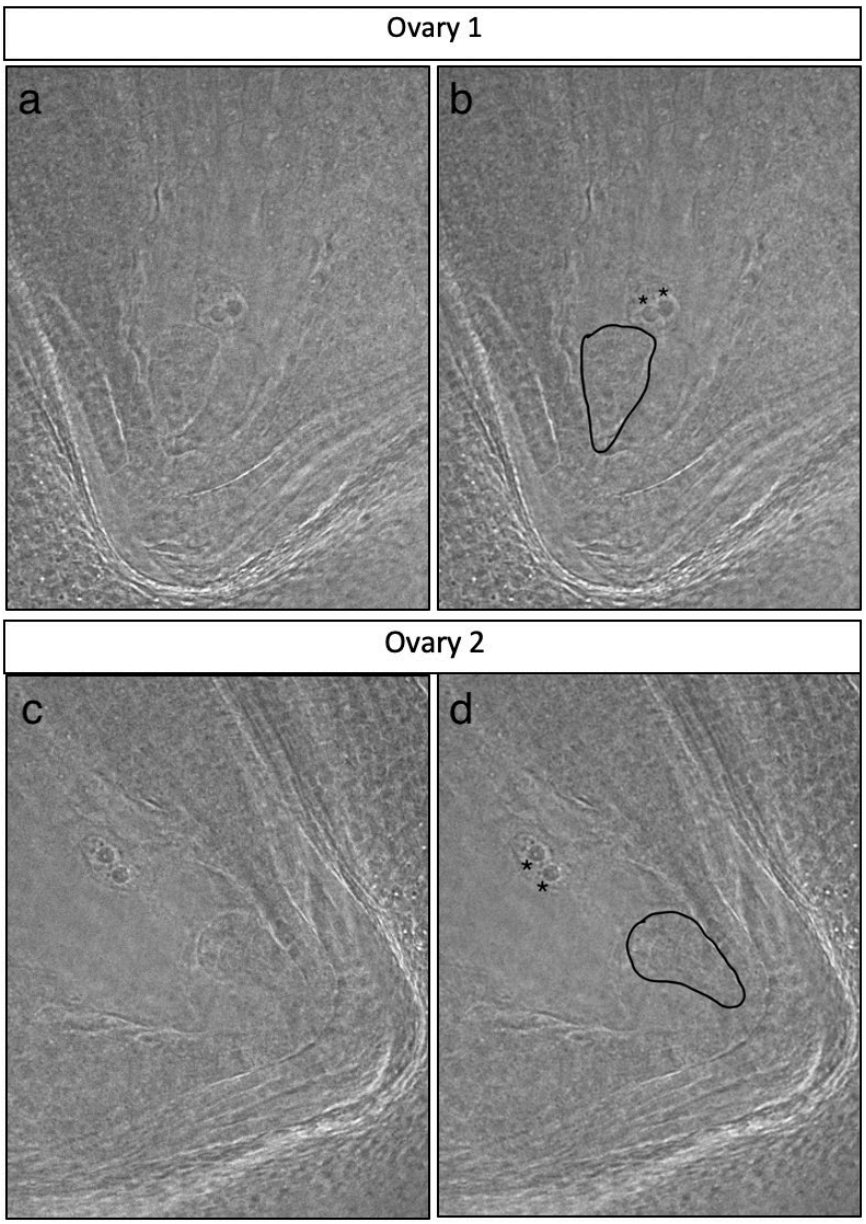


Figure 2.6 Parthenogenesis in *in planta* DEX treated emasculated florets from T_1 15B-12.

Emasculated ovaries from T_1 plant 15B-12 showed parthenogenetic embryo formation (panel a and c) in two different ovaries. The embryo is outlined for distinction (panel b and d). The presence of intact polar nuclei (marked with asterisks in b and d) shows that ovaries are unfertilized.

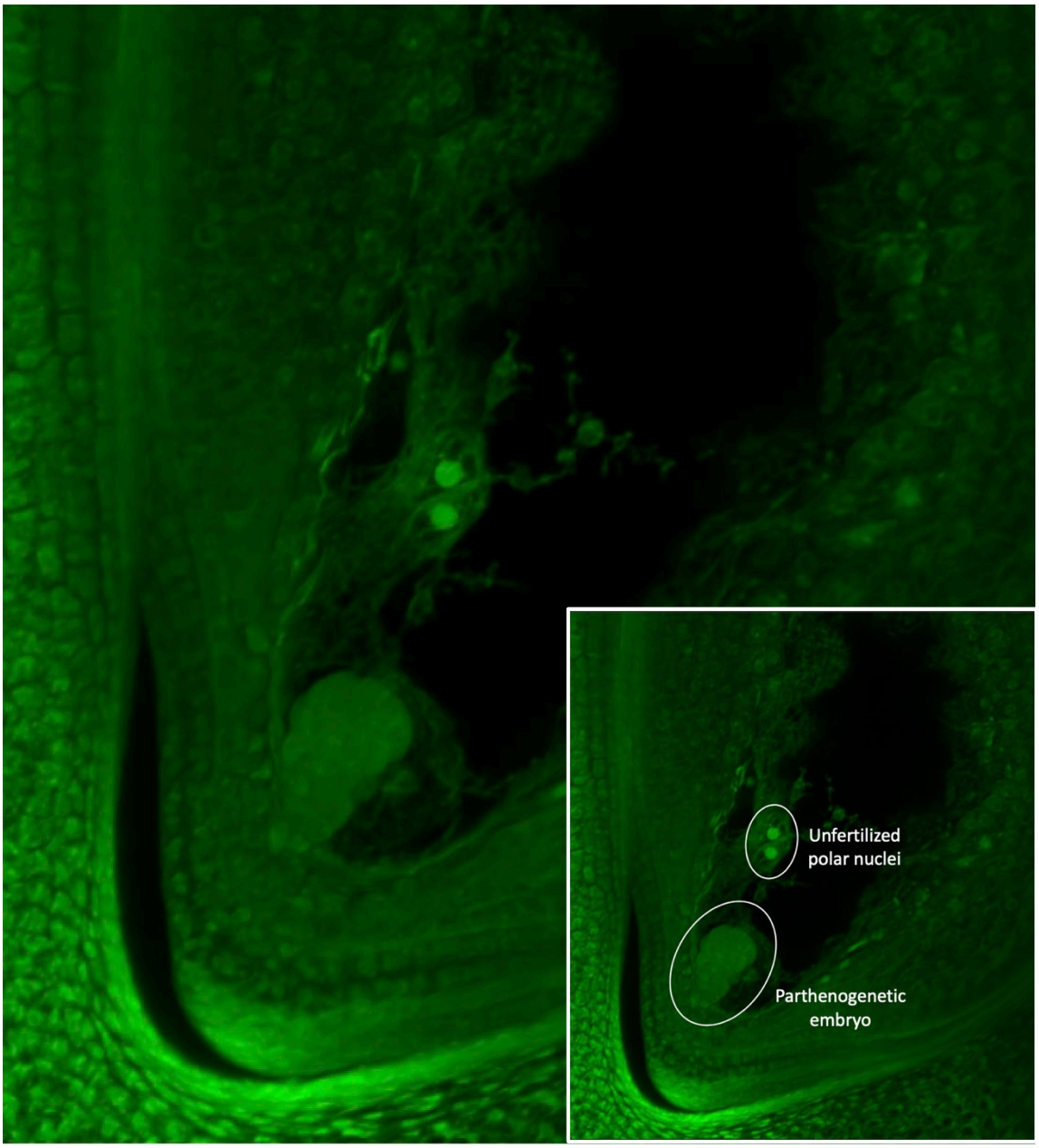


Figure 2.7 Confocal image of a parthenogenetic ovary from T_1 line 15B-12.

Confocal microscope was used to observe *in planta* DEX treated emasculated florets from T_1 plant 15B-12. (Inset) Parthenogenetic embryo and unfertilized polar nuclei are indicated.

CHAPTER 3

THE EFFECT OF EGG CELL PLOIDY LEVEL ON PENETRANCE OF PARTHENOGENESIS[1]

[1]Sidhu, G.S., Conner, J., Ozias-Akins, P. (2021). To be submitted to *Plant Reproduction*.

Abstract

Apomixis can be used to preserve heterotic vigor of hybrids. Natural apomicts exhibiting gametophytic apomixis, undergo apomeiosis to produce non-recombined, unreduced egg cells which initiate embryogenesis without fertilization, also called parthenogenesis. *PsASGR-BBML*, an AP2 transcription factor, induces parthenogenesis in a natural apomict *Pennisetum squamulatum*. As most natural apomicts are polyploid, parthenogenesis occurs in unreduced egg cells, while in transgenic rice transformed with *PsASGR-BBML*, it occurs in reduced egg cells which may be one cause for the observed low penetrance of the trait and ovary developmental abnormalities. Here, we tried to observe the effect of egg cell ploidy level on seed-set and penetrance of parthenogenesis in transgenic rice. Although due to skewed genetic segregation, we didn't generate enough of the desired genotypes, preliminary data shows that higher dosage of PsASGR-BBML could be inhibiting fertilization resulting in only haploid and diploid seeds from plants producing reduced and unreduced egg cells, respectively.

Introduction

In natural apomicts, parthenogenesis usually occurs as a part of apomixis – the other two parts being apomeiosis and pseudogamous/autonomous endosperm formation. Most natural apomictic species are also polyploids, whereas their lower ploidy forms are sexual (Kaushal et al., 2019). For example, diploid cytotypes of *Paspalum notatum*, *Poa pratensis* and *Tripsacum dactyloides* are sexual while polyploid cytotypes are apomicts (Ozias-Akins and van Dijk, 2007). Similarly, apomictic *Pennisetum squamulatum*, from which *PsASGR-BBML* has been cloned and its close relative, *Cenchrus ciliaris* are octoploid and tetraploid respectively. Due to polyploidy and apomeiosis, the egg cells of these natural apomictic plants contain more than one copy of the genome. Mimicking the mitotic division, apomeiosis results in unreduced and non-recombined egg cells, which then initiate parthenogenetic embryogenesis to produce apomictic progeny. In *AtDD45::PsASGR-BBML* transgenic rice, the megaspore mother cells undergo normal meiosis and produce reduced (haploid) egg cells which then initiate parthenogenetic embryogenesis under the influence of PsASGR-BBML.

Expression of *BBM-like* genes is required in rice zygotes to initiate embryogenesis. *OsBBM1*, one of four *BBM-like* genes found in rice, was shown to exhibit paternal-allele dependent expression in zygotes (Anderson et al., 2017; Rahman et al., 2019). Later in globular-like embryos, *OsBBM1* expression becomes bi-allelic, possibly due to auto-activation of the maternal-allele by the paternally expressed *OsBBM1*. The observed dependence of embryogenesis initiation on paternally expressed *OsBBM1* necessitates the fertilization of egg cell by sperm-cell in order to develop into an embryo. These results not only explain the inability of unfertilized ovaries to initiate embryogenesis but also hypothesize that ectopic expression of *BBM-like* genes in these ovaries may lead to fertilization-independent embryogenesis. Supporting this hypothesis,

Khanday et al. (2019) observed that the expression of *OsBBM1* under an egg cell-specific promoter *AtDD45* from *Arabidopsis* (Steffen et al., 2007), leads to parthenogenetic embryo development in rice. Rahman et al. (2019) observed that due to the suppression of the function of *OsBBM1* and its homologs, rice zygotes produced by electro-fusion of isolated gametes could not initiate cell division and ultimately degenerated as compared to control zygotes which developed normally into embryos. Similarly, Khanday et al. (2019) observed that triple knockouts of the rice baby boom genes – *OsBBM1*, *OsBBM2* and *OsBBM3* results in embryo abortion. These observations establish the role of *BBM-like* genes during the initiation of zygotic embryogenesis in rice and during the parthenogenetic embryogenesis in unfertilized egg cells with ectopic expression of these genes. Although *BBM-like* genes are essential for initiating embryogenesis, their over-expression in rice zygotes or unfertilized egg cells doesn't enhance cell-division but appears to be rather inhibitory. When Rahman et al. (2019) over-expressed the *OsBBM1* in rice zygotes by transfecting these with a *p35S-HSP70intron::OsBBM1* plasmid (*p35S-HSP70intron* acts as a strong constitutive promoter), three out of nine zygotes could not initiate cell division and finally degenerated. Five out of six remaining zygotes initiated cell division but showed abnormal embryo development and ultimately ceased. This inhibitory effect of the overexpression of *BBM-like* genes on zygote development indicates that an optimal activity of these genes is required for normal embryogenesis.

During the dissection of unfilled rice florets from three different transgenic lines (transformed with genomic and cDNA based *PsASGR-BBML* under native promoter and genomic *PsASGR-BBML* under *AtDD45* promoter), Conner et al. (2017) observed ovaries with incomplete seed development. These ovaries exhibited multiple developmental abnormalities ranging from no expansion to expanded ovaries filled with a clear solution. Certain ovaries had developed little

green shoot/root like structures while others had just a clump of cells but with no endosperm development. Three caryopses having the shoot/root-like projections were regenerated into plantlets by culturing on Murashige & Skoog (MS) media supplemented with 3 % sucrose. These plantlets were positive for the *PsASGR-BBML* transgene and were found to be haploid after flow cytometry analysis. As an optimal level of PsASGR-BBML activity is required for normal embryogenesis, the observed abnormal phenotypes may have resulted from its ectopic regulation of transcription in reduced egg cells in rice as compared to the unreduced egg cells in natural apomicts. To test this hypothesis, the penetrance of parthenogenesis induced by PsASGR-BBML in haploid and diploid rice egg cells can be compared. Higher penetrance of parthenogenesis in the diploid rice egg cells will imply that the unreduced female egg cell is more permissive for normal embryo development via parthenogenesis than sexually reduced egg cells. Moreover, histological analysis of the developing parthenogenetic ovaries with egg cells of varying ploidy levels can shed more light on the developmental aspects of abnormal phenotypes.

OSD1 stands for *omission of second division* and its role was first established in *Arabidopsis* by d'Erfurth et al. (2009). During normal meiosis, homologous chromosomes undergo crossing-over and are separated during the first meiotic division. The recombined sister chromatids are separated during the second meiotic division producing four haploid daughter cells. In *osd1* mutant plants, the first meiotic division occurs normally but the second division is omitted resulting in two diploid daughter cells (d'Erfurth et al., 2009).

Arabidopsis has one *OSD1* paralog, *UVI4*, but disruption of this gene doesn't affect meiosis but mitosis (d'Erfurth et al., 2009). Rice has two *OSD1* homologues – LOC_Os04g39670 and LOC_Os02g37850 (Mieulet et al., 2016). AMBA12 lines have a 5.1 kb T-DNA insertion in the second intron of LOC_Os02g37850, leading to the formation of diploid gametes which produce

tetraploid progeny upon selfing. The diploid gamete phenotype provided by AMBA12 can be combined with the *PsASGR-BBML* induced parthenogenesis to see if penetrance of the trait is affected by the increased ploidy of the egg cell. In order to achieve this, we crossed AMBA12 plants heterozygous for the *osd1* mutation (ttOo, t – *PsASGR-BBML* absent, O – *OSD1*, o – *osd1*) with pollen from R448 plants homozygous for *PsASGR-BBML* (TTOO). As shown in Figure 3.1, F_2 progeny was taken from TtOo F_1 plants and should have nine different genotypes as observed in a typical di-hybrid cross.

Our ability to assess the effect of egg cell ploidy level on the penetrance of parthenogenesis in F_3 ovaries depends on our ability to effectively separate F_2 plants into various genotypic classes as shown in the Figure 3.1. As F_1 plants which are hemizygous for *PsASGR-BBML* and heterozygous for *osd1* (TtOo) will be taken to the F_2 generation, some portion of the F_2 progeny is expected to be haploid which can be differentiated with flow cytometry. The diploid plants in F_2 will belong to nine genotypic categories observed in a typical dihybrid cross, provided that the *PsASGR-BBML* transgene insertion is not on the same linkage group as *OSD1*. Three out of nine genotypic classes are expected to be *PsASGR-BBML* negative, while the remaining six classes will either be homozygous or hemizygous for the transgene. As the segregation of a hemizygous transgene will impact the penetrance of parthenogenesis, we plan to analyze only those plants which are either homozygous for *PsASGR-BBML* (TTOO and TToo) or where the transgene is expected to be present in each egg cell due to non-reduction, although the transgene might be hemizygous itself (Ttoo). In order to differentiate between *PsASGR-BBML* hemizygous and homozygous F_2 plants, we need to determine the location of the transgene integration in R488 lines, the *AtDD45::PsASGR-BBML* plants used as pollen donors. Once the location is identified,

flanking primers can be designed and used as co-dominant markers to determine the zygosity of *PsASGR-BBML* in F_2 plants.

Several methods can be used for finding the location of a transgene insertion in transgenic plants. Historically, the aim was to isolate, amplify and sequence the flanking regions. Today, high-throughput sequencing of the entire genome followed by bioinformatics approaches to track the transgene integration are possible. In inverse-PCR (abbreviated as i-PCR, depicted in Figure 3.2), genomic DNA is digested with restriction enzymes which don't cut within the T-DNA sequence. Multiple restriction digestion reactions can be carried out simultaneously using different restriction enzymes. The digested genomic DNA is circularized by intra-molecular ligation and amplified by primers designed at the ends of the T-DNA. The sequencing of this PCR-product can generate information about the transgene insertion location. i-PCR has been successfully used to locate transgene insertion sites in transgenic rice (An et al., 2003; Jeong et al., 2006).

Adapter-ligation PCR (depicted in Figure 3.3) is an improved method over i-PCR and is commonly used for isolation of transgene flanking sequences (Devic et al., 1997; O'Malley et al., 2007). This PCR uses nested amplification reactions for enriching the junction regions which are sequenced to determine flanking sequences. This PCR can be used to either isolate the left-border or right-border flanking sequence of the transgene insertion depending upon the primers used for amplification. Genomic DNA is digested with restriction enzymes and the ends of the restricted DNA fragments are ligated to a special adaptor sequence, comprised of a longer upper strand (ADPR1) and a shorter lower strand (ADRP2). ADPR1 contains successive sequences common to two adaptor primers AP1 and AP2. Depending on the left or right-border flanking sequence to be isolated, two successive gene-specific primers from the T-DNA are designed and each of these primers is paired with an adaptor primer (AP1 or AP2) for the nested amplification reaction.

Adaptor-ligation PCR has been successfully used to determine the T-DNA insertion pattern in transgenic rice (Sallaud et al., 2003).

Paired end sequencing data can also be used for determining the transgene insertion site (Kovalic et al., 2012). After sequencing, the reads can be sorted into three types using bioinformatic approaches. Type I reads have both mates from a pair map to genomic regions, type II reads have both mates from a pair mapping to the T-DNA and/or parts of the transformation vector sequence while type III reads have one mate from the pair mapping to the T-DNA/vector sequence and the other mate maps to the genomic region. Type III reads contain the junction between T-DNA/vector sequence and the flanking genomic regions. The mates of type III, which map to the T-DNA/vector sequence, are called orphan reads. Finding the paired mates of orphan reads followed by their mapping onto genome sequence, can elucidate the transgene insertion site. *De novo* assembly of type II and type III reads can be used to infer the structure of the T-DNA.

Materials and Methods

Genetic material

AMBA12 T-DNA insertion line (*OSD1/osd1*) seed was obtained from Genoplante (https://urgi.versailles.inra.fr/Projects/Achieved-projects/Genoplante). Due to poor seed viability, one heterozygous line was regenerated from callus of an AMBA12 seed and allowed to self-pollinate to get F_1 seeds. CTAB isolated DNA from 48 F_1 plants were PCR genotyped with primer-pairs 4254/55 and 4255/56 to distinguish between different segregates. These primers have been adopted from Mieulet et al. (2016) and their sequences are presented in Table 3.1. Heterozygous plants were retained and used as female parents for crossing with pollen from transgenic lines homozygous for *AtDD45::PsASGR-BBML*.

Pollen donor plants were derived from seed (R488) generated from the T_0 transgenic line D25C transformed with AtDD45*::gPsASGR-BBML* (Conner et al., 2017*)*. Three T_1 plants out of fourteen germinated from R448 seed were haploid, from which one haploid plant (R448-11) was doubled with colchicine. Seed from a diploid, transgene-positive T_2 plant (R448-11B) was used to obtain T_3 progeny, from which three diploid, transgene-positive plants (R448-11B-2/4/5) were used as pollen parents in crosses with *OSD1/osd1* plants. Seed from R448-6, a transgene positive, diploid sibling of R448-11 was used to obtain T_2 progeny. All nine T_2 plants were transgene positive and two (R448-6B and R448-6E) were carried to the T_3 generation. Three diploid, transgene positive T_3 plants (R448-6E-7/8 and R448-6B-1) were used as pollen parents in crosses with *OSD1/osd1* plants.

Crossing

Rice crosses were made according to the protocol from McCouch Lab at Cornell University (http://ricelab.plbr.cornell.edu/cross_pollinating_rice). Florets from *OSD1*/*osd1* heterozygous plants were emasculated one day prior to anthesis and covered with a pollination bag. The following day, emasculated panicles were placed in a pollination bag together with a flowering panicle from a pollen donor plant. The bag was tapped 2-3 times a day for the following 2-3 days to assure cross-pollination. One week later, the pollen donor panicles were discarded and *OSD1*/*osd1* panicles were allowed to develop inside the bag, which was removed a few weeks later.

DNA extraction

About 100 mg of young leaf tissue was ground in 200 µL of 2X CTAB buffer (100 mM Tris-Cl (pH = 8), 20 mM EDTA, 1.4 M NaCl, 2 % CTAB, and 0.2 % β-mercaptoethanol) in 1.5 mL eppendorf tubes with an electric mortar-pestle. After grinding, an additional 300 µL CTAB

buffer was added, vortexed and incubated at 65°C in a water bath for 15 minutes. Tubes were taken out and 500 µL of 24:1 Chloroform:Iso-amyl alcohol was added, mixed and centrifuged for 10 minutes at 14,000 rpm. 350 µL of supernatant was transferred to a new eppendorf tube and an equal volume of iso-propanol was added, mixed and centrifuged for 10 minutes at 14,000 rpm. The supernatant was discarded, and the DNA pellet was washed with 70 % ethanol, air dried for 10 minutes and re-suspended in 75 µL 0.5X Tris-EDTA buffer along with 0.1 µL RNase A (working conc. 10 mg mL^{-1}) per sample.

Determining transgene integration location in R448 lines -

Restriction digestion

0.5 µg genomic DNA was digested in a 25 µL reaction volume consisting of 1X CutSmart® Buffer (New England BioLabs, Massachusetts, USA) and 10 U restriction enzyme (*Nhe*I*, Avr*II*, ApaI* and *Hpa*I for i-PCR and *Eco*RV and *Ssp*I for adaptor-ligation PCR). Reaction components were incubated for 1 hour at the recommended incubation temperature for given restriction enzyme followed by heat inactivation. For restriction enzymes which could not be heat inactivated, the restricted DNA was purified with QIAquick Gel Extraction Kit (Qiagen, Maryland). Digestion was verified by running the digested DNA on a gel along with λ DNA cut with Hind-III as a size marker. Purified/heat incubated DNA free of restriction enzyme was used in self-ligation or adaptor-ligation reactions for inverse-PCR and adaptor-ligation PCR respectively.

Self-ligation

Self-ligation reactions consisted of 50 ng of digested, linear DNA along with 1X T4 DNA Ligase Reaction Buffer and 5 U T4 DNA Ligase in a 50 µL reaction followed by incubation for 3 hours – sticky end restriction enzymes at 20°C and blunt end restriction enzymes at 22°C.

Inverse-PCR

Four forward primers (987, 989, 4900 and 4902) were designed inside of the right border and a corresponding three reverse primers (2729, 4901 and 4903) were designed inside of the left border from the *AtDD45::PsASGR-BBML* transformation vector as shown in Figure 3.4. Sequences of these primers are provided in Table 3.2. Purified genomic DNA from a R448 T_3 line 11B-7 was digested with *Nhe*I (G|CTAGC), *Avr*II (C|CTAGG), *Apa*I (GGGCC|C) and *Hpa*I (GTT|AAC). Circularized DNA resulting from self-ligation reaction was used in four separate PCR reactions using the four sets of primers mentioned above – 987F/2729R, 989F/2729R, 4900F/4901R, 4902F/4903R.

Oligo annealing to generate adaptors

1 µM each of two strands of adaptor molecule – ADPR1 and ADPR2 (Table 3.3, Sallaud et al. (2003)), were annealed in a 50 µL reaction consisting of 1X NEBuffer™ 4 (New England BioLabs, Massachusetts, USA) incubated at 94°C for 5 minutes. After incubation, the reaction mixture was allowed to cool to room temperature.

Adaptor-ligation

Two blunt-end restriction enzymes – *Eco*RV (GAT|ATC) and *Ssp*I (AAT|AAT) were used to restrict genomic DNA from four R448 plants (11B-2, 11B-5, 6B-3 and 6E-8). A 20 µL ligation reaction consisted of 100 ng of restricted DNA along with 1X T4 DNA Ligase Reaction Buffer, 40 U T4 DNA Ligase and adaptor molecules with molar amount 20 times that of restricted DNA ends (https://www.neb.com/tools-and-resources/usage-guidelines/tips-for-maximizing-ligation-efficiencies). The reaction mixture was incubated overnight at 16°C followed by Ligase inactivation at 65°C for 10 minutes.

PCR conditions

A 25 µL PCR reaction consisted of 2 µL DNA (genomic DNA for genotyping, restricted and self-ligated DNA for i-PCR and adaptor-ligated restricted DNA for adaptor-ligation PCR), 1X PrimeSTAR GXL Buffer, 200 µM dNTP, 0.2 µM primers, and 0.625 U PrimeSTAR GXL DNA Polymerase (Takara Bio Inc., CA, USA). For left-border adaptor-ligation PCR, the PCR1 was performed with primers AP1/HYG1 and the product was diluted 50 times before using in nested PCR2 performed with primers AP2/CAMB3. PCR2 amplified product was purified with QIAquick Gel Extraction Kit (Qiagen, Maryland), cloned into Zero Blunt™ TOPO™ PCR cloning kit (Thermo Fisher Scientific, Massachusetts), and sequenced by Psomagen, Maryland. The primer sequences AP1, AP2, HYG1 and CAMB3 are presented in Table 3.3.

Whole genome sequencing

Genomic DNA from two R448 T_3 plants – 11B-4 and 6E-7 was sheared (550 bp fragments) using the Covaris E220 Evolution sonicator by the Georgia Genomics and Bioinformatics Core at Athens, GA (https://dna.uga.edu). Library preparation was done using the Illumina TruSeq Nano DNA Low Throughput Library preparation kit followed by paired end sequencing in a NovaSeq S4 Flow Cell at 26 X by Novogene at Sacramento, CA (https://en.novogene.com).

Bioinformatics analysis

1. **Mapping**

Bowtie2 version 2.3.5.1, an ultrafast, memory-efficient short read aligner (Langmead and Salzberg, 2012), was used to align paired end sequences to the transformation vector sequence, to extract paired mates of orphan reads, and to align these extracted mates onto rice genomic sequence. Nipponbare rice reference genome was obtained from Michigan State University rice database release 7 with a total genomic size of 373,245,519 bp

(http://rice.plantbiology.msu.edu/annotation_pseudo_current.shtml). *Sapleo2* cluster at the Georgia Advanced Computing Resource Center (GACRC) at the University of Georgia (https://gacrc.uga.edu) was used for all bioinformatics analysis.

For determining the location of the transgene insertion in R448-T_3 plants – 11B-4 and 6E-7, the paired-end sequencing data from these samples was mapped to the transformation vector sequence by using the following command –

```
module load Bowtie2/2.3.5.1-foss-2018a

time bowtie2 -I 300 -X 3000 -x plasmid -1 'forward-reads-location/file-name.fq' -2 'reverse-reads-location/file-name.fq' -S 'output-file-location/file-name.sam'
```

The '- I' and '- X' arguments respectively define minimum and maximum fragment lengths for valid paired-end alignments. Our sequencing reads have a length of 150 bp, so setting '-I' as 300 bp will output only the non-overlapping reads. Also, the insert size in our sample varies from 200 bp to about 3000 bp, so '- X' was set at 3000 bp.

The SAM files were converted to BAM format using the following command –

```
samtools view -bS 'file-name.sam' > 'file-name.bam'
```

The BAM files were sorted using the following command –

```
samtools sort 'file-name.bam' > 'sorted-file-name.bam'
```

The sorted files were indexed using the following command –

```
samtools index 'sorted-file-name.bam'
```

The BAM index stats were retrieved from the sorted BAM files using the following command –

```
samtools idxstats 'sorted-file-name.bam' > 'sorted-file-name.txt'
```

The unmapped mate-pairs (F264 reads) of the mapped orphan reads (F260 reads) were extracted from the BAM file using the following command –

```
time samtools view -u -f4 -F264 'sorted-file-name.bam' > 'file-name-F264-reads.bam'
```

The BAM file containing the F264 reads was converted to a FASTQ file using the following command –

```
bamToFastq -bam 'file-name-F264-reads.bam' -fq 'file-name-F264-reads.fastq'
```

The resulting FASTQ file, containing the F264 reads, was mapped to the rice reference sequence using Bowtie2 to see how many of these reads map to each rice chromosome. The results were viewed using Integrative Genomics Viewer (Robinson et al., 2011).

2. *De novo* assembly

For determining the insertion structure, *de novo* assembly of type II (reads in which both pairs mapped to vector sequence) and type III reads (the orphan reads – F260 reads, and their paired-mates – F264 reads) was performed.

For assembling the reads, BAM files each containing type II, F260 and F264 reads were merged into a single BAM file using the following command –

```
samtools merge -u 'file-name-merged.bam' 'file-name-input1.bam' 'file-name-input2.bam'
```

The resulting BAM file was sorted according to query name using the following command –

```
samtools sort -n 'file-name-merged.bam' -o 'file-name-merged.qsort'
```

The query sorted file was then converted to two separate FASTQ files, one with forward reads and the other with reverse reads using the following command –

```
bedtools bamtofastq -i merged.qsort -fq1 'file-name-forward.bam' -fq2 'file-name-reverse.bam'
```

These two FASTQ files were then used for *de novo* assembly using Velvet (Zerbino and Birney (2008), https://www.ebi.ac.uk/~zerbino/velvet/) or Geneious (https://www.geneious.com) assemblers.

Results

Rate of parthenogenesis and seed set in 11B and 6B/E R448 lines

The rate of parthenogenesis and seed set was determined in three R448-11B T_3 lines and six R448-6B/E T_3 or T_4 lines. As shown in Table 3.4, 11B lines had very low rates of parthenogenesis (average 2.8 %) but relatively high seed set (average 79.9 %), while 6B/E lines had comparatively high rates of parthenogenesis (average 93.9 %) but low seed set (average 22.3 %).

Determining the transgene insertion location in R448 lines –

Inverse PCR (i-PCR)

Purified genomic DNA from a R448 T_3 line 11B-7 was digested in four separate restriction reactions followed by adaptor ligation and PCR amplification and the results are presented in Figure 3.5. Each of the four primer pairs showed similar band patterns with the band sizes smaller than expected. Due to this unexpected observation, these fragments were not sequenced.

Adaptor-ligation PCR

Genomic DNA from four R448 plants (11B-2, 11B-5, 6B-3 and 6E-8) was restricted with two blunt-end restriction enzymes – *Eco*RV (GAT|ATC) and *Ssp*I (AAT|AAT). The results of Adaptor-ligation PCR are presented in Figure 3.6. In 11B lines, three fragments were observed in *Eco*RV digestion (11B-*Eco*RV-L, 11B-*Eco*RV-M, 11B-*Eco*RV-S) and were sequenced, while in *Ssp*I digestion, two fragments were observed, and one was sequenced (11B-*Ssp*I). In 6B/E lines, two fragments were observed in *Eco*RV digestion with one being sequenced (6B/E-*Eco*RV) while *Ssp*I digestion produced a very small fragment of about 300 bp. The sequencing results are presented figuratively in Figure 3.7. 11B lines seem to have three different complete/incomplete insertions. 11B-*Eco*RV-L and 11B-*Ssp*I represent a portion of an expected insertion which also

includes sequences upstream of the left-border such as the Kanamycin bacterial resistance gene. *KanR* gene includes one restriction site each for *Eco*RV and *Ssp*I and generates the 1430 and 1503 bp bands corresponding to 11B-*Eco*RV-L and 11B-*Ssp*I respectively. 11B-*Eco*RV-M represents a junction of two expected insertions with the first insertion including the 3'UTR region of *PsASGR-BBML* but not the right-border and second insertion starting with a truncated left-border. The eighth exon of *PsASGR-BBML* contains an *Eco*RV restriction site and generates a 1240 bp band. 11B-*Eco*RV-S represents a very similar insertion but the second insertion doesn't include any left-border and directly starts with the CaMV 3'UTR generating a little shorter band of 1032 bp as compared to 11B-*Eco*RV-M. R448-6B/E lines have at least one rearranged insertion comprising of first insertion from Left Border to Right border (including aminoglycoside phosphotransferase CDS, replication origin and STA region from pVS1 plasmid) and a second insertion also from Left border to Right border but including the expected hygromycin selection gene (*Hyg R*) and *PsASGR-BBML* transgene. The second insertion of left border is truncated and has created a *Ssp*I restriction site, which produces a small band of about 300 bp size as observed on a gel. The sequencing results confirmed that the R488 plants analyzed with adaptor-ligation PCR did not contain simple T-DNA left to right border integrations and that the fragments generated did not possess rice genomic flanking sequence.

Whole genome sequencing

1. **Quality check for paired end sequencing data**

Quality check was performed with FastQC. The total number of paired reads in the 11B-4 sample was 519,046,776 (519,046,776 forward and 519,046,776 reverse reads) with length of 150 bp each. Average Phred score for the sample was 36 (encoded on Illumina 1.9 standards) which suggests that sequences are of good quality. Although there were no over-represented sequences

in 11B-4 forward reads, reverse reads had an over-representation of 150 bp GGG….GG sequences (865,269 out of 519,046,776 – 0.166 %). These over-represented reads could be filtered out by using standard software, but as the first step was to map all the reads to the transformation vector sequence, this filtering was not absolutely necessary and was not performed. For 6E-7, the total number of paired-end reads was 551,221,802 with length of 150 bp each. Similar to 11B-4, the 6E-7 sample also had an over-representation of 150 bp GGG….GG sequences (709,513 out of 551,221,802 – 0.128 %).

2. Finding the transgene insertion site

For 11B-4, out of 519,046,776 total paired-end reads, 155,744 reads mapped to the vector sequence out of which 6,916 were orphan reads (F260 reads). For 6E-7, out of 551,221,802 total paired-end reads, 41,805 reads mapped to the vector sequence out of which 1,437 were orphan reads. For both the samples, paired-mates of orphan reads (F264 reads) were mapped to the rice reference genome and the results are presented in Table 3.5. Out of 6,916 F264 reads for 11B-4, 690 (9.97 %) mapped to various rice chromosomes, with the maximum number of reads (400, 57.97 % of the mapped reads) mapping to Chromosome 3. On chromosome 3 (length – 36,413,819 bp), these 400 reads mapped at positions roughly from 33,604,280 to 33,605,550 bp, as shown in a screenshot from IGV in Figure 3.8. There was a gap of about 1,863 bp (33,603,294:33,605,157) between the reads mapping to opposite ends of the transgene insertion, which indicates that this region of rice genomic sequence might be missing. For 6E-7, out of 1,437 F264 reads, 559 (38.90 %) mapped to various rice chromosomes, with the maximum number of reads (507, 90.69 % of the mapped reads) mapping to Chromosome 4. On chromosome 4 (length – 35,502,694 bp), these 507 reads mapped at positions roughly from 33,435,610 to 33,436,900 bp, as shown in an IGV screenshot in Figure 3.9.

3. Determining the transgene insertion structure

De novo assembly of type II and type III reads from the 11B-4 sample was performed using Geneious assembler. Out of 173,572 reads, only 71,794 reads were assembled in 510 contigs, which indicated that this insertion is complex. The BLAST results of contigs showed that this insertion also has two sequences from the pTi plasmid from *Agrobacterium tumefaciens* EHA105 – one large sequence of 4901 bp and one smaller sequence of 1680 bp. De novo assembly of 6E-7 type II and type III reads was done using Velvet assembler and all 43,242 reads were assembled to produce a single contig of length 14,872 bp suggesting that this insertion is much simpler as compared to the insertion in the 11B lines. BLAST of the flanking sequences showed that in 6E-7 line the insertion is located at the end of a putative uncharacterized protein LOC_Os04g56140 on chromosome four.

Combining the *osd1* mutation with the *PsASGR-BBML* transgene

A total of 22 crosses were made generating 18 unique crosses for combining the *osd1* mutation from ttOo lines with the *PsASGR-BBML* transgene from TTOO lines (14 crosses were from 11B parents while four were from 6B/E parents). All the crosses and the number of seeds set per panicle are shown in the Table 3.6.

Genotypic segregation in F_1 progenies

Out of 18 crosses, 7 crosses were selected and 10 F_1 seeds from each selected cross were germinated on MS media. The F_1 plants which resulted from a successful cross were expected to be positive for *PsASGR-BBML* and segregate 1:1 for *OSD1*/*OSD1* and *OSD1*/*osd1*. All F_1 plants (total 66) were genotyped with 1792/1801 for *PsASGR-BBML* and 4254/55 and 4255/56 for *OSD1* (data presented in Table 3.7). All 66 F_1 plants were transgene positive for *PsASGR-BBML* showing

that the crossing was successful. The *OSD1*/*OSD1* and *OSD1*/*osd1* plants in the F_1 segregated in 1:1 ratio, confirming expectations.

***PsASGR-BBML* expression analysis**

PsASGR-BBML expression was analyzed using RT-PCR on RNA extracted from about 30 pre-anthesis ovaries from 10 different F_1 plants and the results are presented in Figure 3.10, Although the same amount of RNA was used for each RT-PCR sample, less intense amplification was identified from the samples with 11B plants as parents as compared to 6B/E.

Genotypic segregation and rate of parthenogenesis in F_2 progenies

Five *OSD1*/*osd1* F_1 plants (Table 3.8) were carried to the F_2 generation in order to get the desired genotypes (TTOO, TToo and Ttoo). Two of the five F_1 plants were from 11B parents (T3-9 X 11B-2 #9 and T3-2 X 11B-4 #5) while the remaining three were from 6B/E parents (T3-6 X 6E-7 #2, T3-6 X 6E-8 #1 and T1-21 X 6B-1 #1). About 50 F_2 seeds were germinated from each F_1 plant and then seedlings were genotyped with the zygosity primers for the *PsASGR-BBML* insertion (11B – 5130/5132 and 5130/1801, 6B – 5126/5128 and 5126/5040) and for *OSD1* (4254/4255 and 4255/4256). F_2 seedling showing heterozygous genotypic signal (TTOo, TtOO, TtOo, Ttoo and ttOo) would indicate the presence of two alleles and therefore imply a diploid plant. Haploid plants should be transgene positive for *PsASGR-BBML*, so plants showing ttOO and ttoo genotypic signal were also considered diploid. Based on this elimination approach, seedlings that were homozygous for *PsASGR-BBML* and homozygous dominant or homozygous recessive for *OSD1* (TTOO, TToo) were subjected to flow cytometry to determine their ploidy level. The number of haploid plants and the number of diploid plants having each of the nine possible genotypes in the F_2 progenies from five F_1 plants analyzed is shown in Table 3.9.

For the F_2 progeny from the 11B plants (T3-9 X 11B-2 #9 and T3-2 X 11B-4 #5), although germination was good (86 and 92 % respectively), the rate of parthenogenesis was very low as we found only one haploid plant in T3-9 X 11B-2 #9 F_2 progeny and none in T3-2 X 11B-4 #5 F_2 progeny. This rate was expected giving the low rate of parthenogenesis in 11B T_3 lines (Table 3.4), which were used as male parents in the original crosses. As the *PsASGR-BBML* insertion in 11B lines (Chr. 3) and *osd1* T-DNA insertion (Chr. 2) are on separate chromosomes, we expected both transgenes to segregate independently of each other and to have TT:Tt:tt and OO:Oo:oo ratios in F_2 progenies as 1:2:1. We observed that for two F_2 progenies derived from 11B TTOO parent, segregation ratios for both transgenes were as expected – T3-9 X 11B-2 #9 F_2 progeny (TT:Tt:tt fits to 1:2:1, $\chi^2 = 5.837$, $p = 0.054$ and OO:Oo:oo fits to 1:2:1, $\chi^2 = 0.395$, $p = 0.820$), T3-2 X 11B-4 #5 F_2 progeny (TT:Tt:tt fits to 1:2:1, $\chi^2 = 2.348$, $p = 0.309$ and OO:Oo:oo fits to 1:2:1, $\chi^2 = 0.261$, $p = 0.877$). This showed that both the transgenes are segregating normally in F_2 progenies from 11B lines, and we were able to select the desired genotypes (TTOO, TToo and Ttoo) for analyzing the rate of parthenogenesis in F_3 seeds.

F_2 progenies from 6B/E plants (T3-6 X 6E-7 #2, T3-6 X 6E-8 #1 and T1-21 X 6B-1 #1) also showed good germination (80, 92 and 86 % respectively). Given the high rate of parthenogenesis in 6B/E T_3 lines (Table 3.4) used as male parents in original crosses, we expected similar results in the F_2 progenies derived from 6B/E TTOO parents, but we observed only five, two and four haploid plants in these three sets of progenies. Also, we observed that the OO:Oo:oo fit 1:2:1 ratio for two progenies (T3-6 X 6E-7 #2, $\chi^2 = 0$, $p = 1.000$ and T3-6 X 6E-8 #1, $\chi^2 = 2.810$, $p = 0.238$) while it didn't fit for the T1-21 X 6B-1 #1 progeny ($\chi^2 = 6.907$, $p = 0.0316$). For *PsASGR-BBML*, the number of homozygous plants was greatly reduced from expected and we got only one TToo plant from all three F_2 progenies generated from crosses with 6B/E. While a

TT:Tt:tt ratio progenies from these three crosses didn't fit to 1:2:1 (χ^2 = 11.400, p = 0.003, χ^2 = 18.174, p = 0.000 and χ^2 = 18.814, p = 0.000 for T3-6 X 6E-7 #2 F_2, T3-6 X 6E-8 #1 F_2 and T1-21 X 6B-1 #1 F_2 respectively), the Tt:tt ratio did fit to 1:1 (χ^2 = 2.077, p = 0.149, χ^2 = 0.783, p = 0.376 and χ^2 = 0.209, p = 0.647 for T3-6 X 6E-7 #2 F_2, T3-6 X 6E-8 #1 F_2 and T1-21 X 6B-1 #1 F_2 respectively). This showed that along with the lower rate of parthenogenesis, the 6B/E F_2 progenies are also showing segregation distortion with a smaller number of *PsASGR-BBML* homozygotes than expected. From these progenies, we were able to select enough Ttoo plants for analyzing the rate of parthenogenesis in F_3 seeds, but for TToo we had only one plant (from T3-6 X 6E-7 #2 F_2 progeny) and for TTOO, we obtained none.

Seed set in F_2 plants

As shown in the Table 3.10, the F_2 plants had very low seed set. Ttoo plants, producing unreduced egg cells had an average of 40, 29 and 22 seeds per plant in T3-6 X 6E-7 #2, T3-6 X 6E-8 #1 and T1-21 X 6B-1 #1 F_2 progenies respectively. The single TToo plant from T3-6 X 6E-7 #2 F_2 progeny produced only 29 seeds from whole plant.

Rate of parthenogenesis in F_3 seeds

Out of 20 Ttoo F_2 plants available from three F_2 progenies derived from 6B/E parents, individual flow cytometry was performed on about 30 F_3 seeds from two F_2 plants (T3-6 X 6E-7 #2 #10 and T3-6 X 6E-8 #1 #40) and on about 30 F_3 seedlings from four F_2 plants (T3-6 X 6E-7 #2 #40, T3-6 X 6E-8 #1 #18, T3-6 X 6E-8 #1 #32 and T1-21 X 6B-1 #1 #3). Individual flow cytometry was also performed on about 20 F_3 seedlings from the single TToo plant (T3-6 X 6E-7 #2 #43), with the idea being that diploid seedlings, if any, can be moved to soil, which will give us additional TToo plants to analyze in the F_3 generation as compared to only one in the F_2. The

flow cytometry results are presented in Table 3.11. Ttoo plants (n = 6) showed a 52.3 % rate of parthenogenesis on average, while the single TToo plant we had showed a rate of 94.1 %.

Discussion

Determining the transgene insertion site

The i-PCR results for 11B lines with each of the four primer pairs showed similar banding patterns, which is only possible if all of the four restriction enzymes used had restriction sites at very close proximity to each other, which was unexpected. The results could also be due to tandem insertion of the transgene where a transformation vector sequence from left border to right border was inserted immediately after the insertion of a similar sequence. This explains the relatively short band sizes (less than 300 bp) observed in our results and also the reason behind the similar pattern of bands observed irrespective of the restriction enzyme used. This experiment showed that there might be more than one tandem complete or incomplete transgene insertion and hence i-PCR was not a useful method for isolating the flanking sequences in the transgenic lines used. Left-border adaptor-ligation PCR was then tried to determine the transgene flanking sequence. The R448-11B lines (11B-2 and 11B-5) showed similar banding patterns both with *Eco*RV and *Ssp*I restrictions, suggesting that the 11B lines have a similar transgene insertion pattern. The multiple bands might have resulted due to multiple insertions of the transgene or due to incomplete restriction digestion although the former seems more likely as supported by i-PCR results. Sequencing of these fragments showed that all the sequences were from within the transformation vector and didn't contain flanking sequences from the rice genome. The insertion pattern is complex as it includes sequences outside the left and right-border in addition to the expected sequences (i.e. from left border to right border including selectable marker and transgene). As i-PCR and adaptor-ligation PCR failed to generate the rice flanking sequences, we proceeded to a

whole genome sequencing approach. The paired-end sequencing data was helpful in determining the transgene insertion site for both the 11B-4 as well as 6E-7 line, although determining the insertion structure was difficult in the 11B-4 line which produced a large number of contigs during *de novo* assembly and also contained some foreign sequences from the pTi plasmid. The 6E-7 line has a simple insertion which was easily determined through *de novo* assembly. Both transgene insertion sites did not appear to affect an endogenous rice gene and therefore the phenotypes and the skewing we observed should not be due to the homozygosity of the transgene integration site.

Comparing the rate of parthenogenesis among desired genotypes

We wanted to determine if the ovary developmental abnormalities and low penetrance of parthenogenesis observed in the sexual diploid rice transformed with various *PsASGR-BBML* transgenes could be due to the reduced ploidy level of the egg cells as compared to unreduced egg cells found in natural apomicts. If the PsASGR-BBML transcription factor binds multiple promoters with varying binding affinities, then in the reduced ovaries of rice, it may affect the transcriptional pathways which are normally unrelated to parthenogenesis and hence lead to developmental issues.

Low seed set and high parthenogenesis in R448-6B/E lines, which are homozygous for *AtDD45::PsASGR-BBML*, can be attributed to abnormal ovary development, such that only a few ovaries produce healthy haploid embryos and endosperm to produce functional seeds, while most get aborted at an earlier developmental stage. This is consistent with the observation that 6B/E lines had high rates of parthenogenesis (average 93.9 %) but low seed set (average 22.3 %). The fact that some caryopses had shoot/root-like projections which regenerated into haploid plantlets when placed on MS media (Conner et al., 2017), suggests that transgene carrying ovaries might undergo precocious embryogenesis. This precocious embryogenesis might be inhibiting central-cell fertilization and hence endosperm development as observed in *endospermless1* (*enl1*) mutants

in rice (Hara et al., 2015). In plants heterozygous for the *AtDD45::PsASGR-BBML*, such as the F_1 plants in this study, about half of their egg cells will have the transgene, from which only a few will produce haploid seeds, while most will get aborted. The other half of the egg cells which lack the transgene, will be fertilized normally and produce the seeds with about 1:1 genotypic ratio of *PsASGR-BBML*/- (Tt) and -/- (tt). Due to the presence of these diploid seeds, and only half of the egg cells being *PsASGR-BBML* positive which hold the potential for haploid seed production, the observed rate of parthenogenesis is low in heterozygous plants as compared to homozygous plants while seed set is comparatively high. This is consistent with the fact that we observed only 11 haploids out 140 F_2 seedlings (from three 6B/E F_1 plants), resulting in about 8 % rate of parthenogenesis, while the average seed set on these three F_1 plants was 45.3 % (Table 3.8). These observations can explain why the *AtDD45::PsASGR-BBML* homozygous lines have higher rates of parthenogenesis and low seed set while the heterozygous lines have very low rates of parthenogenesis but comparatively higher seed set.

Plants homozygous for the *osd1* mutation have lower seed set as compared to wild-type plants. Mieulet et al. (2016), observed that *osd1*/*osd1* lines had 25.2 % seed set as compared to 73.8 % in wild type plants. We also observed very low seed set in the TToo and Ttoo lines which might be due to the combination of the effect of the *osd1* mutation and the abnormal embryo development due to *PsASGR-BBML* expression. As was hypothesized in *AtDD45::PsASGR-BBML* homozygous plants producing reduced egg cells (R448-6B/E plants), where only few haploid egg cells develop into haploid embryos and then produce seeds, while most ovaries abort at some developmental stage, a similar phenomenon might be happening in the *AtDD45::PsASGR-BBML* homozygous plants producing unreduced egg cells (TToo plants). Only a few diploid egg cells will develop into diploid embryos to produce seeds, while most will get aborted, probably

due to higher dosage of PsASGR-BBML. This will lead to lower seed set and apparently very high rate of parthenogenesis (94.1 %) as was observed in one TToo plant analyzed in this study. On the other hand, plants heterozygous for *AtDD45::PsASGR-BBML* producing unreduced egg cells (Ttoo plants), will have lower dosage of PsASGR-BBML, and along with producing some diploid seeds through parthenogenesis, will also produce some tetraploid seeds after successful fertilization. This is consistent with the 52.2 % average rate of parthenogenesis observed in Ttoo plants.

Future prospects

Rice *MiMe* genotype (Mieulet et al., 2016) can also be coupled with *AtDD45::PsASGR-BBML* to see the effect of diploid egg cells on the penetrance of parthenogenesis and to synthesize apomixis in rice. As *MiMe* genotype (Mieulet et al., 2016) and one TToo line from this study show 94.1 % penetrance of apomeiosis and parthenogenesis respectively, combining *MiMe* genotype with *PsASGR-BBML* in homozygous form can be used to achieve highly penetrant synthetic apomixis. Although we obtained only one TToo plant from F_2 progeny, very high rate of parthenogenesis (94.1 %) observed in this genotype suggests that more plants of the same genotype can be obtained by germinating F_3 seeds, which can be used to get better estimates of seed-set and penetrance of parthenogenesis. If, due to higher dosage of *PsASGR-BBML*, seed set in TToo lines is found to be considerably lower than Ttoo lines, then incomplete penetrance of parthenogenesis in Ttoo lines (52.3 % on average) might produce more diploid seeds per plant as compared to TToo lines showing full penetrance. This suggests that combining *MiMe* with *PsASGR-BBML* in a heterozygous state might produce more apomictic seeds per plant as compared to using *PsASGR-BBML* in the homozygous state, although an efficient method of differentiating between apomictic diploid and sexual tetraploid progenies will need to be developed. Reciprocal crosses between

plants heterozygous for *PsASGR-BBML* and wild type plants can help elucidate the exact reason behind the low frequency of *PsASGR-BBML* homozygotes. Histological analysis of developing ovaries will also provide more insights into the developmental abnormalities and any difference for that in lines producing reduced and unreduced egg cells.

References

An, S., Park, S., Jeong, D.-H., Lee, D.-Y., Kang, H.-G., Yu, J.-H., Hur, J., Kim, S.-R., Kim, Y.-H., Lee, M., Han, S., Kim, S.-J., Yang, J., Kim, E., Wi, S.J., Chung, H.S., Hong, J.-P., Choe, V., Lee, H.-K., Choi, J.-H., Nam, J., Kim, S.-R., Park, P.-B., Park, K.Y., Kim, W.T., Choe, S., Lee, C.-B., An, G. (2003) Generation and Analysis of End Sequence Database for T-DNA Tagging Lines in Rice. Plant Physiology, **133**(4), 2040-2047.

Anderson, S.N., Johnson, C.S., Chesnut, J., Jones, D.S., Khanday, I., Woodhouse, M., Li, C., Conrad, L.J., Russell, S.D., Sundaresan, V. (2017) The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. Developmental Cell, **43**(3), 349-358.e344.

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

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., Mercier, R. (2009) Turning Meiosis into Mitosis. PLOS Biology, **7**(6), e1000124-e1000124.

Devic, M., Albert, S., Delseny, M., Roscoe, T.J. (1997) Efficient PCR walking on plant genomic DNA.

Hara, T., Katoh, H., Ogawa, D., Kagaya, Y., Sato, Y., Kitano, H., Nagato, Y., Ishikawa, R., Ono, A., Kinoshita, T., Takeda, S., Hattori, T. (2015) Rice SNF2 family helicase ENL1 is essential for syncytial endosperm development. The Plant Journal, **81**(1), 1-12.

Jeong, D.-H., An, S., Park, S., Kang, H.-G., Park, G.-G., Kim, S.-R., Sim, J., Kim, Y.-O., Kim, M.-K., Kim, S.-R., Kim, J., Shin, M., Jung, M., An, G. (2006) Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. The Plant Journal, **45**(1), 123-132.

Kaushal, P., Dwivedi, K.K., Radhakrishna, A., Srivastava, M.K., Kumar, V., Roy, A.K., Malaviya, D.R. (2019) Partitioning Apomixis Components to Understand and Utilize Gametophytic Apomixis. Frontiers in Plant Science, **10**(256).

Khanday, I., Skinner, D., Yang, B., Mercier, R., Sundaresan, V. (2019) A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. Nature, **565**(7737), 91-95.

Kovalic, D., Garnaat, C., Guo, L., Yan, Y., Groat, J., Silvanovich, A., Ralston, L., Huang, M., Tian, Q., Christian, A., Cheikh, N., Hjelle, J., Padgette, S., Bannon, G. (2012) The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology. The Plant Genome, **5**(3).

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods, **9**(4), 357-359.

Mieulet, D., Jolivet, S., Rivard, M., Cromer, L., Vernet, A., Mayonove, P., Pereira, L., Droc, G., Courtois, B., Guiderdoni, E., Mercier, R. (2016) Turning rice meiosis into mitosis. Cell Research, **26**, 1242-1242.

O'Malley, R.C., Alonso, J.M., Kim, C.J., Leisse, T.J., Ecker, J.R. (2007) An adapter ligation-mediated pcr method for high-throughput mapping of t-dna inserts in the Arabidopsis genome. Nature Protocols, **2**(11), 2910-2917.

Ozias-Akins, P. and van Dijk, P.J. (2007) Mendelian Genetics of Apomixis in Plants. Annual Review of Genetics, **41**(1), 509-537.

Rahman, M.H., Toda, E., Kobayashi, M., Kudo, T., Koshimizu, S., Takahara, M., Iwami, M., Watanabe, Y., Sekimoto, H., Yano, K., Okamoto, T. (2019) Expression of Genes from Paternal Alleles in Rice Zygotes and Involvement of OsASGR-BBML1 in Initiation of Zygotic Development. Plant and Cell Physiology, **60**(4), 725-737.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P. (2011) Integrative genomics viewer. Nature Biotechnology, **29**(1), 24-26.

Sallaud, C., Meynard, D., Van Boxtel, J., Gay, C., B›s, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., Portefaix, M., Ouwerkerk, P.B.F., Rueb, S., Delseny, M., Guiderdoni, E. (2003) Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L.) functional genomics. Theor Appl Genet, **106**, 1396-1408.

Steffen, J.G., Kang, I.-H., Macfarlane, J., Drews, G.N. (2007) Identification of genes expressed in the Arabidopsis female gametophyte. **51**(2), 281-292.

Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research, **18**(5), 821-829.

Tables

Table 3.1 Sequences of primers used to determine the zygosity of *PsASGR-BBML* and *osd1* transgenes in F_2 progeny.

Primers 5130/5132 and 5130/1801 were used to determine the zygosity of *PsASGR-BBML* in 11B lines (insertion on chromosome 3) primers 5126/5128 and 5126/5040 were used for 6B/E lines (insertion on chromosome 4). Primers 4254/4255 and 4255/4256 were used to determine the zygosity of *osd1* and these primers are adopted from Mieulet et al. (2016).

Primer	Sequence (5' to 3')
5126	TGATGCCGCAGATGGTGATT
5128	GCTAGGCACAAAACGTAGCG
5040	TGTCATACCACTTGTCCGCC
5130	GCCACGTAGGATCCTGAAGG
5132	CATCTCCTCCTCCCCCTGAT
1801	TTCTCATGGCTCCTAGACTCCCAC
4254	ACAGCAGGGATGGAAGAAGA
4255	TAGTTTACCGGGTGCGATTT
4256	GTCTGGACCGATGGCTGTGTAGAAG

Table 3.2 Sequences of primers used for i-PCR.

Primer	Sequence (5’ to 3’)
987	TCCCAACAGTTGCGCAGCCTGAATGG
989	CCTGAATGGCGAATGCTAGAGCAGCTTG
2729	GGGTTTCGCTCATGTGTTGAGCA
4900	AGATTGTCGTTTCCCGCCTT
4901	GGGTTTCGCTCATGTGTTGAGCA
4902	TCGTTTCCCGCCTTCAGTTT
4903	GGGTTTCGCTCATGTGTTGAGCA

Table 3.3 Sequences of primers and adaptors used in adaptor-ligation PCR.

Primers (AP1, AP2, HYG1 and CAMB3) and adaptors (ADPR1 and ADPR2) used in adaptor-ligation PCR. After an initial digestion reaction, the restricted DNA was ligated with adaptor sequences consisting of annealed ADPR1 and ADPR2, which were then amplified in a nested PCR with PCR1 (AP1 and HYG1) followed by PCR2 (AP2 and CAMB3). All the primers were used here were previously described by Sallaud et al. (2003).

Oligo name	Sequence (5’ to 3’)
AP1	GGATCCTAATACGACTCACTATAGGGC
AP2	CTATAGGGCTCGAGCGGC
HYG1	ATCAGAGCTTGGTTGACGGCAATTT
CAMB3	AGATGCCGACCGGATCTGTC
ADPR1	CTAATACGACTCACTATAGGGCTCGAGCGCCGCCCGGGGAGGT
ADPR2	P-ACCTCCCC-NH_2

Table 3.4 Rate of parthenogenesis and seed set in homozygous R448 – 11B and 6B/E lines (*AtDD45::PsASGR-BBML*).

Twenty-four to twenty-six seeds from seven T_3 and two T_4 plants were analyzed with individual seed flow cytometry to determine the rate of parthenogenesis which was calculated according to the formula = [n(H) + n(H+D)]/total no. of seeds analyzed*100. Seed set data for each of these plants is also presented in the last column of the table.

Plant	Generation	Total seeds analyzed	Haploid (H)	Haploid twins (H+D)	Diploid	Parthenogenesis	Seed set
11B-2	T_3	24	1	1	22	8.3 %	84.4 %
11B-4	T_3	26	-	-	25	0 %	83.0 %
11B-5	T_3	25	-	-	24	0 %	72.2 %
Average						2.8 %	79.9 %
6E-7	T_3	25	16	4	2	80.0 %	25.6 %
6E-10	T_3	24	20	3	1	95.8 %	34.9 %
6E-25	T_3	25	25	-	-	100.0 %	17.4 %
6B-1	T_3	25	23	2	-	100.0 %	14.7 %
6E-10 A	T_4	24	19	2	2	87.5 %	31.2 %
6B-1 A	T_4	23	23	-	-	100.0 %	10.0 %
Average						93.9 %	22.3 %

Table 3.5 Number of F264 reads mapped to various rice chromosomes in 11B-4 and 6E-7 samples.

A total of 6916 and 1437 F264 reads were extracted from 11B-4 and 6E-7 samples respectively, which were then mapped to the rice reference genome. The number of these reads mapping to each of the rice chromosome is presented in the following table. Column two represents the length (bp) of each chromosome. * represents the number of reads mapping to neither of the chromosomes.

Rice chromosome	Length (bp)	No. of F264 reads mapped	
		11B-4	6E-7
Chr1	43270923	18	0
Chr2	35937250	12	5
Chr3	36413819	400	4
Chr4	35502694	8	507
Chr5	29958434	7	1
Chr6	31248787	13	0
Chr7	29697621	9	2
Chr8	28443022	14	5
Chr9	23012720	8	0
Chr10	23207287	6	0
Chr11	29021106	190	35
Chr12	27531856	5	0
ChrUn	633585	0	0
ChrSy	592136	0	0
*	0	6226	878
Total F264 reads		6916	1437

Table 3.6 Seed set in ttOo X TTOO crosses.

A total of 22 crosses were made generating 18 unique crosses for combining the *osd1* mutation from ttOo lines with the *PsASGR-BBML* transgene from TTOO lines (14 crosses with 11B parents and four with 6B/E parents). T1, T2, and T3 plants were derived from an original ttOo seed taken through multiple generations. Resulting progeny were genotyped and heterozygous (ttOo) plants were used as female parents in these crosses. Seed/plants from these ttOo plants were maintained separately. Number of seeds set per panicle for each cross are shown in the last column.

Cross no.	Cross	No. of seeds set per panicle
1	T3-9 X 11B-2	39
2	T3-2 X 11B-4	44
3	T3-3 X 11B-4	32
4	T1-17 X 11B-4	47
5	T3-2 X 11B-5	39
6	T3-16 X 11B-6	27
7	T1-18 X 11B-6	29
8	T2-10 X 11B-8	22
	T2-10 X 11B-8	7
9	T2-9 X 11B-9	18
10	T2-14 X 11B-9	22
11	T3-6 X 11B-10	5
12	T2-9 X 11B-11	29
13	T2-14 X 11B-11	16
14	T1-18 X 11B-12	8
15	T3-6 X 6E-7	28
	T3-6 X 6E-7	28
	T3-6 X 6E-7	25
	T3-6 X 6E-7	7
16	T3-6 X 6E-8	36
17	T3-9 X 6E-10	11
18	T1-21 X 6B-1	17

Table 3.7 Germination and genotyping data of F_1 progeny of *OSD1/osd1* X R448 crosses.

10 seeds each of 7 selected crosses (column 1) were sown and the number of germinated seedlings is presented in column 2. The third column represents the number of *PsASGR-BBML:GR* transgene positive plants (successful crosses) out of germinated. Fourth and fifth columns represent the number of *OSD1/osd1* and *OSD1/OSD1* plants respectively out of successful crosses.

Cross	F_1 progeny			
	Germinated/10	*PsASGR-BBML:GR* positive	*OSD1/osd1*	*OSD1/OSD1*
T3-9 X 11B-2	10	10	5	5
T3-2 X 11B-4	10	10	6	4
T3-2 X 11B-5	9	9	5	4
T3-3 X 11B-4	9	9	4	5
T3-6 X 6E-7	9	9	5	4
T3-6 X 6E-8	9	9	4	5
T1-21 X 6B-1	10	10	6	4

Table 3.8 Bulk Seed Flow Cytometry (BSFC) and seed set in *OSD1*/*osd1* F_1 plants from five selected crosses.

F_1 progeny was taken from seven ttOo X TTOO crosses where *OSD1*/*osd1* and *OSD1*/*OSD1* plants segregated in 1:1 genotypic ratio. Five F_1 plants (marked with stars in the table below) were carried to the F_2 generation to get desirable genotypes for the analysis of rate of parthenogenesis. Table below shows BSFC results (about 4 sets of 5 seeds each) from *OSD1*/*osd1* along with their seed set.

T3-9 X 11B-2 *OSD1*/*osd1* F_1 progeny		
Plant	BSFC	Seed set
T3-9 X 11B-2 #1	N/A	0.0 %
T3-9 X 11B-2 #4	N/A	24.0 %
T3-9 X 11B-2 #5	N/A	26.2 %
T3-9 X 11B-2 #9*	N/A	51.5 %
T3-9 X 11B-2 #10	N/A	58.6 %
Average seed set		32.1 %

T3-2 X 11B-4 *OSD1*/*osd1* F_1 progeny		
Plant	BSFC	Seed set
T3-2 X 11B-4 #1	0/3	1.7 %
T3-2 X 11B-4 #2	0/4	51.7 %
T3-2 X 11B-4 #3	0/4	60.4 %
T3-2 X 11B-4 #5*	0/4	39.8 %
T3-2 X 11B-4 #9	0/4	61.8 %
T3-2 X 11B-4 #10	0/4	58.5 %
Average seed set		45.6 %

T3-6 X 6E-7 *OSD1*/*osd1* F_1 progeny		
Plant	BSFC	Seed set
T3-6 X 6E-7 #2*	2/4	51.2 %
T3-6 X 6E-7 #4	2/4	45.1 %
T3-6 X 6E-7 #5	3/4	34.9 %
T3-6 X 6E-7 #6	1/4	33.9 %
T3-6 X 6E-7 #9	2/4	34.5 %
Average seed set		39.9 %

T3-6 X 6E-8 *OSD1*/*osd1* F_1 progeny		
Plant	BSFC	Seed set
T3-6 X 6E-8 #1*	3/4	47.9 %
T3-6 X 6E-8 #2	2/4	32.3 %
T3-6 X 6E-8 #3	2/4	43.1 %
T3-6 X 6E-8 #5	2/4	28.1 %
Average seed set		37.8 %

T1-21 X 6B-1 *OSD1*/*osd1* F_1 progeny		
Plant	BSFC	Seed set
T1-21 X 6B-1 #1*	3/4	37 %
T1-21 X 6B-1 #3	2/4	32.1 %
T1-21 X 6B-1 #5	1/4	45.7 %
T1-21 X 6B-1 #6	2/4	35.4 %
T1-21 X 6B-1 #7	1/4	30.9 %
Average seed set		36.7 %

Table 3.9 Genotypic segregation data in the F_2 generation from 11B and 6B/E parents.

About 50 F_2 seeds were germinated and then seedlings were genotyped for zygosity with primers for *PsASGR-BBML* (11B – 5130/5132 and 5130/1801, 6B – 5126/5128 and 5126/5040) and for *OSD1* (4254/4255 and 4255/4256). The table below shows the segregation ratios of 9 possible genotypes in the F_2 generation from five of the F_1 *PsASGR-BBML*/-, *OSD1*/*osd1* (TtOo) plants. The number of haploid plants observed is shown in parentheses adjoining the name of each F_1 parent plant. Only those F_2 plants which showed TTOO or TToo signal during genotyping were subjected to flow cytometry as other genotypes can't be haploids because of heterozygosity or at least one transgene (TTOo, TtOO, TtOo and Ttoo) or lack of *PsASGR-BBML* transgene (ttOO, ttOo and ttoo).

T3-9 X 11B-2 #9 F_2 (1 Haploid)				
	TT	Tt	tt	Total
OO	TTOO 5	TtOO 4	ttOO 2	11
Oo	TTOo 7	TtOo 13	ttOo 3	23
oo	TToo 5	Ttoo 3	ttoo 1	9
Total	17	20	6	43

T3-2 X 11B-4 #5 F_2 (0 Haploid)				
	TT	Tt	tt	Total
OO	TTOO 2	TtOO 4	ttOO 4	10
Oo	TTOo 7	TtOo 9	ttOo 8	24
oo	TToo 4	Ttoo 5	ttoo 3	12
Total	13	18	15	46

T3-6 X 6E-7 #2 F_2 (5 Haploid)				
	TT	Tt	tt	Total
OO	TTOO 0	TtOO 5	ttOO 5	10
Oo	TTOo 0	TtOo 13	ttOo 7	20
oo	TToo 1	Ttoo 6	ttoo 3	10
Total	1	24	15	40

T3-6 X 6E-8 #1 F_2 (2 Haploid)				
	TT	Tt	tt	Total
OO	TTOO 0	TtOO 6	ttOO 5	11
Oo	TTOo 0	TtOo 15	ttOo 13	28
oo	TToo 0	Ttoo 5	ttoo 2	7
Total	0	26	20	46

T1-21 X 6B-1 #1 F_2 (4 Haploid)				
	TT	Tt	tt	Total
OO	TTOO 0	TtOO 7	ttOO 3	10
Oo	TTOo 0	TtOo 7	ttOo 8	15
oo	TToo 0	Ttoo 9	ttoo 9	18
Total	0	23	20	43

Table 3.10 Seed set in TToo, Ttoo, TtOO, ttoo and ttOO genotypes in three 6B/E F_3 progenies.

All seeds from a plant (Whole plant) were harvested from genotypes with low seed set (TToo, Ttoo and ttoo), while seeds from a given number of panicles (column 3) were collected from genotypes TtOO and ttOO and used to calculate seed set (last column).

T3-6 X 6E-7 #2 F_2 progeny				
F_2 Plant	Genotype	Panicles	Seeds harvested	Seed set
43	TToo	Whole plant	29	-
TToo – Average 29 seeds from each plant				
2	Ttoo	No seed set	0	-
7	Ttoo	No seed set	0	-
10	Ttoo	Whole plant	51	-
11	Ttoo	Whole plant	35	-
34	Ttoo	Whole plant	30	-
40	Ttoo	Whole plant	43	-
Ttoo – Average 40 seeds from each plant				
1	TtOO	5	74	30.2 %
3	TtOO	5	82	30.0 %
19	TtOO	5	64	29.4 %
23	TtOO	5	37	14.9 %
25	TtOO	No seed set	0	-
TtOO – Average seed set 26.1 %				
15	ttoo	Whole plant	68	-
35	ttoo	No seed set	0	-
42	ttoo	Whole plant	102	-
ttoo – Average 85 seeds from each plant				
4	ttOO	5	40	19.8 %
5	ttOO	No seed set	0	-
12	ttOO	5	121	49.6 %
18	ttOO	No seed set	0	-
26	ttOO	5	97	49.2 %
ttOO – Average seed set 39.5 %				

T3-6 X 6E-8 #1 F_2 progeny				
F_2 plant	Genotype	Panicles	Seeds harvested	Seed set
14	Ttoo	Whole plant	14	-
18	Ttoo	Whole plant	39	-
26	Ttoo	Whole plant	6	-
32	Ttoo	Whole plant	42	-
40	Ttoo	Whole plant	40	-
Ttoo – Average 29 seeds from each plant				
1	TtOO	5	48	22.6 %
6	TtOO	No seed set	0	-
10	TtOO	5	54	26.9 %
13	TtOO	5	64	36.2 %
38	TtOO	6	56	27.7 %
39	TtOO	No seed set	0	-
TtOO – Average seed set 28.3 %				
15	ttoo	Whole plant	38	-
37	ttoo	No seed set	0	-
ttoo – Average 38 seeds from each plant				
8	ttOO	5	86	42.2 %
27	ttOO	No seed set	0	-
30	ttOO	5	126	68.1 %
36	ttOO	5	134	52.1 %
42	ttOO	5	92	49.5 %
ttOO – Average seed set 52.9 %				

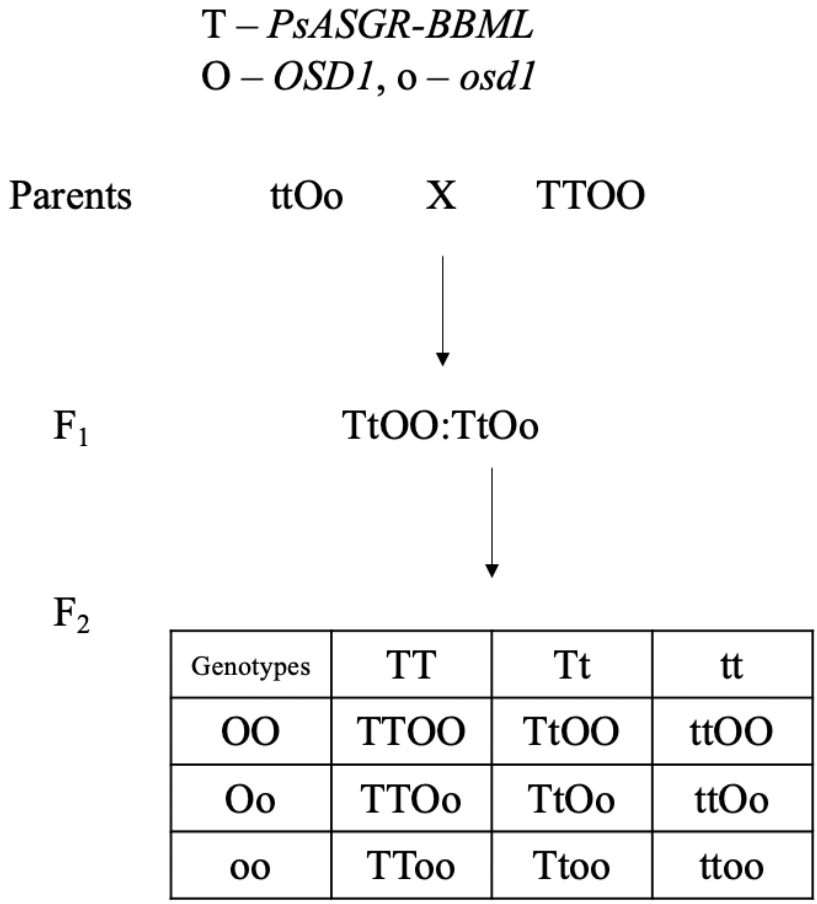
T1-21 X 6B-1 #1 F_2 progeny				
F_2 Plant	Genotype	Panicles	Seeds harvested	Seed set
3	Ttoo	Whole plant	51	-
22	Ttoo	No seed set	0	-
26	Ttoo	Whole plant	16	-
28	Ttoo	Whole plant	20	-
29	Ttoo	Whole plant	5	-
32	Ttoo	No seed set	0	-
35	Ttoo	Whole plant	18	-
39	Ttoo	Whole plant	30	-
43	Ttoo	Whole plant	15	-
Ttoo – Average 22 seeds from each plant				
1	TtOO	8	37	12.9 %
4	TtOO	6	36	19.8 %
17	TtOO	6	32	12.4 %
23	TtOO	7	40	13.4 %
34	TtOO	5	81	41.3 %
38	TtOO	No seed set	0	-
45	TtOO	7	60	17.2 %
TtOO – Average seed set 19.5 %				
5	ttoo	No seed set	0	-
8	ttoo	Whole plant	12	-
9	ttoo	Whole plant	94	-
11	ttoo	Whole plant	25	-
ttoo – Average 44 seeds from each plant				
12	ttOO	5	58	29.9 %
19	ttOO	6	76	35.0 %
33	ttOO	6	89	34.4 %
ttOO – Average seed set 33.1 %				

Table 3.11 Individual flow cytometry data on F_3 seeds or seedlings from 6B/E parents.

About 30 F_3 seeds or seedlings from 8 different F_2 plants were analyzed with individual flow cytometry for the determining the penetrance of parthenogenesis. 'Diploid' indicates parthenogenesis and 'Tetraploid' indicates sexual reproduction. Parthenogenesis is calculated according to formula = n(Diploid)/ [n(Diploid) + n(Tetraploid)]/*100.

F_2 progeny	F_2 plant (Genotype)	Flow on	Total	Tetraploid	Diploid	Un-germinated	Parthenogenesis
T3-6 X 6E-7 #2	#43 (TToo)	F_3 seedlings	19	1	16	2	94.1 %
	#10 (Ttoo)	F_3 seeds	28	14	13	1	48.1 %
	#40 (Ttoo)	F_3 seedlings	29	4	12	3	46.2 %
T3-6 X 6E-8 #1	#18 (Ttoo)	F_3 seedlings	31	12	13	6	52.2 %
	#32 (Ttoo)	F_3 seedlings	30	11	16	3	59.3 %
	#40 (Ttoo)	F_3 seeds	30	12	18	-	60.0 %
T1-21 X 6B-1 #1	#3 (Ttoo)	F_3 seedlings	30	14	13	3	48.1 %
	#9 (ttoo)	F_3 seedlings	30	27	-	3	0.0 %

Figures



Genotypes	TT	Tt	tt
OO	TTOO	TtOO	ttOO
Oo	TTOo	TtOo	ttOo
oo	TToo	Ttoo	ttoo

Figure 3.1 Crossing scheme to generate desired genotypes in F_2 progeny.

Plants heterozygous for *osd1* mutation (ttOo) were crossed with pollen from plants homozygous for *PsASGR-BBML* (TTOO). F_2 progeny was taken from TtOo F_1 plants and should theoretically have nine genotypes as shown in the box above.

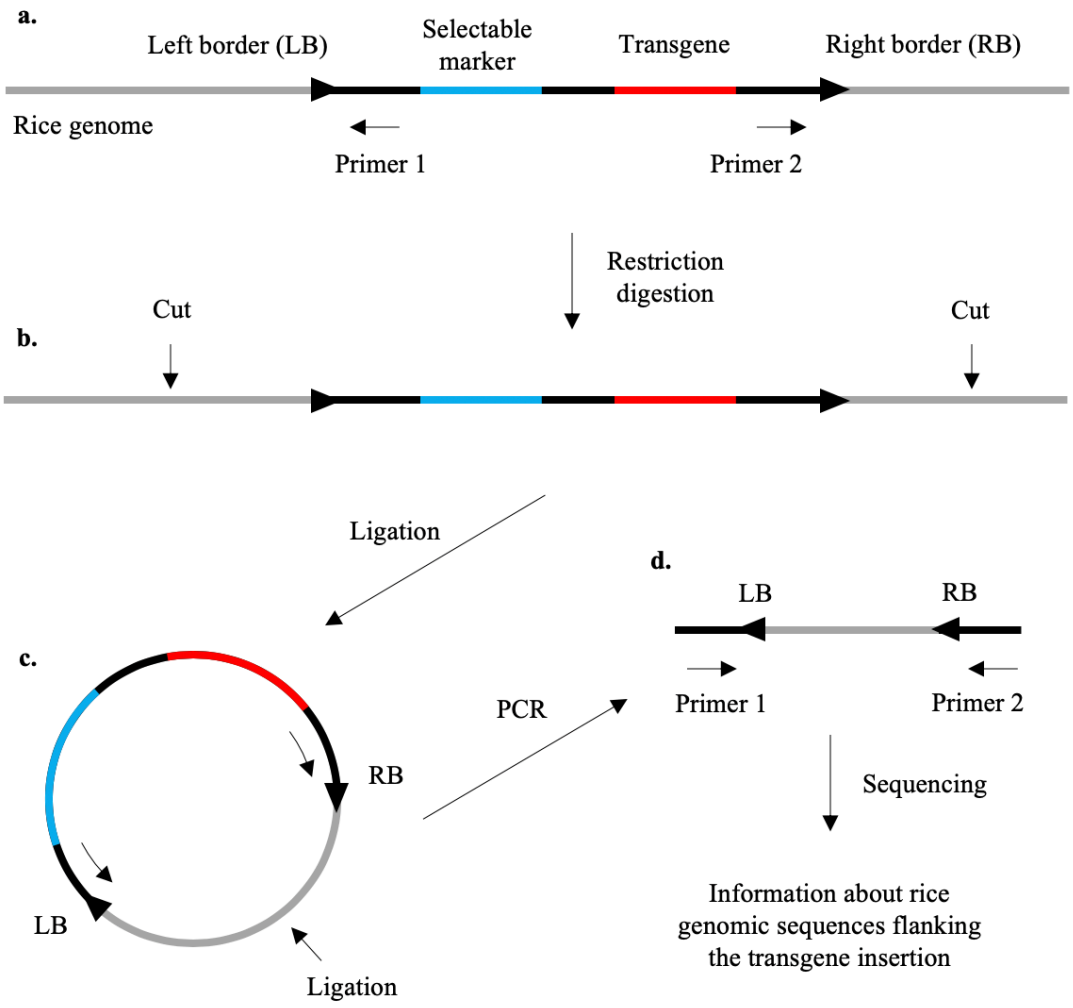


Figure 3.2 Workflow for inverse-PCR.

Inverse-PCR is used to identify the flanking sequence of the transgene insertion site, (a.) the expected transgene insertion where the sequence from transformation vector between the left-border and right-border will be inserted in the rice genome. Primers for i-PCR direct towards opposite directions, (b.) genomic DNA from transgenic plant is cut with a restriction enzyme cutting outside the inserted sequence followed by ligation (c.). Primer pair is used to amplify the flanking sequence and the amplified produce is sequenced providing us with the information about the sequences flanking the insertion site.

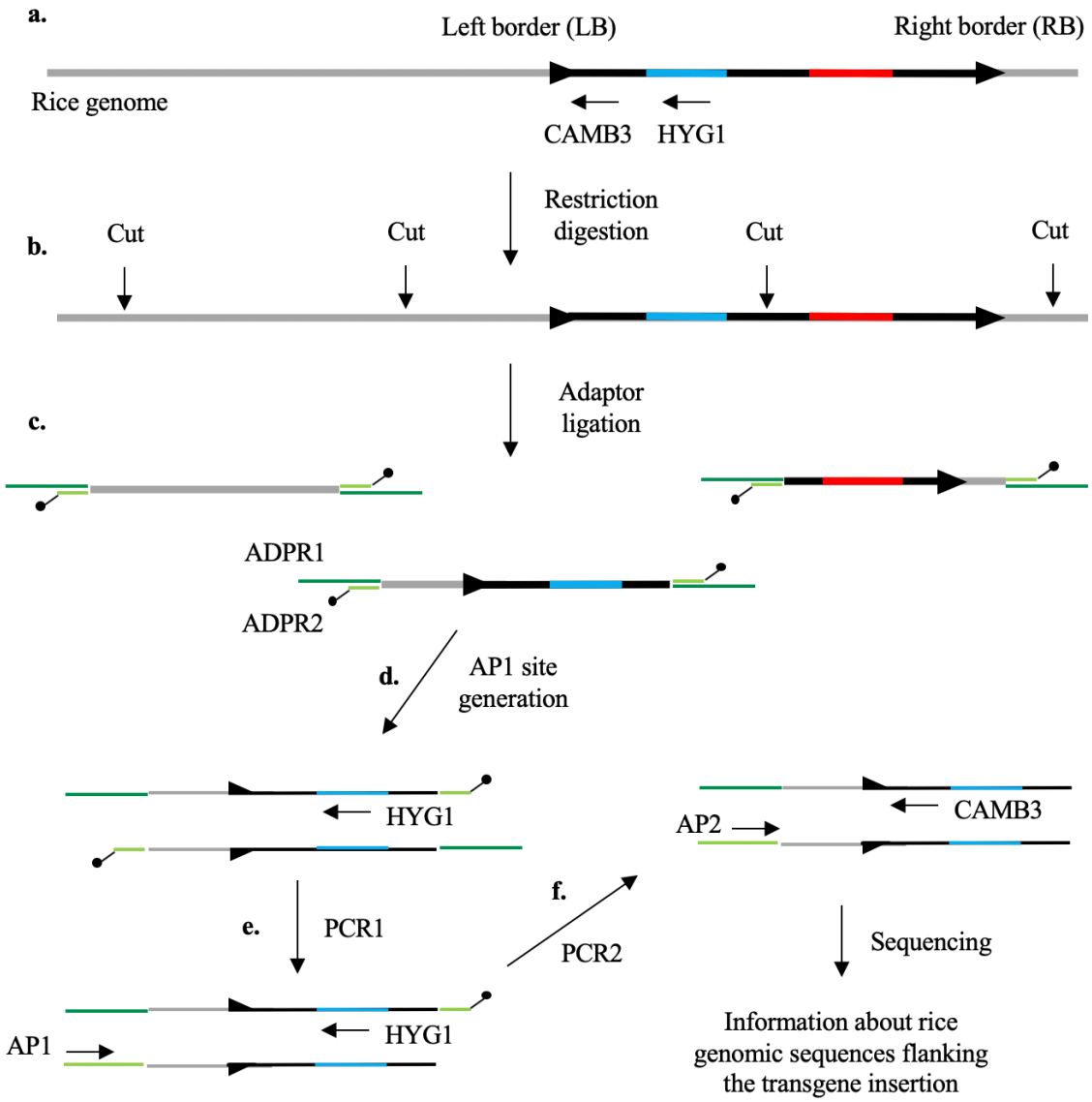

a.
Left border (LB)
Right border (RB)
Rice genome
CAMB3
HYG1
b.
Cut
Cut
Restriction digestion
Cut
Cut
c.
Adaptor ligation
ADPR1
ADPR2
d.
AP1 site generation
HYG1
AP2
CAMB3
e.
PCR1
f.
PCR2
Sequencing
AP1
HYG1
Information about rice genomic sequences flanking the transgene insertion

Figure 3.3 Workflow for left-border adaptor-ligation PCR.

Adaptor-ligation PCR is used to identify the flanking sequence of the transgene insertion site, (a.) the expected transgene insertion pattern, (b.) genomic DNA is digested with a restriction enzyme followed by (c.) adaptor ligation at the ends. The adaptor consists of long upper strand (ADPR1) and a short lower strand (ADPR2). ADPR1 sequence contains the successive sequences common to adaptor primers - AP1 and AP2. Only one sequence from the restricted, adaptor-ligated products will contain the desired left-border junction sequence, (d.) HYG1 primer will extend from the upper-strand and generate an AP1 primer binding site, (e.) primers AP1 and HYG1 will amplify the desired product in PCR1 and the amplified product will then be used for (f.) nested PCR 2 using primers AP2 and CAMB3. Final amplified product will be sequenced to get information about the sequences flanking the transgene.

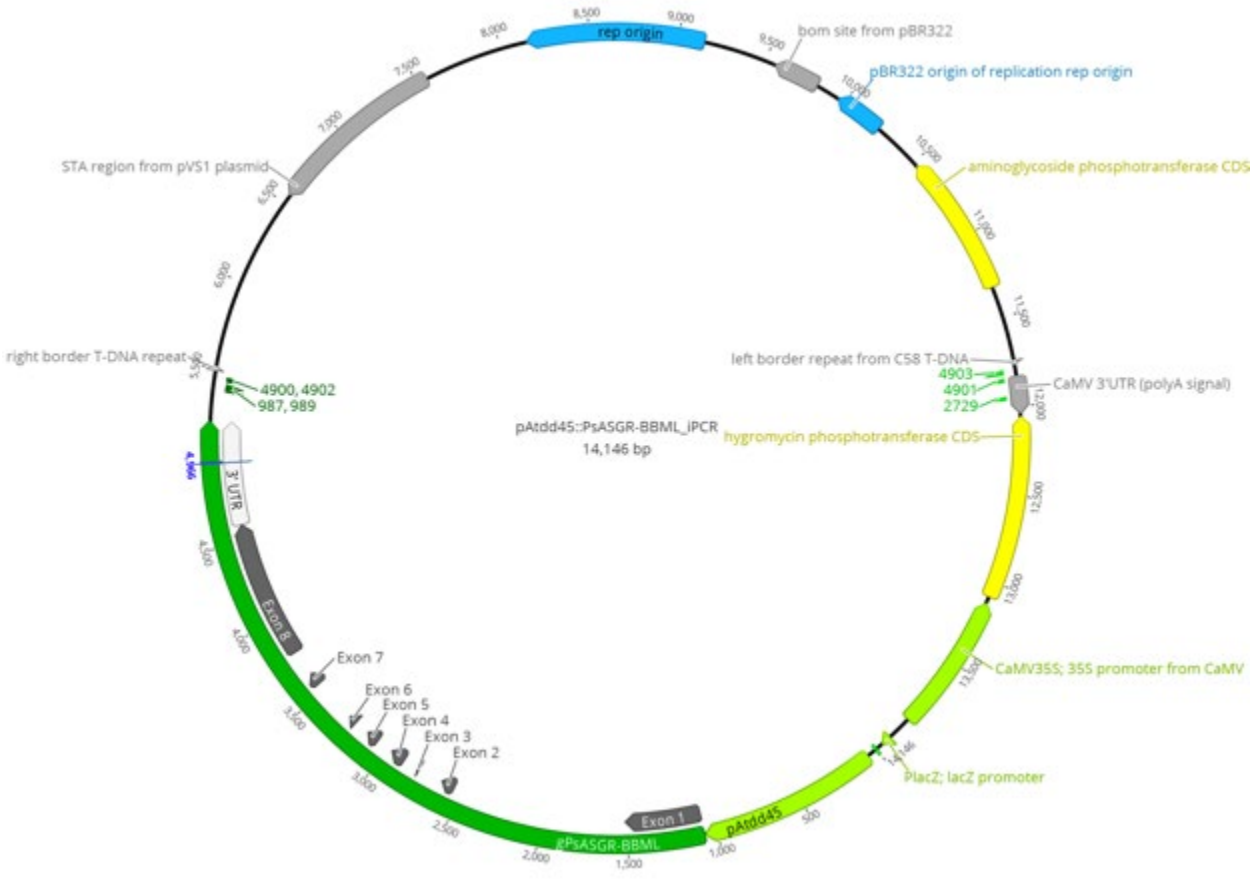


Figure 3.4 *AtDD45::PsASGR-BBML* transformation vector.

The *AtDD45::PsASGR-BBML* transformation vector is shown where four forward primers (987, 989, 4900, 4902) were designed left of the right border and three reverse primers (2729, 4901, 4903) were designed right of the left border for the purpose of i-PCR.

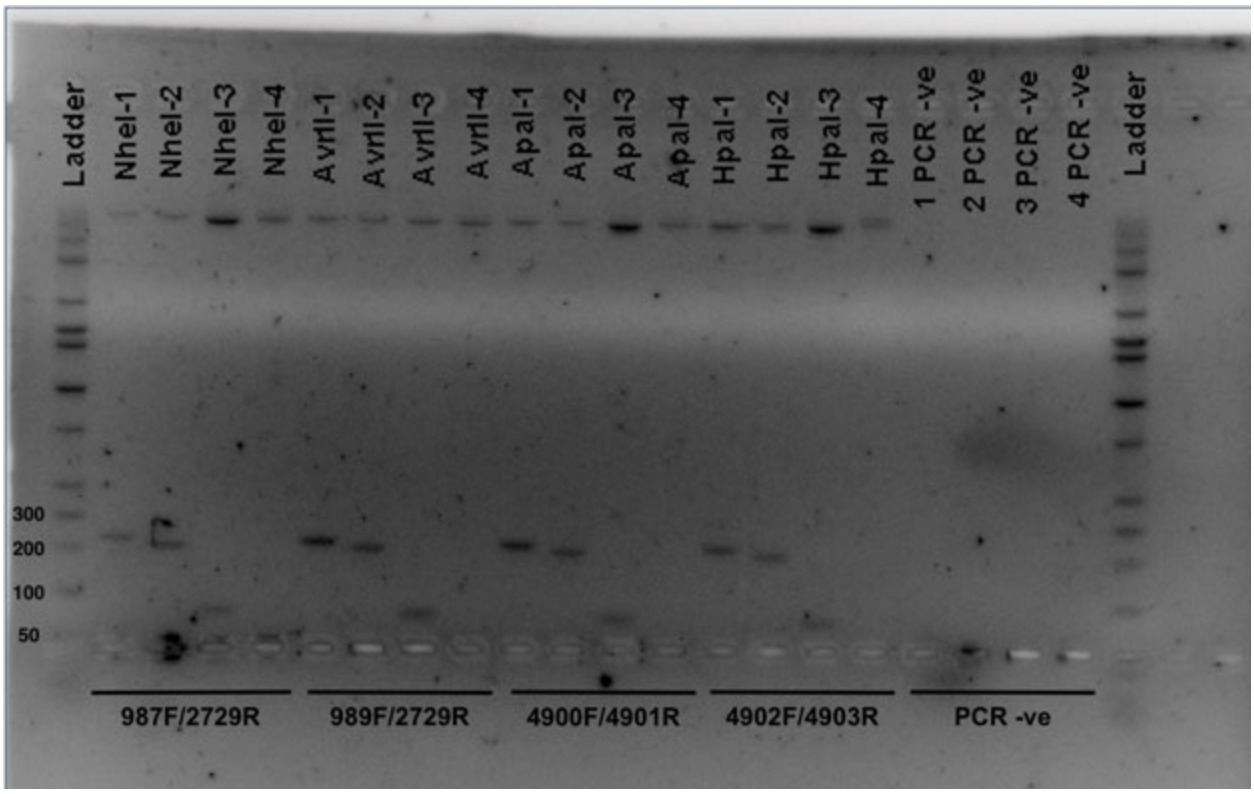


Figure 3.5 i-PCR results with four different primer-pairs.

Purified genomic DNA from R488 plant 11B-7 was restricted with four different restriction enzymes NheI, AvrII, ApaI and HpaI, followed by cirularization and PCR with four different primer-pairs – 987F/2729R (lanes 2-5), 989F/2729R (lanes 6-9), 4900F/4901R (lanes 10-13) and 4902F/4903R (lanes 14-17). Each primer pair was also used in a PCR negative reaction – 987F/2729R (lane 18), 989F/2729R (lane 19), 4900F/4901R (lane 20) and 4902F/4903R (lane 21). Lanes 1 and 22 represent the molecular ladder.

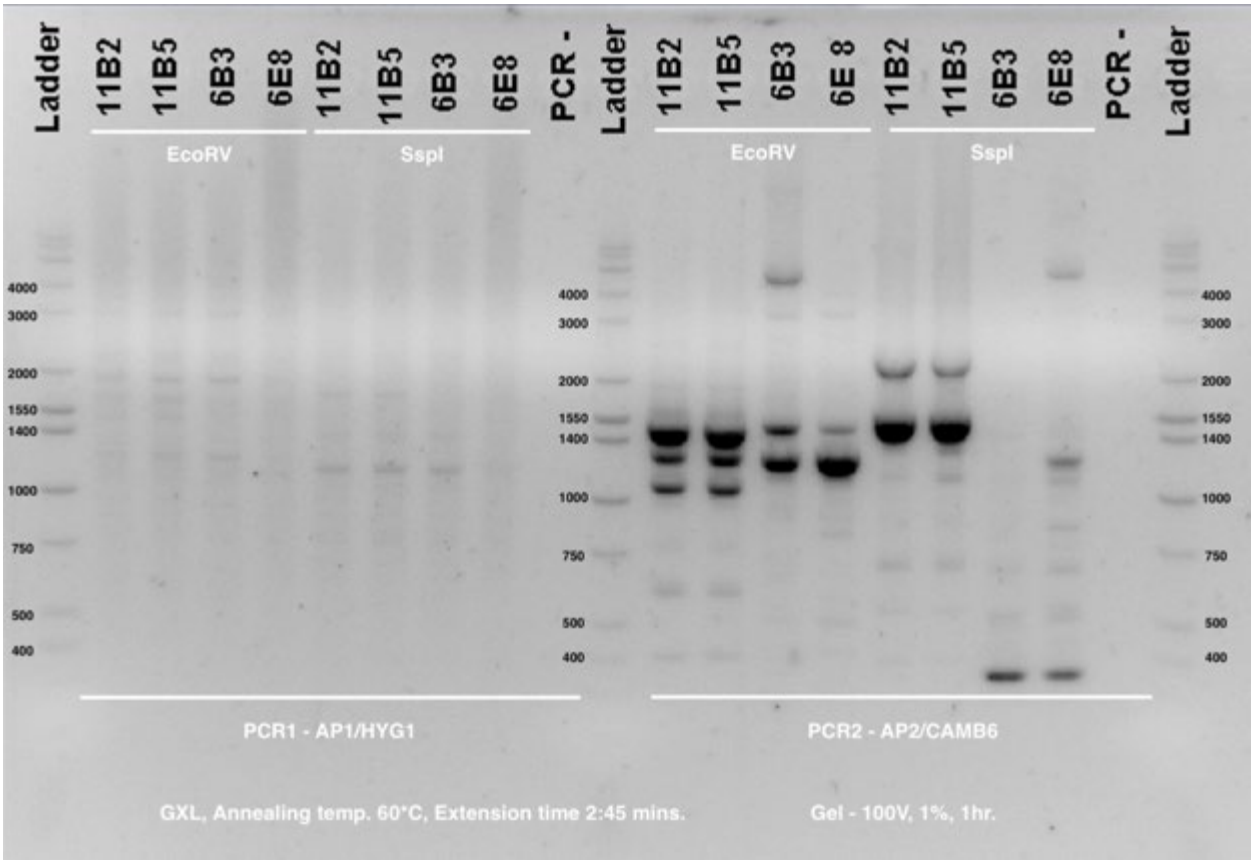


Figure 3.6 PCR1 and PCR2 results for adaptor-ligation PCR.

Genomic DNA from four R448 plants (11B-2, 11B-5, 6B-3 and 6E-8) were restricted with two blunt-end restriction enzymes – *Eco*RV and *Ssp*I, followed by adaptor-ligation and two subsequent nested PCR reactions. The results of these two reactions – PCR1 (left) and PCR2 (right) are presented in the above picture.

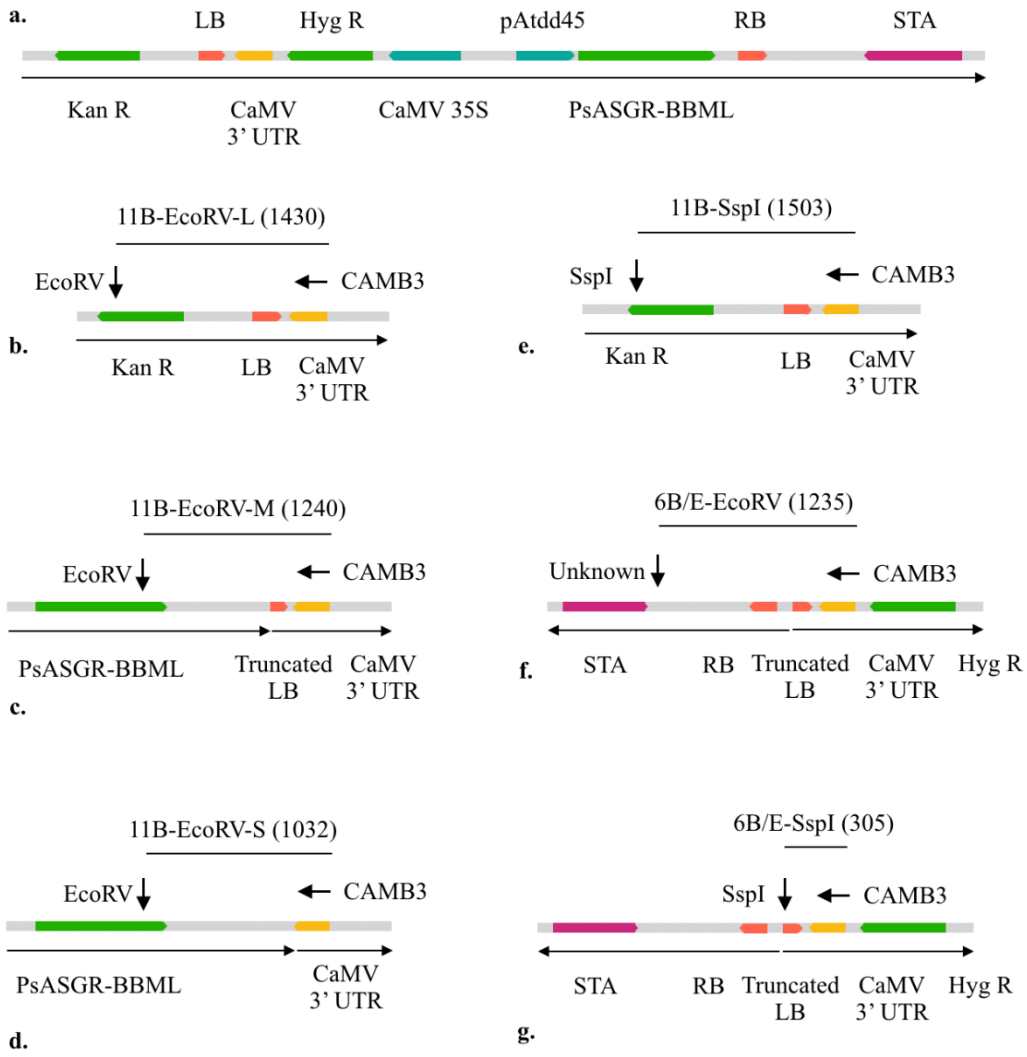


Figure 3.7 Sequencing results of the adaptor-ligation PCR.

(a.) a portion of *PsASGR-BBML* transformation construct. The expected insertion would be from left-border (LB) to right-border (RB) including selection marker gene (Hyg R) and transgene (*PsASGR-BBML*). The arrow below defines the expected integration direction. Insertion pattern in (b.) 11B-*Eco*RV-L, (c.) 11B-*Eco*RV-M, (d.) 11B-*Eco*RV-S, (e.) 11B-*Ssp*I, (f.) 6B/E-*Eco*RV and (g.) 6B/E-*Ssp*I.

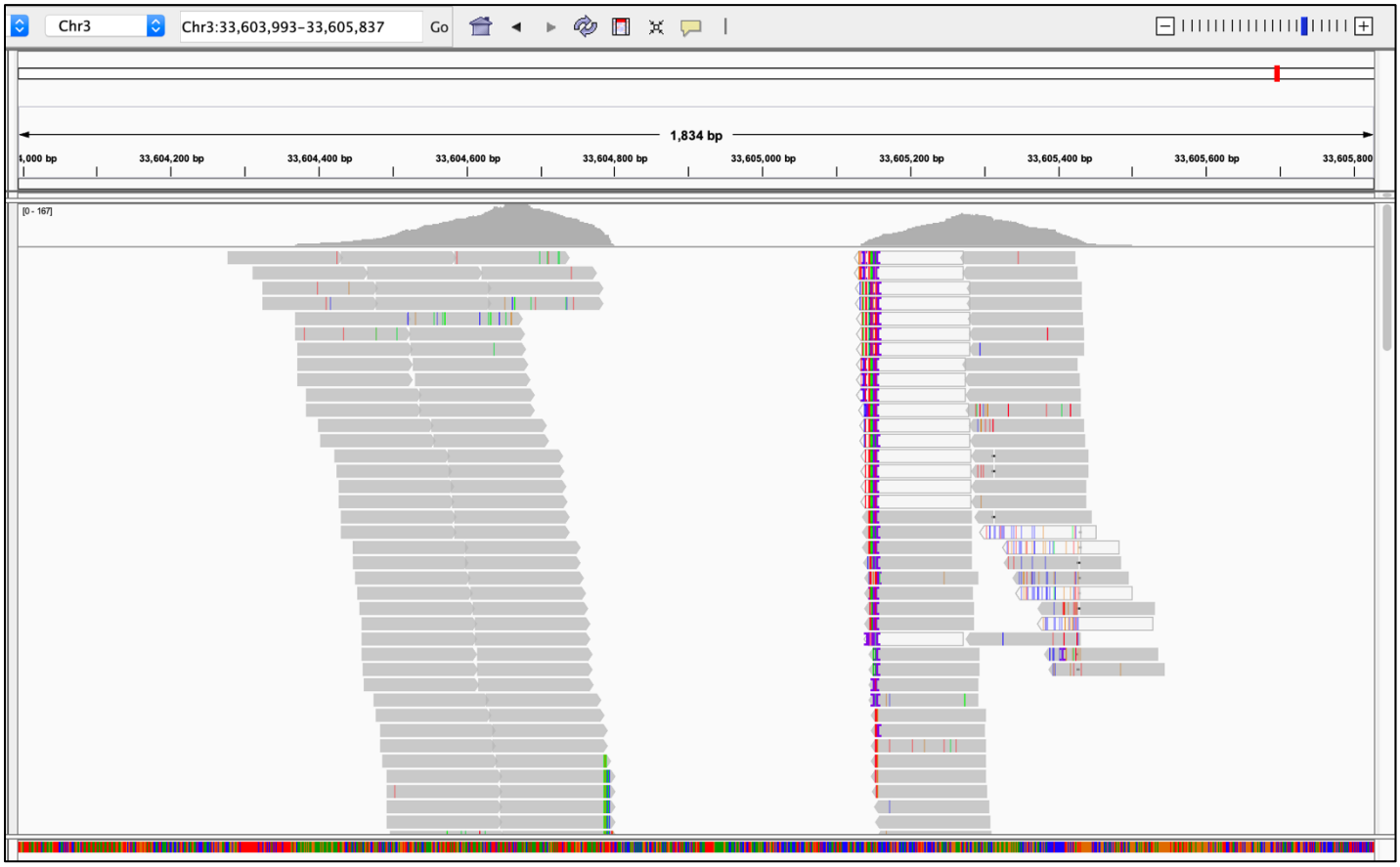


Figure 3.8 IGV screenshot showing some of the F264 reads mapping to rice chromosome 3.

Out of 6,916 F264 reads for 11B-4, 690 mapped to various rice chromosomes, out of which 400 mapped to Chromosome 3 at positions roughly from 33,604,280 to 33,605,550 bp. There is a gap of about 1,863 bp (33,603,294:33,605,157) between the reads mapping to opposite ends of this transgene insertion.

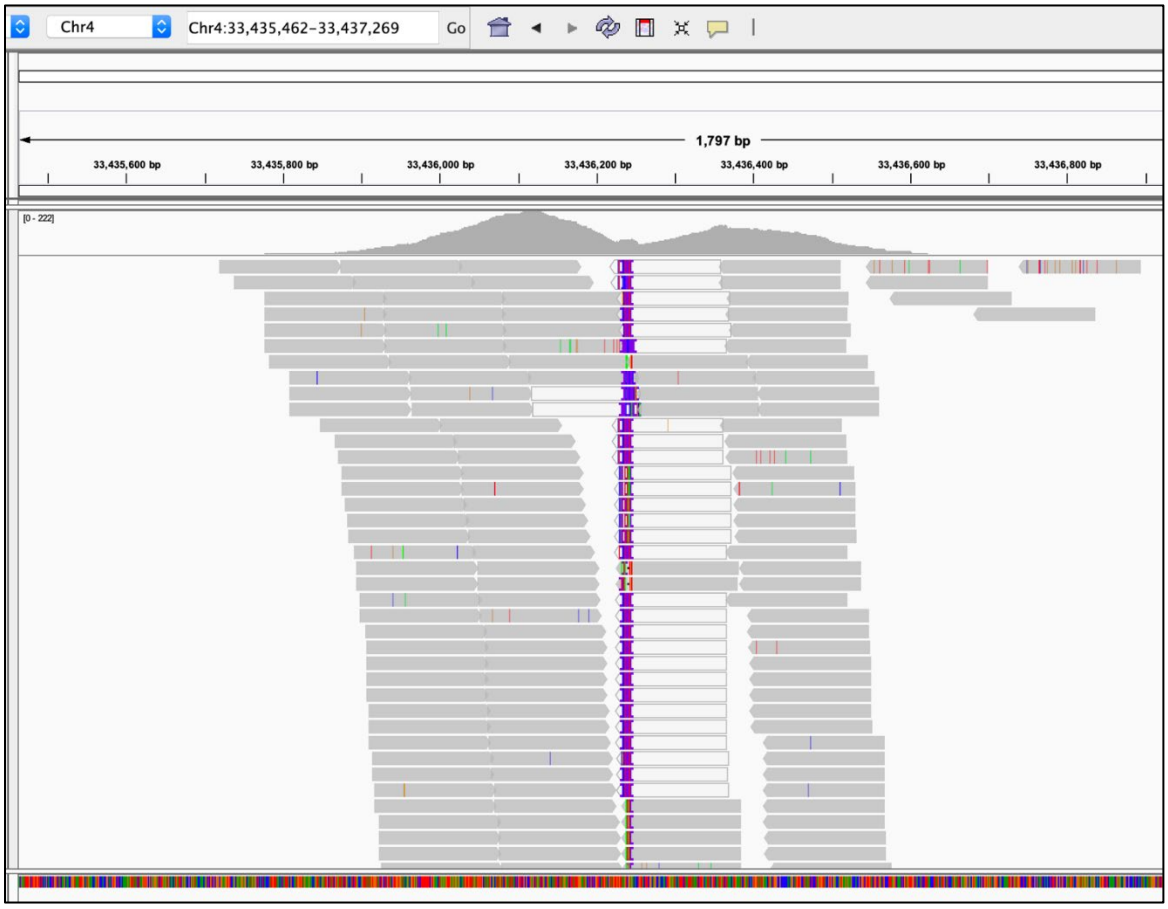


Figure 3.9 IGV screenshot showing some of the F264 reads mapping to rice chromosome 4.

Out of 1,437 F264 reads for 6E-7, 559 mapped to various rice chromosomes, with the maximum number of reads (507) mapping to Chromosome 4 at positions roughly from 33,435,610 to 33,436,900 bp. There was a negligible gap between the reads mapping to opposite ends of the transgene insertion.

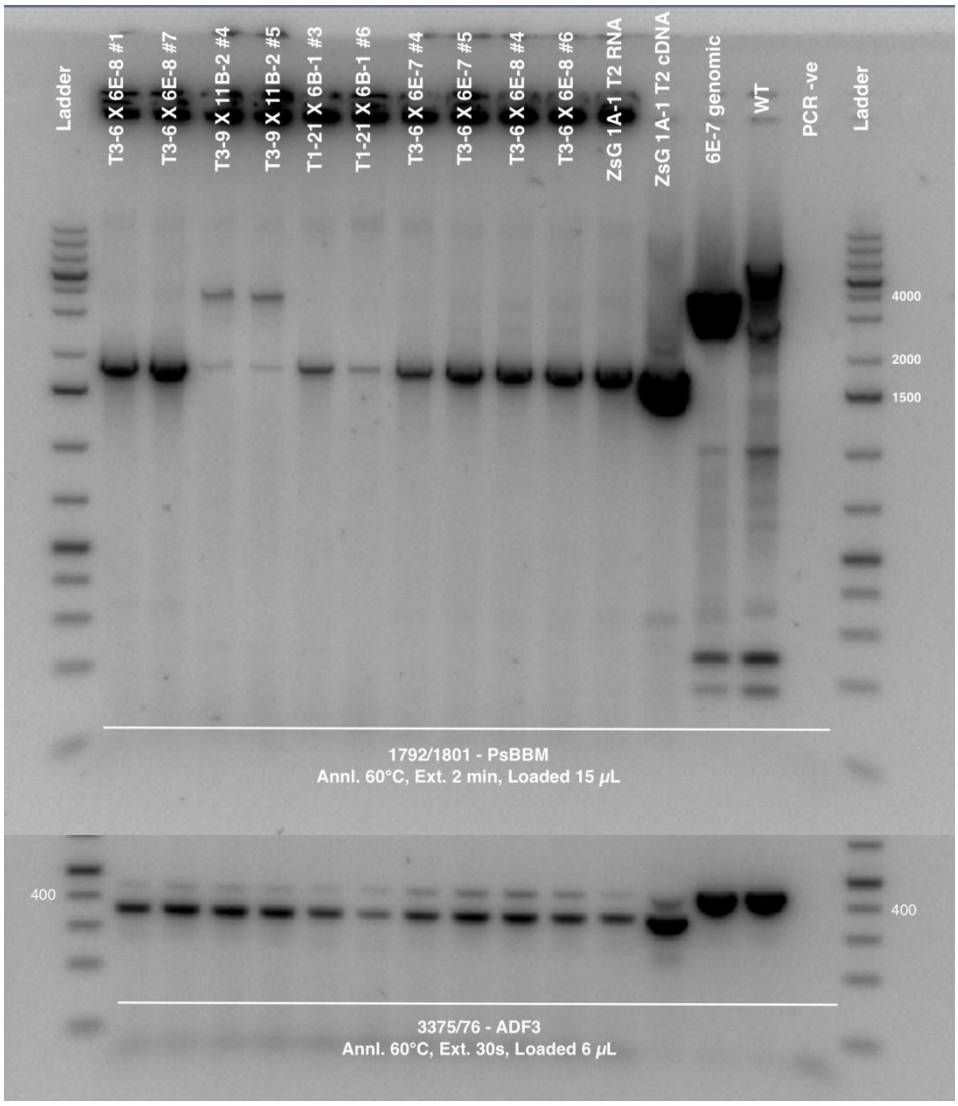


Figure 3.10 *AtDD45::PsASGR-BBML* expression analysis in F_1 plants.

AtDD45::PsASGR-BBML expression was determined through RT-PCR on RNA extracted from about 30 pre-anthesis ovaries from 10 different F_1 plants. ZsG 1A-1 T_2 was used as positive control; ZsG 1A-1 RNA sample was subjected to same RT-PCR reaction while for ZsG 1A-1 cDNA sample, cDNA synthesized in a previous reaction was used as a control only during PCR.

CHAPTER 4

SUMMARY

Doubled Haploid (DH) technology significantly shortens the process of inbreeding, which can take five to six generations with traditional methods (Dunwell, 2010). This technology involves haploid induction followed by chromosome doubling to generate homozygous doubled haploid lines. Haploids can be induced by *in vitro* methods where the gametophytic development of haploid spores is redirected to sporophytic development (Forster et al., 2007). Androgenesis, the process of regenerating haploid plants from male gametic cells using tissue culture, is widely used for haploid induction in rice (Mishra and Rao, 2016). However, high genotype dependency and the laborious nature of this technique limits its wider adoption (Forster et al., 2007). *In vivo* methods of haploid induction are based on uni-parental genome elimination, where the genome of one parent gets eliminated from the zygote formed by crossing two parents, or parthenogenesis, where egg cell initiates embryogenesis without fertilization. Uni-parental genome elimination can be achieved with the use of inducer lines, CENH-3 mediated haploid induction, wide hybridization or inter-specific crosses (Watts et al., 2018). Although these approaches produce transgene free haploids, the need for crossing makes it difficult to induce haploids in autogamous species such as wheat and rice (Jacquier et al., 2020). Parthenogenesis, which is found as a component of apomixis in nature, can be employed as a cross-free technique of haploid induction. In a natural apomict *P. squamulatum*, an AP2 transcription factor, *PsASGR-BBML* is shown to induce parthenogenesis and various *PsASGR-BBML* transgenes induce parthenogenesis in pearl millet (a close relative of

P. squamulatum), rice and maize (Conner et al., 2015; Conner et al., 2017). However, the dominant nature of *PsASGR-BBML* hinders its use in DH technology as the resulting lines will reduce back to haploids. This demands that *PsASGR-BBML* expression is regulated such that it can be turned on during haploid induction and kept off otherwise. In this study, we used a post-translational regulation system (Yamaguchi et al., 2015) involving the fusion of *PsASGR-BBML* to the hormone binding domain from rat Glucocorticoid Receptor (GR), such that the PsASGR-BBML:GR fusion protein remains inactive until the application of a synthetic steroid Dexamethasone (DEX). We used *pAtDD45*, an egg cell-specific promoter (Ohnishi et al., 2014; Steffen et al., 2007), to drive *PsASGR-BBML:GR* expression and tested three different methods of DEX application – through watering, *in planta* floral and *in vitro* floral treatment. Through flow cytometry and histology, we have shown that haploids can be induced by each type of DEX treatment. A self-pollinated line, homozygous for *AtDD45::PsASGR-BBML:GR* showed an haploid induction rate of 92 % and seed set of 17.2 % when treated with 100 µM DEX through watering at the flowering stage. Analysis of more homozygous lines is currently underway and will provide additional estimates of induction rate and seed set. Post-translational regulation can also be used for understanding the molecular mechanisms involved in *PsASGR-BBML* induced parthenogenesis by comparing RNA-seq profiles of DEX treated and untreated plants. GR specific antibodies can also be used for protein pull down studies to determine if PsASGR-BBML is part of a protein transcriptional complex and for performing ChIP-seq to identify its direct transcriptional targets. Understanding the molecular underpinnings of parthenogenesis will also help to engineer synthetic apomixis in crop plants.

Although many polyploids reproduce sexually, almost all natural apomicts are polyploid (Kaushal et al., 2019). Polyploid megaspore mother cells from these apomicts undergo apomeiosis to produce unreduced embryo sacs, possessing polyploid egg cells. The unreduced egg cells

undergo parthenogenesis to produce apomictic progeny at a high penetrance i.e., good seed set with all seed being clonal to the maternal parent. When diploid sexual rice was transformed with *PsASGR-BBML* (Conner et al., 2017), megaspore mother cells from these plants underwent meiosis to produce reduced embryo sac with the egg cells having a haploid genome. *PsASGR-BBML* induces parthenogenesis in these reduced egg cells at a lower penetrance as compared to natural apomicts and unfilled ovaries show abnormal developmental phenotypes (Conner et al., 2017). We speculated that one of several reasons for this can be the reduced ploidy level of egg cells in transgenic rice as compared to the unreduced egg cells in natural apomicts. We aimed to test this hypothesis by combining the *PsASGR-BBML* transgene through crossing with the rice *osd1* mutation (Mieulet et al., 2016) which produces unreduced egg cells. Higher penetrance of parthenogenesis in unreduced egg cells as compared the reduced egg cells will indicate that the former are more permissive to parthenogenesis than the latter. The *osd1* mutation was combined with two separate lines homozygous for the *PsASGR-BBML* transgene – 11B with a transgene insertion on chromosome 3 and 6B/E with a transgene insertion on chromosome 4. As in a typical dihybrid cross, we expected nine different genotypes in the F_2 progenies with a 1:2:1 segregation for both TT:Tt:tt (T - *PsASGR-BBML*) and OO:Oo:oo (O – *OSD1* and o - *osd1*). F_2 progenies from 11B parents segregated as expected for both transgenes. Progenies from 6B/E parents segregated normally for *osd1* but were highly skewed against the *PsASGR-BBML* transgene, with the Tt:tt ratio more closely fitting to 1:1. We obtained only one plant which was homozygous for the *PsASGR-BBML* transgene and producing unreduced egg cells (TToo). The observed unexpected segregation ratios are probably due to the inability of PsASGR-BBML transgene carrying ovaries to develop into seeds. This possibility can be tested by making reciprocal crosses among plants heterozygous for the *PsASGR-BBML* transgene (Tt) and WT (tt). Although seed set in F_2 progenies

was lower, we were able to perform individual flow cytometry on F_3 seed/seedlings. We observed a 94.1 % rate of parthenogenesis in one TToo plant while six Ttoo plants showed an averge parthenogenesis rate of 52.3 %. This suggests that variations in egg cell ploidy level and *PsASGR-BBML* transgene copy number can lead to different rates of parthenogenesis and an optimal level of *PsASGR-BBML* activity will be required to achieve full penetrance. Due to the high rate of parthenogenesis, F_3 progeny from TToo plant will mostly be diploid possessing the same genotype (TToo) and we plan to analyze the rate of parthenogenesis and seed set in these F_3 plants to get better estimates for these parameters.

References

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

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

Dunwell, J.M. (2010) Haploids in flowering plants: Origins and exploitation. Plant Biotechnology Journal, **8**(4), 377-424.

Forster, B.P., Heberle-Bors, E., Kasha, K.J., Touraev, A. (2007) The resurgence of haploids in higher plants, Elsevier Current Trends: pp 368-375.

Jacquier, N.M.A., Gilles, L.M., Pyott, D.E., Martinant, J.-P., Rogowsky, P.M., Widiez, T. (2020) Puzzling out plant reproduction by haploid induction for innovations in plant breeding. Nature Plants, **6**(6), 610-619.

Kaushal, P., Dwivedi, K.K., Radhakrishna, A., Srivastava, M.K., Kumar, V., Roy, A.K., Malaviya, D.R. (2019) Partitioning Apomixis Components to Understand and Utilize Gametophytic Apomixis. Frontiers in Plant Science, **10**(256).

Mieulet, D., Jolivet, S., Rivard, M., Cromer, L., Vernet, A., Mayonove, P., Pereira, L., Droc, G., Courtois, B., Guiderdoni, E., Mercier, R. (2016) Turning rice meiosis into mitosis. Cell Research, **26**, 1242-1242.

Mishra, R. and Rao, G.J.N. (2016) In-vitro Androgenesis in Rice: Advantages, Constraints and Future Prospects. Rice Science, **23**(2), 57-68.

Ohnishi, Y., Hoshino, R., Okamoto, T. (2014) Dynamics of Male and Female Chromatin during Karyogamy in Rice Zygotes. Plant Physiology, **165**(4), 1533-1543.

Steffen, J.G., Kang, I.-H., Macfarlane, J., Drews, G.N. (2007) Identification of genes expressed in the Arabidopsis female gametophyte. **51**(2), 281-292.

Watts, A., Kumar, V., Raipuria, R.K., Bhattacharya, R.C. (2018) In Vivo Haploid Production in Crop Plants: Methods and Challenges. Plant Molecular Biology Reporter, **36**(5), 685-694.

Yamaguchi, N., Winter, C.M., Wellmer, F., Wagner, D. (2015) Identification of direct targets of plant transcription factors using the GR fusion technique. Methods in molecular biology (Clifton, N.J.), **1284**, 123-138.