EXPRESSION ANALYSIS AND MAPPING OF THE *SEMIGAMY* GENE IN COTTON (*GOSSYPIUM BARBADENSE* L.)

A Thesis

by

ANDREA BORGES VIEIRA MAEDA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	David M. Stelly
Committee Members,	David H. Byrne
	Russell W. Jessup
Head of Department,	David D. Baltensperger

May 2015

Major Subject: Plant Breeding

ABSTRACT

Several naturally occurring apomicts rely on various forms of semigamous reproduction, where female and male nuclei fail to fuse after fertilization. Heritably high frequencies of haploids among monoembryonic seed were reported in 1963 in the cotton cultivar 'Pima S-1' (*Gossypium barbadense* L.). Later, the doubled haploid line '57-4' derived from 'Pima S-1' was developed. The combined production of haploids, maternal/paternal chimeras and zygotic progeny subsequently led to the hypothesis of reproduction by semigamy. Incomplete expression, gene dosage relationships and patterns of inheritance led to the hypothesis of control by a single, incompletely dominant gene, *Semigamy* (*Se*). The gene has a number of practical uses in breeding and science that would likely benefit from additional knowledge about the gene's location, mode of gene action, penetrance and expression. Our lab has studied the cytology, expression and chromosomal localization of *Semigamy*, but some results have been incongruous. Here, I re-map *Se* using new methods, markers and germplasm, and compare findings to previous results; I also characterize maternal versus paternal *Se* expression. Genetically marked reciprocal crosses between *Semigamy* homozygous and heterozygous plants allowed the tracking of allelic inheritance in maternal and paternal haploid sectors in chimeric progeny. SNP markers were developed for the *Semigamy* region, and then used

for genotyping sectors, linkage analysis, expression analysis and marker-assisted backcrossing.

Genotypes of haploid sectors and backcross individuals confirmed the SNP selection procedures and differentiated between previous *Se* mapping results. Crosses with female versus male *Sese* heterozygotes demonstrated markedly different effects on reproduction. Transmission of *Se*-linked SNPs from heterozygous females to normal progeny was random (1:1), whereas all haploid sectors had *Gb* (*Se*) alleles, indicating that semigamy requires megagametophytic expression of the *Se* locus (or lack thereof). *Se*-linked SNPs of haploid sectors from heterozygous males, on the other hand, occurred in ~3:1 distribution of *Gb* and *Gh* allele. The differences between male and female results was statistically significant ($p < 0.05$), but additional studies are needed to differentiate among possible causes for those differences.

DEDICATION

This work is dedicated to my mom, my dad and my husband who gave me all the love and support necessary to get here.

You are everything to me…

ACKNOWLEDGEMENTS

I would like to acknowledge all the wonderful people who directly or indirectly contributed to my growth during these years.

My lab mates and friends Amanda, Ammani, Chi, Jiale, Kevin, Mariana, Ting and Yu-Ming, it's funny how friendship develops in the most variable forms and situations. You will be forever remembered, simply because y'all are the best!

Bo, thanks for your friendship and late afternoon "chit-chatting" about all kinds of stuff in the lab, I enjoyed!

Fei I'm sorry for all the times I "made you nervous" with my "particular ability" to work around the lab with all my awkward clumsiness… You taught me so much and I'll be forever grateful!

Thanks to the student workers that without them, my life would be way more difficult. Clint, Keanan, Roger, Joshua and special thanks to Dustin! Seriously, you guys are awesome!

Wayne, many thanks to you! You're the best "plant-caretaker, planter, harvester, crosser, greenhouser, ginner, delinter, bud chooser, microscoper"… ever! All our "fights" were so small close to all our laughter! And Robert, the same extends to you! Thank you so much for continuing Wayne's work with such attention to details. I don't know what I would have done without y'all!

Jared, thank you for always being there for me when I needed to use USDA facilities!

Dr. Byrne and Dr. Jessup, thank you for being part of my committee! You certainly contributed to make my work better with your advices, suggestions and discussions.

Dr. Stelly, thank you for always find some time to answer questions and discuss my research. Semigamy is a challenging subject, but you always made me think outside the box. I'm so thankful for all the scientific advice, insightful discussions and suggestions. The most important thing that you carry with you in life is knowledge and I learned so much with you and I grew so much as a person and as a professional! Thank you so much!

I would also like to thank my husband Murilo and my friend Henrique for always offering a hand when I needed help with some of my field work. I know Texas summer is not the most inviting environment to work outside, but you always were there for me! Thanks guys!

. Thanks to all my family and Murilo's family, especially my dad, mom, brother, sister and my dearest parents-in-law, for all the support and late night Skype conversations that helped me so much staying focused emotionally to get through all the obstacles and difficulties of living in a different country.

All of you, I'm truly grateful for everything.

NOMENCLATURE

SNP	Single Nucleotide Polymorphism
Se	*Semigamy*
G. barbadense	*Gossypium barbadense*
G. hirsutum	*Gossypium hirsutum*
Gb allele	*Gossypium barbadense* allele
Gh allele	*Gossypium hirsutum* allele
Gh-Gb SNP	Single Nucleotide Polymorphism between *Gossypium hirsutum* and *Gossypium barbadense*

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Cotton (*Gossypium sp*.) has been cultivated in both tropical and subtropical regions of the world since prehistoric times. It was domesticated from wild woody perennial types and is now most widely cultivated as an herbaceous annual crop to meet human needs (Poehlman and Sleper 1995). Currently, it is the world's most important natural textile fiber crop, accounting for around 35 percent of total world fiber use. Also, cotton is a major crop in the food oil industry (www.usda.gov). Cotton is grown in both developed and developing countries, guaranteeing the income of millions of farmers around the world (Lacape, Nguyen et al. 2005). The United States, India and China collectively provide two-thirds of the world’s cotton production. The USA is the leading exporter and accounts for over one-third of global trade in raw cotton (www.usda.gov).

Including the two newly identified species (Grover *et al.* unpublished, Wendel *et al*. unpublished), there are 52 species recognized in the genus *Gossypium*, 7 being tetraploids (2n=4x=52) and 45 being diploids (2n=2x=26) (Han, Guo et al. 2004). The tetraploids originated from an ancient polyploidization event estimated by "molecular clocks" to have occurred 1-2 million years ago, in which an A-genome diploid species native to Africa and D-genome species native to Mexico hybridized followed by a chromosome

doubling event, resulting in an "AD" allotetraploid species that diverged evolutionarily into at least 7 extant species (Paterson, Wendel et al. 2012). Four species were domesticated, including two diploids, *Gossypium arboreum* and *Gossypium herbaceum*, and two tetraploids, *Gossypium hirsutum* and *Gossypium barbadense*. The tetraploids account for 90% and 5% of the world cotton production, respectively. *Gossypium hirsutum* has the biggest economic impact among all cotton species.

Plant breeding is one of civilization's oldest activities and started with the need of improving plants for human benefit. Today, the world's food production relies on breeding programs to create exceptional genetic types that excel in productivity to meet population needs. As a result of the expansion of breeding programs and an increase of demand by the world's population, breeders are always searching for new methods to increase production and convenient techniques to facilitate the breeding process. But, despite its significance, plant breeding is a relatively slow process that typically relies on hybridization, genetic recombination and/or inbreeding. To develop and bring to market new elite genetic types or "cultivars" typically requires multiple years. One of the most time-consuming steps in this process is inbreeding, which is often used to attain homozygosity and uniformity.

Apomixis

Sexual reproduction is an evolutionarily important trait that entails DNA recombination and independent assortment of most genes, and serves to maintain and create variability. However, circumventing sexual reproduction can be advantageous in certain situations, e.g., when breeders seek immediate homozygosity or to mass-produce a uniformly heterozygous genotype through seed. Some types of asexual reproduction or "apomixis" can lead to the formation of haploid progeny from a single gametic nucleus (Forster, Heberly-Bors et al. 2007). In disomic organisms, chromosome doubling of haploids immediately yields genetically uniform homozygous lines, which reduces the time otherwise required to achieve homozygosity through multiple generations of self-pollination. In addition, such homozygous lines are especially valuable for selection of quantitative traits by breeders, since high homozygosity promotes greater levels of additive variance and absence of dominance and segregation events (Griffing 1975, Snape 1989). Other types of apomixis lead to transmission of the entire maternal nuclear genome and could enable breeders to fix heterosis in seed-propagated hybrid crops, which could have a big economic impact worldwide (Asker and Jerling 1992). Apomixis could also be used for genetically isolating transgenic crops. An apomictic transgenic plant that is male-sterile or otherwise designed to be incapable of outcrossing with non-transgenic plants would be effectively isolated. These possibilities have

aroused great interest among plant scientists due to time- and cost-saving applications in plant breeding.

In 1908, Hans Winkler introduced the term "apomixis" as “replacement of sexual reproduction in which plant multiplication occurs asexually without cell and nuclear fusion” (Winkler 1908). The definition established by Winkler encompassed vegetative and seed-based modes of asexual reproduction, including recurring and non-recurring forms (Jacobs 2000, Pupilli and Barcaccia 2012). In most forms recurrent apomixis, meiosis does not take place or is modified such that the maternal genotype and chromosome complement are transmitted into a “2n” megaspore, megagametophyte or embryo; the resulting embryo is typically a clonal copy of the maternal parent. In non-recurrent apomixis, meiosis is usually normal and a haploid cell of the megagametophyte, usually the egg cell, gives rise to a haploid embryo. In non-recurrent forms, the gametophytes have half the number of chromosomes of the maternal sporophyte and the process does not repeat from one generation to the other (Maheshwari 1950, Stebbins 1941, Battaglia 1963). More recently, most authors have defined apomixis more narrowly, essentially equivalent to “agamospermy”, or asexual reproduction through seed (Hanna and Bashaw 1987, Richards 1997). Furthermore, some contemporary researchers, especially those working on recurrent forms, have equated apomixis with forms of agamospermy in which meiosis is functionally modified or entirely avoided and a genetically unreduced

egg cell develops without being fertilized (Koltunow 1993, Carman 1997, Richards 2003, Bicknell and Koltunow 2004).

Apomictic reproduction system is typically not obligate. It often occurs simultaneously with sexual reproduction and, in some cases, even within the same ovule. Thus, they are not mutually exclusive. Relative frequencies vary widely. Most of the apomictic plants today are considered "facultative", meaning that the same plant is capable of producing apomictic and zygotic seeds (Koltunow, Bicknell et al. 1995, Biddle 2006).

Mechanisms of Apomixis

In recurring forms of apomixis, embryos can originate gametophytically or sporophytically (Asker and Jerling 1992, Koltunow 1993, Richards 1997). In sporophytic apomixis, embryos develop by a process called adventitious embryony. In this process, one or more embryos form directly from the somatic cells and gametophytic generation is completely absent, such that the embryo is genetically and chromosomally unreduced relative to the maternal parent. There are two types of somatic tissues that can give rise to the embryo: the nucellus and the integument, the first being the most typical and well described in *Citrus sp*. (Wilms, Van Went et al. 1983, Wakana and Uemoto 1987). There are two gametophytic pathways that can lead to formation of apomictic embryos. Depending on the origin of the cells that form the megagametophyte, those apomictic mechanisms are classified in diplospory or apospory. In diplospory, a

genetically and chromosomally “2n” megagametophyte develops from the megasporocyte (megaspore mother cell) by either mitotic events or modified meiosis. In each diplosporic system, meiosis is typically interrupted or modified at a specific stage by unknown molecular mechanisms. In apospory, genetically and chromosomally unreduced “2n” gametophytes develop directly from nucellar somatic cells of the ovule and the unfertilized egg cells subsequently give rise to clonal embryos (Nogler 1984, Hanna and Bashaw 1987, Koltunow 1993, Jessup, Burson et al. 2002).

In non-recurrent apomixis (also called reduced parthenogenesis), the megasporocyte undergoes normal meiosis and a haploid cell of the subsequently formed megagametophyte, the egg cell (or non-egg cell in case of apogamety), develops into a haploid embryo (Stebbins 1941, Battaglia 1963, Jacobs 2000).

Apomixis in Modern Agriculture

Apomixis is already an important plant breeding tool and could prospectively revolutionize agricultural practices and economics. If reproductively efficient apomixis could be developed in or transferred to important crops, the possibility to select and clonally propagate individuals with superior characteristics would lead the world to a whole new type of agricultural system. The impact of increased productivity, efficiency and consistency of

agricultural production could be comparable to the Green Revolution (Calzada, Crane et al. 1996, Pupilli and Barcaccia 2012).

Despite the existence of more than 400 apomictic species widely spread among the plant kingdom, apomixis is not easily found in agricultural crops with rare exceptions like *Citrus sp*. and tropical forage grasses. There have been several attempts to introgress apomictic traits of non-cultivated species into related sexual crop plants via conventional breeding, but most of them were unsuccessful (Ramulu, Sharma et al. 1999, Spillane, Curtis et al. 2004). Recently, Akiyama et al. provided new insights on the origin, evolution and variations in “Apospory-Specific Genomic Region (ASGR) carrier chromosome” among *Cenchrus* and *Pennisetum* genera, the latter being the genus of *Pennisetum glaucum* (L.) or pearl millet. They suggested that apomixis originated once and spread among species by repeated hybridization (Akiyama, Goel et al. 2011). Such research leads to valuable information toward a better comprehension of apomixis.

Spillane et al. (2004) described the biological, ecological and economic challenges that apomixis researchers face. According to the authors, introducing recurrent apomixis into sexual crops is still problematic because (1) apomictic traits are absent or extremely rare in most crop gene pools, (2) epigenetic barriers (“parent-of-origin effects due to genomic imprinting, cytoplasmic effects or gene dosage sensitivity” can hinder asexual reproduction), (3) the potential of a dominant apomictic transgene to become uncontrollably widespread among

related species, and also (4) the concern among seed companies about intellectual property rights, since an apomictic seed could be reproduced for indefinite generations with no need to renew the seed stock (Spillane, Curtis et al. 2004). Nevertheless, the potential of apomictic crops is so significant that those issues should be rapidly addressed in order to develop an apomictic crop production.

Doubled Haploids

Haploids and doubled haploids (DH) have aroused great interest among plant scientists due to the many benefits they bring to plant breeding i.e., (1) they provide systems by which genetically uniform homozygous doubled haploid lines can be produced quickly, reducing the time required for inbreeding, (2) selection for desired traits or markers at the haploid or DH level affords significant genetic advantages because all subsequent gametes will reflect the selected genotype and will not segregate. Furthermore, the low additive variance and presence of dominance that usually hamper the efficiency of selection in early generations are solved by the use of doubled haploids where additive variance is increased and dominance is absent (Snape 1989), (3) the homozygous lines are well suited to facile gene and DNA marker mapping.

In 1922, Blakeslee et al. reported a haploid mutant for the first time in flowering plants. It occurred in *Datura stramonium* (Jimson weed) when they were testing cold temperature to induce chromosomal abnormalities. Later on, in

1924, during extensive research on *Datura* plants, Blakeslee and Belling reported the first doubled haploid plant (Blakeslee, Belling et al. 1922, Blakeslee and Belling 1924) and since then the potential of doubled haploids in plant breeding has been well recognized among the scientific community.

Doubled haploid technology has developed at a fast rate throughout the years and today it has been used in more than 200 species including major crop species like wheat, cotton, maize, citrus, soybean and barley (Maluszynski, Kasha et al. 2003). Typically, doubled haploid plant production is done by the creation of a haploid embryo either in vitro (tissue culture) or in vivo via genetic induction (parthenogenesis, pseudogamy, chromosome elimination or haploid-inducer lines) followed by chromosome doubling techniques usually involving a mitotic inhibitor, like colchicine (Forster and Thomas 2005).

In 1959, Coe reported a genetic strain (Stock 6) in maize that upon selfing, could produce as high as 3.2% of haploids (Coe 1959). Since then, several other haploid-inducers emerged with higher haploid-induction rates i.e., WS14 (Lashermes and Beckert 1988), ZMS (Chalyk 1994), MHI (Eder and Chalyk 2002), RWS (Roeber, Gordillo et al. 2005). Although the haploid induction rate might change according to which inducer line is used and the environment (Rober, Gordillo et al. 2005), currently, haploid-inducers in maize have a haploid induction rate of 8% or more (Melchinger, Schipprack et al. 2014). The development of a system in which haploid-inducer lines are used to produce haploids that can be easily identified at seed stage by an anthocyanin

color marker, previously used in potato called “embryo-spot” (Hermsen and Verdenius 1973), allowed a worldwide adoption of the doubled haploid technology in maize targeting a cost- and time-effective production of inbred lines (Prasanna, Chaikam et al. 2012).

In 2002, Liu et al. reported an efficient method of haploid production in wheat via induced microspore embryogenesis. In this report, up to 50% of the microspores in a spike were converted from gametophytic to sporophytic pathway by a chemical inducer. According to their results, 50 to 5500 green plants were produced by each spike from a wide diversity of wheat genotypes, making this system very promising for wheat research and breeding (Liu, Zheng et al. 2002).

In cotton, there has been interest in developing doubled-haploid breeding methods using the *Semigamy* (*Se*) gene of *G. barbadense* (Chaudhari 1979, Stelly, Lee et al. 1988) and also by creating a genetically engineered system based on the Ravi and Chan CenH3-replacement strategy (Ravi and Chan 2010, Van Deynze and Stelly, unpublished).

Semigamy

The term "semigamy", perhaps better called "hemigamy" or "gynandroembryony" (Battaglia 1980), was first coined by Battaglia in 1946 during research with *Rudbeckia spp*. Battaglia defined semigamy as a type of fertilization in which the sperm cell fuses with the egg cell, i.e., syngamy, but the

sperm nucleus does not fuse with the egg nucleus, i.e., karyogamy, and both nuclei undergo mitotic divisions independently. Therefore male and female cells coexist separately in the embryo (Battaglia 1946), which leads to high frequencies of mono-embryonic haploids and chimeric products composed of sectors of maternal and paternal origin (Turcotte and Feaster 1967, Chaudhari 1978, Zhang, Stewart et al. 1999, Biddle 2006) (Fig 1.1). After that, several authors reported the occurrence of semigamy in different species i. e., Coe reported semigamy fertilization while working with *Cooperia pedunculata* (Coe 1953), Turcotte and Feaster reported that haploid production in *Gossypium barbadense* was caused by semigamy (Turcotte and Feaster 1967), Rao and Narayana identify semigamy in *Coix aquatic* (Rao and Narayana 1980), Lanaud reported semigamy in *Theobroma cacao* L. (Lanaud 1988).

Figure 1.1. Monoembryonic chimeric products in cotton composed of maternal and paternal sectors.

The *Semigamy* gene was first reported in cotton by Turcotte and Feaster (1967). In 1963, haploid plants were found and reported by Turcotte and Feaster in a field of the Pima (*Gossypium barbadense*) variety 'S-1'. Later, upon chromosome doubling, a double haploid line (57-4) was created from those haploids and it started to produce a high frequency of haploids from single-embryo seeds, a phenomenon not yet described in cotton. Further investigations, including crossing the 57-4 line with different parental stocks and the production of chimeric progeny, led to the conclusion in 1967 that they were

observing the same semigametic fertilization phenomenon as Battaglia and Coe (Turcotte and Feaster 1963, Turcotte and Feaster 1967). In 1969, Turcotte and Feaster transferred the *virescent-7*, a recessive mutant gene for foliage color (*v7v7*), to semigamous line 57-4, developing a phenotypic marker for *Semigamy*. This marker allowed the identification of *v7* haploids and chimeric products when the semigametic line was crossed to any cotton plant with normal foliage color (*V7V7*) (Turcotte and Feaster 1969). The ability to produce chimeras in F1 plants and in their F2 progenies led to the report of semigamy as a heritable trait (Turcotte and Feaster 1975).

There is no reported natural occurrence of *Semigamy* gene in another cotton species but, as suggested by Chaudhari, it has the potential to be used in all cotton species for the means of haploid production. As mentioned by Turcotte and Feaster, Brown succeeded in producing haploids in *Gossypium klotzschianum* and *Gossypium tomentosum* for the first time via semigamy. After that, several cases of haploid production via semigamy in different species of cotton were reported (Turcotte and Feaster 1974, Chaudhari 1978).

Semigamy has been used in different types of research for its ability to produce haploid and chimeric progeny. Dolan and Poethig utilized periclinal and mericlinal chimeras produced via semigamy to study layer-specific expression of the Okra leaf shape in cotton, concluding that Okra ($L2^o$) mutant is active in all three tissue layers of the leaf (Dolan and Poethig 1998). Zhang et al. used haploid and doubled haploid populations originated from a cross between

Gossypium hirsutum and *Gossypium barbadense*, and Vsg semigamous lines to construct an allotetraploid cotton molecular linkage map with SSRs and RAPDs (Zhang, Guo et al. 2002). Several other researchers analyzed the stability and performance of doubled haploids produced via semigamy in cotton and all concluded that doubled haploids are as competitive and as stable as the cultivars from which they originated (Feaster and Turcotte 1973, Chaudhari 1979, Mahill, Jenkins et al. 1984).

Previous studies have established that semigamous lines commonly produce between 30 and 60 % haploids after self-pollinations or crosses with another semigamous line, and about 0.7 to 1.0% haploids when crossed as female parents with non-semigametic cotton lines (Turcotte and Feaster 1967, Chaudhari 1978, Chaudhari 1979). Haploids were never observed when non-semigamous lines were used as females in a cross with semigamous lines, suggesting a possible type of maternal effect (Biddle 2006). When heterozygous for Semigamy (*Sese*) were self-pollinated, F2 population segregated in a 3:1 of haploid-producing:non-haploid-producing plants. When heterozygotes were backcrossed to a non-semigamous line, the backcrossing population segregated in a 1:1 ratio of haploid-producing:non-haploid-producing plants. However, crosses between homozygous *Se* lines, i.e., *SeSe* x *SeSe*, gave rise to a higher haploid frequency than *SeSe* x *Sese* or *SeSe* x *sese* crosses (Biddle 2006). Given those observations, *Semigamy* was characterized as a partially dominant gene (Turcotte and Feaster 1974, Chaudhari 1978, Zhang, Stewart et al. 1999).

Before semigamy, haploids in cotton were primarily obtained through polyembryony (Chaudhari 1978). Polyembryony is the general term for a variety of mechanisms in which more than one embryo is present where only one would be expected and often, one of these embryos is haploid (Batygina and Vinogradova 2007). Most, if not all, plants spontaneously produce haploids in this manner, but only at a very low level, making it inefficient for use in research and requiring the use of very large populations to get a small number of haploids. By using a semigamous system, researchers can use a much smaller number of plants to obtain an equal number of haploids, increasing the efficiency and efficacy of their work (Chaudhari 1978). Today, doubled-haploid lines in cotton are usually derived via semigamy followed by chromosome doubling (Turcotte and Feaster 1967, Turcotte and Feaster 1974, Chaudhari 1979, Stelly, Lee et al. 1988).

Despite its practical use by the scientific community, to this date, the precise mechanism and genomic location of the *Semigamy* gene remain unclear. Further knowledge about gene localization, its mode of gene action and expression could be very valuable for breeding programs, since the gene can be potentially used for mass production of doubled haploids and to facilitate genetic mapping.

Semigamy and Apomixis

Semigamy has been related to apomictic systems by many authors (Battaglia 1946, Coe 1953, Battaglia 1955, Nogler 1984, Stelly, Lee et al. 1988, Wendel, Stewart et al. 1991). The production of maternal and/or paternal haploid progeny is explained by the occurrence of syngamy with no nuclear fusion (karyogamy) between the parental gametes. Both sperm and egg nucleus divide independently during mitotic divisions in the egg cytoplasm. (Battaglia 1955, Stelly, Lee et al. 1988). Asker and Jerling described semigamy (hemigamy) as "a peculiar form of pseudogamy" where the egg cell has to be penetrated by the sperm cell in order to stimulate development even though egg and sperm nuclei do not fuse (Asker and Jerling 1992). Coe pointed out that semigamy would be better described as an additional feature of some pseudogamous apomictic species and should not be considered an exclusive type of apomixis (Coe 1953). Nonetheless, this process is genetically reminiscent of non-recurrent apomixis in which a reduced cell is stimulated to give rise to a haploid embryo with the same characteristics as one of the parents.

Even in a non-recurring form, to have a naturally occurring apomictic-like gene in such an important crop like cotton is very promising for the future of agriculture, whereby a haploid producing gene have the potential to be artificially introgressed to several other crops enabling the propagation of selected desirable traits to the next generation with no mixture of parental types. Furthermore, if combining of semigamy trait with efficient production of

genetically unreduced “2n” gametes becomes feasible, a whole new model of crop production could be developed, where semigamy could be compared to a recurrent form of apomixis.

The challenge ahead is that this gene still remains poorly understood and most of its mechanisms are unclear, limiting its manipulation. A deeper comprehension of the *Semigamy* gene’s mode of action and expression would address the questions about instability of haploid production, especially in reciprocal and interspecific crosses. In addition, the localization of the gene to a well-characterized genetic marker region in the cotton genome would enable reliable marker-assisted selection within a segregating population as well as taking one step further towards the possibility of cloning the gene for use in different agronomic crops.

CHAPTER II

LOCALIZATION OF THE *GOSSYPIUM BARBADENSE SEMIGAMY* GENE AND DEVELOPMENT OF *SE*-LINKED SNP MARKERS

Introduction

Genetic markers are not new to the scientific community. Mendel's notable experiments in the nineteenth century for example, are well known for their reliability and were based on genetic markers, although phenotypic not molecular. To meet the need for rapid and more reliable ways to identify desirable traits, scientists developed molecular DNA markers, i.e., directly related to the genetic material of the organism. This revolutionized scientific research since those markers are not affected by the environment and they can be found in all tissues and at any life stage of the organism (Agarwal, Shrivastava et al. 2008). Genetic resources in cotton such as molecular markers, genomic sequences, linkage maps and bacterial artificial chromosomes (BACs) are valuable tools for genomic sequence analysis and assembly. The first comprehensive genetic map of cotton was assembled using restriction fragment length polymorphisms (RFLPs) and it was applied to the analysis of cotton chromosome organization and evolution (Reinisch, Dong et al. 1994). Among the other various types of molecular DNA markers developed and mapped in cotton are SSRs, AFLPs and SNPs.

Single nucleotide polymorphisms (SNPs) are the most abundant type of marker and are found throughout genomes. They are characterized by variations of a single nucleotide among the genomes of individuals of a population (Agarwal, Shrivastava et al. 2008). SNP markers have become popular among scientists for their relative ease of use, amenability to large-scale analyses and their diverse applications, including construction of high-resolution linkage maps.

Semigamy (Se) arose as a spontaneous mutant of the cotton species *Gossypium barbadense* (Turcotte and Feaster 1963, Turcotte and Feaster 1967). In cotton, this trait leads to an array of different kinds of embryos, including chimeras, maternal/paternal haploids and normal zygotes (Turcotte and Feaster 1967, Chaudhari 1978, Zhang, Stewart et al. 1999). When semigamous fertilization occurs in cotton, the gametes undergo syngamy (fusion) but skip karyogamy (nuclear fusion) and the non-fused maternal and paternal nuclei of the zygotic heterokaryon undergo mitotic divisions more or less independently (Hodnett 2006, Biddle 2006).

The construction of a high-resolution linkage map around the *Semigamy* gene in the cotton genome would likely benefit not only cotton breeders, but also the entire plant breeding community. It would lead to a better characterization of the *Semigamy* allele and enable marker-assisted selection of semigamous plants in cotton. In addition, mapping the *Se* gene would facilitate its cloning and increase the likelihood of the gene or its sequence information being used to

modify reproductive behavior in other economically important species, e.g., for producing doubled haploids or engineering recurrent apomicts.

In previous work, our lab crossed homozygous *Semigamy* lines from *Gossypium barbadense* species as the pollen parent to a set of monosomic and monotelodisomic plants of *G. hirsutum*, of which the chromatin deficiencies collectively covered approximately 60% of the cotton genome. The corresponding hypoaneuploid F1 hybrids were cytologically identified and backcrossed to *Se* lines. The *Se*-locus genotypes of multiple backcross progeny from each hypoaneuploid were determined phenotypically by cytological analysis of each plant's reproductive behavior during fertilization (Biddle 2006). A larger interspecific backcross population from *Sese* * *SeSe* was classified cytologically and also genotyped with SSR markers (Stelly Lab, unpublished data). Close linkage was detected between the *Semigamy* locus with a couple of SSRs in chromosome 12, but segregation distortion in the mapping population was so extreme that the results were regarded with strong apprehension. Therefore a radically different experimental approach was developed to map the *Semigamy* gene. SSR analysis was combined with evaluation of haploid sectors of chimeras, specifically the haploid sectors from gametes of *Sese* parents. Given heterozygosity for nearby SSR loci, the *Sese* parent would concomitantly segregate for *Se* and the markers. Preliminary tests of this method by another former student, Leslie Kendall, were limited in number, but seemed to confirm its utility and that the *Se* gene indeed resides in the long arm of chromosome 12 of

G. barbadense (Stelly lab, unpublished data). Because the *Se* gene does not behave in a "normal" manner, i.e. haploid frequency is not stable in reciprocal crosses among semigamous and non-semigamous parents, and because early work on this gene in the 1980s by Gwyn had indicated that the locus might exist in chromosome 4 (Gwyn and Stelly 1990), it was deemed especially important that our more recent mapping results (localizing *Se* to chromosome 12) be well confirmed by at least two or three independent methods. With that goal in mind, this research work has the following objectives:

1. Develop SNP markers near *Se* that would enable analyses that are time- and cost-effective, e.g., for marker-assisted selection and analysis of gene action.

2. Independently map the *Se* gene relative to SNP markers by linkage analysis of a newly created population of haploid segregates, and then compare to previous mapping results.

Materials and Methods

SNP Development

From the data available at Cotton Microsatellite Database (Blenda, Scheffler et al. 2006), SSR markers believed to be at or flanking the *Se* region were selected, and then mapped according to sequence similarity by BLASTn analysis along the D5 reference genome of *G. raimondii* (Paterson, Wendel et al. 2012). Then, the physical locations of those SSR markers were used to target

a BWA Next-Generation Sequencing approach to identify SNPs between *G. hirsutum* and *G. barbadense* in that same genomic region. Primers were designed for all of those SNPs and tested by KASP assays of a screening panel (*G. hirsutum* TM-1, *G. barbadense* 3-79, F1, and several *G. hirsutum* x *G. barbadense* RILs) to detect polymorphisms between *G. hirsutum* and *G. barbadense* (*Gh-Gb* SNPs). This strategy was considered sufficient to identify SNPs at high level of proficiency for the semigamous germplasm relative to *G. hirsutum*, because prior research during SNP development revealed that virtually all polymorphisms between *G. barbadense* and *G. hirsutum* are common to all interspecific genotypic combinations, i.e., any *G. barbadense* versus any *G. hirsutum* (Stelly, personal communication). To test and verify the applicability of these *Gh-Gb* SNPs to the research materials, the respective genotypes were determined for the semigamous parental lines as well.

Se **Localization**

To verify the linkage with *Semigamy*, the four best primers were tested in haploid sectors of chimeric seedlings originated from a cross between a F1 population heterozygous for *Semigamy* (*Sese*) and a semigamous line (*SeSe*). Different leaf colors were used as a phenotypic marker to detect parent-of-origin of the haploid sectors. For a second verification, a G*. hirsutum* x *G. barbadense Se* heterozygous F1 hybrid (*Sese*) was crossed as the pollen parent with a population of *G. hirsutum* red lines (R1 TM-1 and DeRidder Red). Seeds were

germinated to establish the backcross population (BC1F1) and screened with the four previously used SNPs. For genotyping, DNA was extracted from cotyledonary tissue of BC1F1 non-germinated seeds using a non-destructive high-throughput DNA extraction method developed by Zheng et al. (in press). Next, 2 µL of diluted DNA were loaded into wells of 96-well plates and dried in the oven at 60ºC for 1 hour. After optional storage of plates containing dried DNA samples, 8µL of PCR mixture (containing 4.0µL of Reaction Mix including polymerase, 3.8µL of sterile deionized water, 0.11µL of Primer Mix, and 0.064µL of 50mM MgCl2) was added to each well. After short centrifugation, the plates were submitted to a thermocycler machine (Eppendorf Mastercycler Ep Gradient S Thermal) set according to KASP assay recommendations: a preliminary acclimation step of 94°C for 15 minutes followed by 10 cycles at 94°C for 20 seconds, then 65°C for 1 minute decreasing 0.8°C per cycle to an annealing temperature of 57°C for the final cycle. This was followed by 28 cycles of denaturation at 94°C for 20 seconds, and annealing at 57°C for 1 minute. More cycles were added when additional amplification was required (Zheng et al. in press). The plates were briefly centrifuged and then submitted to a Pherastar plate reader (BMG LABTECH). Reads were imaged and interpreted using KlusterCaller software (LGC Genomics). Seed that were genotyped as heterozygotes (*Sese*) by the SNPs were planted and grown in the field.

A phenotypic test was performed to confirm the genotypic results given by the SNP markers. BC1F1 segregates identified as prospectively heterozygous

(*Sese*) according to SNP genotyping were testcrossed as females to *Sev7,* a semigamous line with a yellow-green leaf phenotype. The testcross seed were harvested, dried in a seed-drying oven (3 days at 38ºC), cold-treated (5 days at 5.5ºC to 8.8ºC, 12% to 15% humidity) and then planted in 25.4 x 50.8-centimeter trays in the greenhouse. Seedlings were screened visually for chimerism (red/green/yellow-green) at approximately 1-2 weeks after germination.

To obtain additional linkage data, the above *Sese* BC1F1 plants were backcrossed to *G. hirsutum*. BC2F1 seedlings were SNP-genotyped to identify prospective *Sese* segregates, and these were backcrossed (winter greenhouse of 2013-2014) to create BC3F1 seed. About 75 BC3F1 seeds were randomly selected, planted in Jiffy-7® pellets in the greenhouse and seedlings were genotyped to identify prospective *Sese* heterozygotes. The BC3F1 plants that were recombinant among the SNPs were testcrossed as females to a semigamous line (*Sev7*). Seeds derived from the crosses were planted in the greenhouse in 25.4 x 50.8-centimeter trays to identify chimeras among the seedlings.

Results and Discussion

SNP Development

Thirty SNPs between *G. hirsutum* and *G. barbadense* were identified computationally as marking sequences in the same genomic region as the *Se*-linked SSR markers, according to assembly to the D5 genome. We were able to

design primers for 25 of the 30 SNPs. The 25 primers were tested using a screening panel containing DNA from *G. hirsutum* TM-1, *G. barbadense* 3-79, F1, and several *G. hirsutum x G. barbadense* RILs. The primers were subjected to selection on the basis of their ability to detect polymorphism between the parents and their ability to cluster samples of the same genotype on the KlusterCaller software (Figure 2.1).

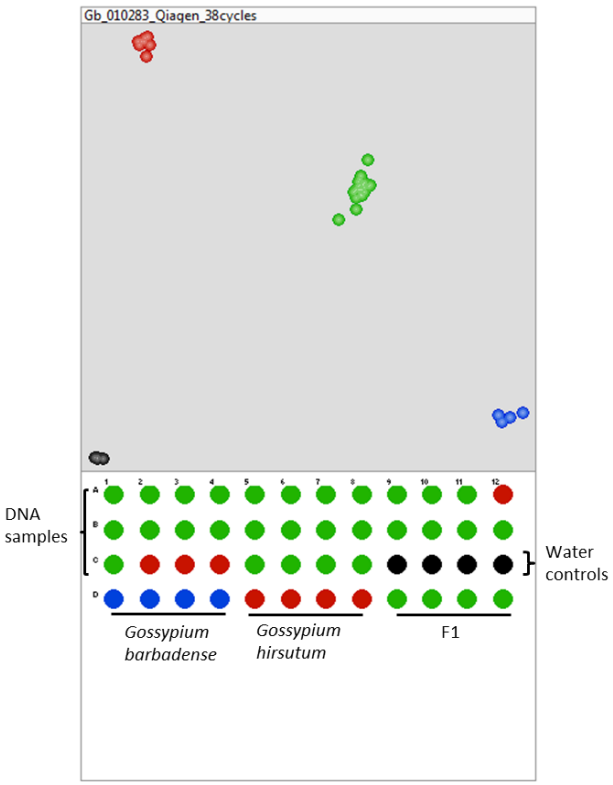

Figure 2.1. KlusterCaller figure showing polymorphism between *G. hirsutum* and *G. barbadense* according to SNP Gb_010283. The red dots represent the *G. hirsutum* (*Gh*) genotype, the blue dots represent the *G. barbadense* (*Gb*) genotype and the green dots represent the F1 between *Gh* and *Gb*. Genotypic results of DNA samples will be defined by the position in which they clustered.

Four SNPs were chosen according to these criteria, i.e., their D5 position, clustering characteristics for genotype calling, ability to discriminate well between *G. hirsutum* and *G. barbadense*, and to segregate well in RILs: Gb_010283, Gb_010482, Gb_013164 and Gb_016965. Data on the SNPs suggested that they would mark loci in the homeologous chromosomes 12 and/or 26 of the AD species.

Se **Localization**

The four *Gh-Gb* SNP markers associated by sequence with the target region on D5 genome Scaffold_8 were used to genotype the semigamous and non-semigamous lines to be used as parents. The results verified that the semigamous lines, all of which are *G. barbadense*, shared the same SNP genotype as line 3-79, and that the non-semigamous red-pigmented *G. hirsutum* lines both shared the same genotypes with TM-1 (data not shown).

Two tests were conducted to verify the relationship between these four markers and the *Semigamy* gene. The first consisted of genotyping the haploid sectors from chimeric seedlings produced by the cross *Sese x SeSe*. We were interested specifically in the SNP genotypes of haploid sectors arising from the maternal *Sese* parent, because semigamous reproduction requires megagametophytic presence of the *Se* allele (Biddle 2006). Given this strong megagametophytic requirement, we anticipated that genotyping of haploid sectors arising from *Sese* female parents using *Se*-linked markers would reveal

only *Gb* SNP alleles, or nearly so, within the limits of recombination. In contrast, a 50:50 *Gb* (*Se*):*Gh* (*se*) allelic distribution was expected for markers not linked with *Se* (*Se*-independent). Data in Table 2.1 show that in all maternally derived haploid sectors, only the *Gb allele* was present when genotyped with *Se*-linked SNPs, whereas both *Gb* and *Gh allele*s were found for the control markers (*Se*-independent). The KlusterCaller genotypic data did not show good amplification and clusters for some of the *Se*-independent control markers, but it was visible that several samples clustered close to the non-semigamous parent genotype (*sese*), i.e., in the opposite side of the semigamous parent genotype (*SeSe*). Moreover, when we examined SNP genotypes of non-chimeric tetraploid seedlings, both *Se*-linked and *Se*-independent SNPs were found to follow a ~50:50 ratio (Table 2.1) indicating the same ratio for *SeSe:Sese* (*GbGb*:*GbGh*) in tetraploid. These vastly different maternal transmission rates of *Se* to maternal haploid sectors versus non-chimeric seedlings indicate that semigamous reproduction is contingent on megagametophytic expression of the *Se* locus and also excludes the argument that the distortion in *Se* versus *se* transmission rate could be due to a *se*-linked lethal gene.

Table 2.1. Genotypes of maternally derived chimeric haploid sectors and normal tetraploid seedlings originated from a *Sese* (*GbGh*) x *SeSe* (*GbGb*) cross according to *Se*-linked and *Se*-independent SNP markers.

	Se-linked SNPs				*Se*-independent SNPs		
Plant_ #	Gb_010283	Gb_010482	Gb_013164	Gb_016965	UCcq10680_162	UCcg11310_419	UCcot10015_139
22	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*	*Gb*
36	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
47	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*
161	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
238	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*
243	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	Failed	*Gh*
255	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*
290	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*
307	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*
308	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	Failed	*Gb*
319	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	Failed	*Gb*
379	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*
386	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
398	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	Failed	*Gb*
*TET 9	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGh*
*TET 14	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGb*	*GbGb*	*GbGh*
*TET 15	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGh*
*TET 18	*GbGh*	*GbGh*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*
*TET 30	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGh*
*TET 35	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGh*	*GbGh*	*GbGb*
*TET 41	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGb*
*TET 70	*GbGb*	*GbGb*	*GbGb*	*GbGb*	*GbGh*	*GbGh*	*GbGb*
*TET 77	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGb*	*GbGh*
*TET 81	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*GbGb*	*GbGb*	*GbGh*

*TET - tetraploid normal seedlings

In the second experiment to verify SNP – *Se* relationships, a *G. hirsutum* x *G. barbadense Se* heterozygous F1 hybrid (*Sese*) was crossed as the pollen

parent onto *G. hirsutum* lines R1 TM-1 and DeRidder Red. Seed were tissue sampled for DNA extractions and germinated to establish the backcross population (BC1F1). DNA samples from 84 seed were screened with the four previously used SNPs, so that the SNP genotypes could be used for marker-assisted selection. Being unsure as to which markers were most closely linked to *Se*, all seedlings containing one or more heterozygous SNPs were selected. From the 84 BC1F1 samples, 39 were genotyped as heterozygous for at least one of the markers and 31 were genotyped as heterozygotes by all of the markers. A total of 42 seeds were sown in the greenhouse, including all the above, plus three seedlings homozygous at all four SNP loci, as controls. From these, 36 seedlings were obtained and later transplanted to the field. Four plants did not survive transplanting to the field. Six of the 10 genotypes that were lost were recombinant for SNP loci, which could be of biological significance. The 32 BC1F1 plants available at flowering were testcrossed as female parent to the *Gb* semigamous line *Sev7*. We were able to produce testcross seed from 27 BC1F1 plants as female parents, including 22 that were heterozygous (*GbGh*) for all four markers, two that were SNP-recombinants and three non-semigamous (*GhGh*) for all four markers as controls. Testcross seedlings resulting from these crosses were screened for chimeras to determine maternal *Se*-locus genotypes (Table 2.2). Out of the 22 BC1F1 plants that were classified as *Sese* on the basis of all four markers, 16 produced a chimera and/or a haploid plant when crossed as female parent to *Sev7.* No chimeras and/or haploids were found

among testcross seedlings from the recombinants or non-semigamous parents (Table 2.2).

Table 2.2. Chimera/haploid production from testcrosses of BC1F1 plants putatively heterozygous for *Se (GbGh)* with a homozygous semigamous line as pollen parent.

Plant_ #	#_seed_	#_germinated	Chimera_#	Haploid_#	Gb_010283	Gb_010482	Gb_013164	Gb_016965	Classification
5	165	111	0	5	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
8	116	31	3	2	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
14	83	48	1	2	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
18	56	30	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
20	51	19	1	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
21	150	96	0	3	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
*28	33	31	0	0	*GbGh*	*GhGh*	*GhGh*	*GhGh*	recombinant
29	185	153	10	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
32	92	71	0	1	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
*43	39	36	0	0	*GhGh*	*GhGh*	*GbGh*	*GbGh*	recombinant
45	27	22	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
46	79	70	0	1	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
50	37	32	1	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
53	23	23	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
55	142	133	0	1	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
58	94	42	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
61	109	77	0	2	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
62	64	35	0	1	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
63	156	121	0	1	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
65	136	109	7	2	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
67	117	103	2	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
68	47	28	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
70	45	33	0	0	*GhGh*	*GhGh*	*GhGh*	*GhGh*	non-semigamous control
71	35	33	0	0	*GhGh*	*GhGh*	*GhGh*	*GhGh*	non-semigamous control
76	13	7	0	0	*GhGh*	*GhGh*	*GhGh*	*GhGh*	non-semigamous control
36	34	26	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*
81	57	22	0	1	*GbGh*	*GbGh*	*GbGh*	*GbGh*	*Sese*

*Recombinants

Biddle (2006) established that the *Semigamy* gene must be present in the female parent to be expressed. Accordingly, production of chimeras and/or haploids will only occur when the female parent has at least one copy of the *Semigamy* gene. *Sese* heterozygous BC1F1 females would thus be expected to produce chimeras, but not non-semigamous BC1F1 homozygotes (*sese*). It is also known that in terms of seed production, levels of *Se* expression are less than 100%, i.e., normal tetraploid progeny are also produced by *SeSe* homozygous mutants. Thus, the results of experiment-2 also establish a correlation between the SNP genotypic data and production of chimeras and haploids. If a monogenic factor were responsible for semigamous reproduction and not linked to the SNPs, a 1:1 ratio would be expected among SNP homozygous genotypes. However, a Chi-square test on the SNP-selected (putative *Sese*) BC1F1 plants demonstrated a non-1:1 ratio of plants that produced chimeras to those that did not produce chimeras and/or haploid ($p < 0.05$).

There were two single-crossover events between markers, one between Gb_010283 and Gb_010482, and the other between Gb_010482 and Gb_013164 (Table 2.2). Given the understanding that only *Sese* BC1F1 females could produce chimeras, chimeric/haploid progeny arising from the BC1F1 seed parents would only have come from Sese BC1F1 plants, so SNP alleles that are preferentially associated with the chimera-producing BC1F1 plants would be closer to the gene than the ones that genotyped the same plants as non-

semigamous. Unfortunately, neither of the plants where the crossovers occurred produced chimeras and/or haploids and there was no satisfactory conclusion on this matter. One of the explanations for the lack of chimeras and/or haploids among the progeny of those plants is the fact that there were not as many seeds produced from the testcross using those specific plants, one produced 33 seeds and the other produced 39 seeds. Upon planting, only 31 and 36 seeds germinated respectively so, the number of seeds used for screening chimeras/haploids was very low. It was observed (in another work from this same research project) that the percentage of chimeric progeny derived from a *Sese* x *SeSe* cross is less than 2%, therefore there was a high possibility of not detecting any chimeras/haploids among the seedlings produced by those plants. Another explanation is a possible crossover event between the gene and the closest marker. In this last case, despite that the marker pointed those plants as heterozygotes, they did not possess the gene in their genome.

The same four markers were used to genotype the advanced BC3F1 population: Gb_010283, Gb_010482, Gb_013164 and Gb_016965. Upon marker-assisted selection, 37 out of 75 samples were genotyped as clear heterozygotes (*Sese*) for at least one of the markers. Plants that were recombinant for these markers were used as females in a testcross with *Sev7* mutants following the same rationale for the test with BC1F1 plants. Upon seedling screening, recombinants that produced chimera(s) or haploid(s) would indicate that the SNP marker that genotyped that plant as *Sese* is closer to the

Semigamy gene than the SNPs that genotyped that same plant as non-semigamous (*sese*). A total of 1431 seedlings were evaluated and screened for chimeras or haploids. According to table 2.3 nine seedlings were described as “chimera-like”, eight of them were from plants with a recombination between Gb_010283 and Gb_010482 and between Gb_010482 and Gb_016965. One “chimera-like” occurred in one of the non-semigamous controls.

Table 2.3. Rate of chimera/haploid production originated from the testcross between recombinant BC3F1 plants genotyped as *Sese (GbGh)* for at least one of the SNP markers and a semigamous line.

Plant_#	#_seed	#_germinated	"CHI-like"	Haploids	Gb_010283	Gb_010482	Gb_013164	Gb_016965	Classification
A2	224	214	2	0	*GbGh*	*GbGh*	*GbGh*	*GhGh*	recombinant
A3	172	167	0	0	*GhGh*	*GhGh*	*GhGh*	*GbGh*	recombinant
A4	117	111	0	0	*GhGh*	*GbGh*	*GbGh*	*GbGh*	recombinant
A9	198	189	1	0	*GbGh*	*GbGh*	*GbGh*	*GhGh*	recombinant
C4	120	98	1	0	*GhGh*	*GhGh*	*GhGh*	*GhGh*	non-semigamous control (*GhGh*)
*C7	53	35	0	0	*GbGh*	*GhGh*	*GhGh*	*GhGh*	recombinant
*C9	45	37	1	0	*GbGh*	*GhGh*	*GhGh*	*GhGh*	recombinant
*C10	147	104	3	0	*GbGh*	*GhGh*	Failed	*GhGh*	recombinant
*D2	187	120	1	0	*GbGh*	*GhGh*	*GhGh*	*GhGh*	recombinant
*D5	27	20	0	0	*GbGh*	*GhGh*	*GhGh*	*GhGh*	recombinant
E1	167	146	0	0	*GbGh*	*GbGh*	*GbGh*	*GbGh*	semigamous control (*GbGh*)
E2	109	103	0	0	*GhGh*	*GhGh*	*GhGh*	*GhGh*	*non-semigamous control* (*GbGh*)
F11	91	38	0	0	*GhGh*	*GbGh*	*GbGh*	*GbGh*	recombinant
F12	25	23	0	0	*GhGh*	*GhGh*	*GbGh*	*GbGh*	recombinant

*Recombination had occurred in previous generation

The “chimera-like” description was used to denote a tentative classification for seedlings that seemed to present subtle chimerism on the cotyledons, but no chimerism on the hypocotyl. They lacked clear sectorial chimerism. According to Dolan and Poethig (1998), chimeras could involve any of the three tissue layers of the leaves so, a subtle change of color in the cotyledonary surface could indicate haploidy in one of the histogenic layers of the cotyledon. The seedlings were transplanted to a bigger pot and left to grow in the greenhouse to confirm chimerism, but none could be confirmed as chimeric or haploid. Possible explanations might be a mixture of tetraploid and haploid tissue and the haploid tissue did not develop after seedling stage, or that no chimeras were produced from those recombinants. The latter can imply that a recombination event occurred between the closest marker and the gene and the recombinant plants do not have the *Semigamy* gene in their genome. Further tests are needed on this population to clarify those questions.

In addition, two of the SNP markers were used in a secondary work, described in the third chapter of this thesis. Suffice it, here, to say that the genotypic results given by those SNP markers in Chapter 3 were consistent with the hypothesis that *Semigamy* expression differs in male and female parents. Those results suggest that semigamous reproduction is contingent on megagametophytic expression of the *Se* locus and that the SNP markers are associated with the *Semigamy* gene.

Incongruities

The SNPs used in this work were selected in 2012, when the first crosses involving semigamous plants were performed to generate the populations used in the present research. I used the program BLAST to map the *Se*-linked SSRs according to sequence similarity to Scaffold_8 of the D5 reference genome from *Gossypium raimondii*. In previous work, our lab has shown that the D5 Scaffold_8 corresponds to homeologous chromosomes 12 and 26 of *G. hirsutum*. The physical locations of those SSR markers were used to identify SNPs between *G. hirsutum* and *G. barbadense* in that same genomic region, using the BWA alignment software (Li and Durbin, 2009) for SNP sequences identified by our laboratory in collaborative research (Hulse-Kemp, Ashrafi et al., 2014). The genotypic results of chimeric haploid sectors and tetraploid seedlings for MAS given by those referred SNPs were satisfactory so far, in terms of phenotypic and genotypic expectations.

In 2014, a robust SNP-chip involving more than 60,000 non-redundant gene-associated SNPs for several cotton species was produced via Illumina by our lab and associates (Hulse-Kemp, Ashrafi et al., 2014, Hulse-Kemp et al., unpublished). Two of the four SNPs used throughout this work were added to the chip, Gb_010283 and Gb_016965, and renamed according to the hierarchical SNP-naming procedure as Gb379_009473 and Gb379_014995 respectively. In the interspecific *G. hirsutum* x *G. barbadense* F2 map created with the Chip, those SNPs were mapped to chromosome 26, which is the D-

subgenome homeolog for chromosome 12. The chromosome-26 location seemingly contradicts the chromosome-12 *Semigamy* mapping position found previously. In an attempt to understand the incongruity between these two mapped positions, we performed a linkage analysis by genotyping the same BC1F1 and BC3F1 populations used in this present work with SNP Sc08_43919741, which was recently identified and localized to the D5 Scaffold-8 (Zhu et al. 2014). We determined that this SNP is in the D5 Scaffold-8 region that contains all of the previously identified *Se*-linked SSRs and SNPs. Analysis of *G. hirsutum* x *G. barbadense* F2 linkage maps based on the Chip shows that this SNP is in chromosome 12 (Hulse-Kemp et al. unpublished).

The KlusterCaller results of the BC1 and BC3 populations genotyped by the SNP Sc08_43919741 are shown in Figure 2.1. The clusters' positions relative to the controls (parents and F1 marked with an asterisk) indicate a multi-locus SNP. The green dots represent the F1 control and all samples that were heterozygous for all loci on that SNP. The yellow dots represent samples that are heterozygous for all but one locus and the orange dots indicate samples that are heterozygous for all but two loci for that specific SNP. Since we cannot determine if the samples clustering on the yellow or orange positions are heterozygous for the same loci as the control, we did not consider those samples for the linkage analysis. Moreover, the SNP Sc08_43919741 was effective for populations involving the R1 TM1 parent (red dots) but not for the DeRidder Red population. When genotyping the BC1 and BC3 samples that

originated from crosses with the recurrent parent DeRidder Red, the SNP was found to be non-polymorphic between DeRidder Red and *G. barbadense* (blue dots), because the former seems to have the same Sc08_43919741 SNP allele as *G. barbadense* (Figure 2.2). Thus, we were also unable to use the samples originating from DeRidder Red as part of the analysis of linkage relationships between markers.

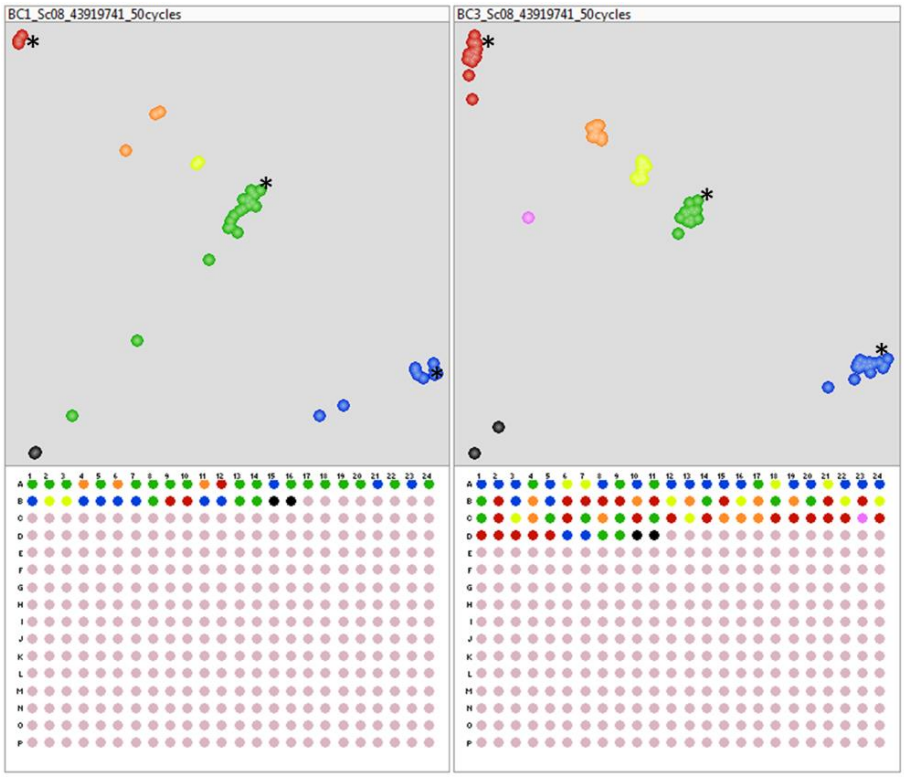

Figure 2.2. KlusterCaller figure representing genotypic results of BC1F1 and BC3F1 populations according to SNP Sc08_43919741. Asterisks represent parents and F1 position. Only samples that clustered on the same position as the parents or F1 were considered for linkage analyses.

Linkage analysis of the BC1 and BC3 population samples that clustered in the same position as the parental controls indicate that Sc08_43919741 is strongly linked with the SNPs previously identified as *Se*-linked based on segregation analysis (Table 2.4). The range in map distances between the markers at BC1F1 population with 10 samples was between 0 cM to 20 cM and at BC3F1 population with 32 samples was between 0 cM and 6.25cM.

Table 2.4. Genetic distances between SNP Sc08_43919741 and previously used SNPs.

BC1F1 population (10 samples)			
Genetic distances between:			
Gb_010283 and Sc08_43919741	Gb_010482 and Sc08_43919741	Gb_013164 and Sc08_43919741	Gb_016965 and Sc08_43919741
0 cM	10 cM	20 cM	20 cM
BC3F1 population (32 samples)			
Genetic distances between:			
Gb_010283 and Sc08_43919741	Gb_010482 and Sc08_43919741	Gb_013164 and Sc08_43919741	Gb_016965 and Sc08_43919741
0 cM	3.125 cM	6.25 cM	6.25 cM

The genetic distances presented in Table 2.4 display a level of linkage among the SNPs that would not be expected if they were in different chromosomes.

Three hypotheses were made to explain the incongruity of the SNP linkage analysis results in the BC1 and BC3 populations versus the interspecific F2 population mapping using the SNP-chip. The former indicated that all of the five SNPs are linked, whereas the latter indicated that one SNP (Sc08_43919741) is in chromosome 12 and two (Gb_010283 and Gb_016965) are in chromosome 26 (the other two (Gb_010482 and Gb_013164) were not used in the SNP-chip). Hypothesis 1 is that the genomic segment where the SNPs are localized is duplicated in homeologous chromosomes 12 and 26. Therefore, the SNPs and possibly *Semigamy* are in both chromosomes 12 and 26. Although this is a possible explanation, mapping results of the interspecific F2 population using the SNP-chip did not show Gb_010283 and Gb_016965 as homeo-SNPs i.e., they appear to be in only one of the homeologs. Another possibility is that a duplication at the *Se* region occurred only at the semigamous line and ectopic homeologous recombination occurred between chromosome 12 and 26. In this case, only the semigamous line would have the duplicated region in both chromosomes and SNP genotyping results would differ according to the population screened i.e., semigamous versus non-semigamous populations. Hypothesis 2 is that the backcross populations were not large enough to properly identify genetic distances among SNPs. In this scenario, previous SNPs are in chromosome 26 and the SNP newly tested is in chromosome 12. The zero recombination rate between Gb_010283 and Sc08_43919741 makes this hypothesis probabilistically challenging since the random chance of having no

recombination in 42 samples between two SNPs (as shown on table 2.4) in different chromosomes would normally be infinitesimally low. However, if expression of the semigamy trait depends on two genes, one in 12 and one in 26, that would explain the correlation between genotypic and phenotypic data in my BC1F1 population (both genes are still present in this early generation), and the lack of confirmed chimeras or haploids on my BC3F1 population (selection occurred just for 26, not for 12). The last hypothesis is that previous SNPs are in chromosomes 12 and they have been mistakenly mapped to chromosome 26.

Despite all the potential that this gene has for the breeding and genetic community, there is still much to be learned for it to be fully understood. I hope that this work contributes to future studies.

Conclusions

The results from this present work, despite being less than fully conclusive, show a clear correlation between the *Semigamy* gene and several SNP markers that are sequence-localized to the same D5 genomic region as SSRs that were previously believed to be linked to *Se*. Thus, the data reinforce previous segregation based and sector-based mapping results from our lab (Stelly lab, unpublished). There is still some inconsistency in where the SNPs used are located on the *Gossypium barbadense* genome, whether it is in chromosome 12, in its homeolog 26 or in both chromosomes. The production of chimeras/haploids from BC1F1 plants genotyped as heterozygotes by the

markers establishes a connection between phenotypic and genotypic data, since only heterozygous (*GbGh*) females would be able to express the gene as opposed to non-semigamous plants (*GhGh*). Plants genotyped as *GhGh* by all four markers did not produce chimeras or haploids when testcrossed as females with a semigamous line. When the same test was performed on an advanced BC3F1 population, we did not observe well-defined chimeras among seedlings originated from plants genotyped as heterozygotes by the same markers; some "chimera-like" seedlings were tentatively identified, but when they were grown in the greenhouse, chimerism could not be confirmed. This could be explained by: (1) the relatively low number of seedlings screened; (2) the fact that "chimera-like" seedlings were chimeric for a mixture of haploid and tetraploid tissue, but that haploid tissue did not develop after seedling stage; (3) a recombination event during population advancement between the markers and the gene or, (4) the semigamy trait would be expressed by two genes, one in chromosome 12 and one in 26. If the SNPs used for MAS during backcrossing used are, in fact, in chromosome 26, as suggested by independent Chip-based mapping of a different interspecific population, then the MAS for *Se* during population advancement was made only for the respective segment of c26, and the homeologous segment of c12 was lost.

A high-resolution map of the *Semigamy* region is a goal of our lab. Once the map is created, a segregating population can be genotyped, recombinants between SNPs can be phenotyped by cytology or progeny scoring to establish a

genotype/phenotype relationship and the gene can be precisely localized. Moreover, if we have SNPs on both sides of the *Se* gene, BC1F1 and BC3F1 populations used in this work can be genotyped by a set of markers distal and proximal to the gene to characterize any recombination event that could have happened during population advancement. If there was no recombination and the gene(s) is (are) still in BC3F1 population, the factor(s) causing a lack of chimeras and haploids among progeny could be of interest e.g., allelic interactions of two or more genes could be determining the semigamy trait.

In addition, the four SNP markers used in this present work, Gb_010283, Gb_010482, Gb_013164 and Gb_016965 will be used to genotype the interspecific *G. hirsutum* x *G. barbadense* F2 population used in the SNP chip. Genotypic results given by those markers will be analyzed and compared to the SNP chip results used to create the map and perhaps, this will bring clarification to the contradiction between linkage analysis and mapping position of those SNPs.

In conclusion, the precise localization of *Semigamy* to a well-characterized genetic marker region in the cotton genome is much desired and it has been one research focus of our lab. A deep understanding of the *Se* gene and its localization would enable reliable marker-assisted selection within a segregating population as well as to bring us one step closer to cloning the gene, and use it more effectively, in cotton as well as other crops and/or model research organisms.

CHAPTER III

SNP-BASED INFERENCES ON EXPRESSION, REPRODUCTIVE RAMIFICATIONS AND INHERITANCE OF THE *SEMIGAMY* GENE IN COTTON

Introduction

Reproductive mutants have a tremendous impact on sexual and asexual reproduction research because they enable biological and genetic investigations of the reproductive system.

Semigamy is an unusual reproductive feature that causes formation of haploid and chimeric progeny by the non-fusion of female and male gametic nuclei (Battaglia 1946). In cotton, semigamy naturally occurs in *Gossypium barbadense* species (Turcotte and Feaster 1967). Semigamy's striking apomixis-like features and effects have drawn attention for decades, but the difficulty of accurately classifying genotypes on the basis of phenotypes and the variability of haploid production rates has complicated and delayed research progress on this gene. Until recently, all the knowledge about the *Semigamy* mutant's time and mode of action in cotton were based only on progeny scoring, i.e., the information about a plant carrying the *Semigamy* gene was determined based on reproductive phenotype, specifically on the frequency of haploids and/or chimeras among its progeny. Such determinations are indirect, and have some limitations. Semigametic plants can also produce normal tetraploid progeny due

to the orientation of spindle fibers during mitotic divisions of the zygote (Hodnett 2006) (Figure 3.1). The frequencies of haploids and chimeras can also be influenced by background segregation and/or the environment, which can differentially affect tetraploid and haploid (diploid) progeny germination by imposing different levels of inbreeding depression or environmental effects on germination rates. For example, extensive inbreeding depression would expectedly depress survival of haploids derived from interspecific F1 hybrids.

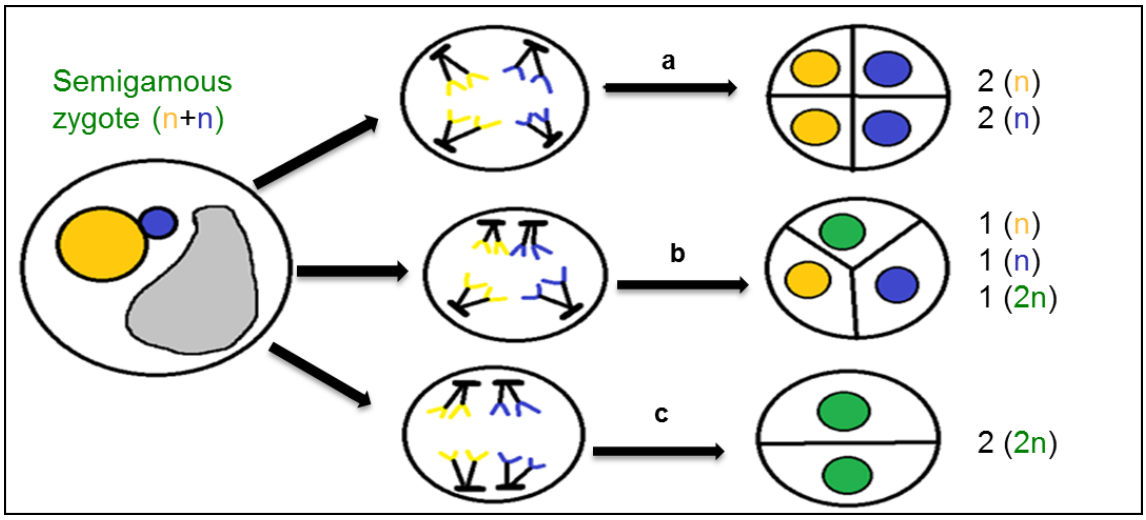

Figure 3.1. First mitotic divisions of semigamous zygote. Spindle number and orientation determine possible outcomes: (a) tetrapolar → 4 haploid cells; (b) tripolar → 1 normal + 2 haploid cells, and (c) bipolar (single spindle or fused/parallel spindles), → two normal cells (modified from Biddle 2006).

To circumvent the above problems, our laboratory developed a novel strategy that emphasized direct phenotypic evaluation using cytological

reproductive analysis. We used a new clearing method developed in our lab, which allowed the direct examination of the egg cell, before, during and after fertilization and provided reliable differentiation between unfertilized, sexual and semigamous ovules (Hodnett, Crane et al. 1997, Biddle 2006, Hodnett 2006). Cytological analysis of ovules resulting from several intercrosses among homozygous semigamous (*SeSe*), heterozygotes and homozygous non-semigamous (*sese*) plants revealed that the sperm cell penetrates the egg cell, and that the nuclei do not fuse (Biddle 2006, Hodnett 2006) (Figure 3.2). In 1967, Turcotte and Feaster suggested that the source of haploids in Pima cotton was due to semigamous reproduction, based on fact that some progeny were chimeric, but all were monoembryonic (Turcotte and Feaster 1967). However, cytological proof was never developed until the work of Hodnett (2006) and Biddle (2006).

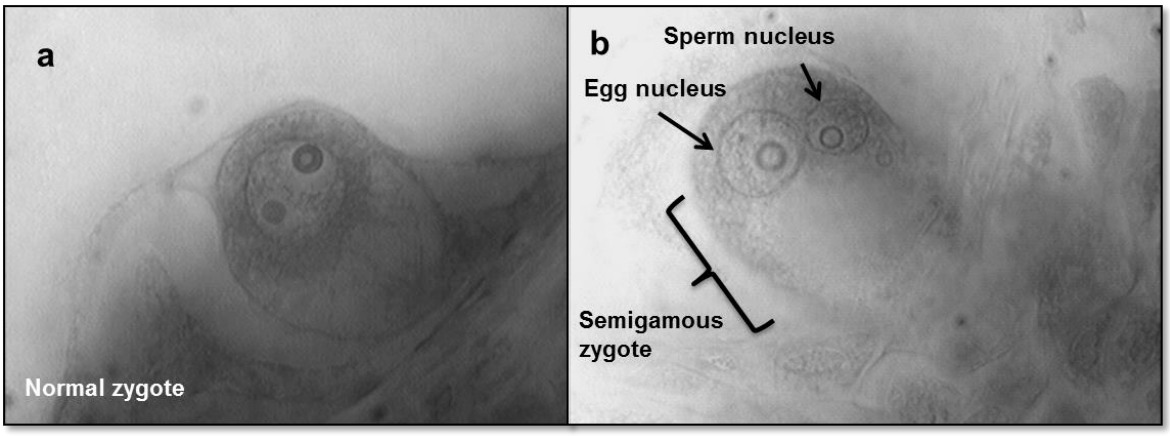

Figure 3.2. Normal versus semigamous zygote as depicted in Biddle 2006. (a) Sexual zygote. The presence of two nucleoli indicates that karyogamy has recently occurred. (b) Semigamous zygote. Though the egg and sperm nuclei are adjacent to each other, no fusion has occurred (Biddle 2006).

The overall results from Biddle's (2006) research showed that the actual frequencies of semigamous ovule production differ markedly from the rates that had been previously reported, because those had been deduced from progeny scoring rather than by direct cytological observation. From her results, two possible hypotheses about *Semigamy* expression were raised: either *Se* is a recessive gene and acts at the zygotic stage or *Se* is expressed at the gametophytic stages, in which case the type of gene action cannot be inferred from available data. However, neither of those hypotheses could be eliminated by her statistical analysis. Those conclusions have one contradiction also mentioned by Biddle, which is that the occurrence of chimeras from crosses between homozygous semigamous lines and non-semigamous lines indicate

that the expression of the *Semigamy* allele could not be exclusively zygotic recessive.

Although *Se* gene is transferred through both parents, it seems that *Semigamy* is only expressed when present in the female parent. In crosses between non-semigamous females (*sese*) and a heterozygous (*Sese*) or homozygous (*SeSe*) males, all or nearly all zygotes develop normally, i.e., non-semigamously (Biddle 2006). The contrasting results from reciprocal crosses led our lab to suggest possible differences in *Se* expression in mega- versus micro-gametophytes, and/or the resulting zygotes.

A deeper understanding of *Semigamy* reproductive ramifications and inheritance is likely to be achieved through a more detailed characterization of the gene's time(s) of expression. Moreover, a better characterization of *Semigamy* expression may be essential to its large-scale use in research and breeding, given the complicated interplay between this gene's inheritance, expression, reproductive ramifications and their heritability. Therefore, this present study has the objective of determine and compare rates of *Semigamy* expression according to allelic inheritance from maternal and paternal heterozygous (*Sese*) parents.

Material and Methods

The populations used throughout this work came from crosses between two species of cotton, *Gossypium hirsutum* and *Gossypium barbadense*. Three

different lines of *Gossypium barbadense* that were homozygotes for *Semigamy* were used in the crosses: 57-4, *Sev7* and *R1* Pima. *Sev7* and *R1* Pima have phenotypic markers i.e., light green and red leaf color respectively. The *Gossypium hirsutum* parents were TM-1, *R1* TM-1 and DeRidder Red. The last two lines have a phenotypic marker that confers a red color to the leaves and bracts.

Interspecific hybrid F1 and BC1F1 plants were bred to create plants that were heterozygous for *Semigamy* (*Sese*), highly heterozygous for DNA sequence-based markers, and also carrying certain cell-autonomous phenotypic (color) markers. These F1 and BC1 plants were used as parents in genetically marked crosses described below to generate chimeric progeny seedlings with maternal and paternal haploid sectors. The phenotypic markers were used to facilitate detection of seedling chimerism, haploidy, zygotic composition, and, for haploid sectors, the parent-of-origin. The chimeric seedlings were produced by reciprocally crossing genetically marked *Sese* plants and semigamous lines (*SeSe*).

Two batches of crosses performed, as shown below in Tables 3.1 and 3.2. The *Sese* heterozygous plants were used reciprocally as the female parent (Table 3.1) or male parent (Table 3.2). Parental combinations enhanced parental differences in allelic content for phenotypic markers, to facilitate inferences about parent-of-origin for specific sectors of chimeric and non-chimeric seedlings.

Table 3.1. First batch of crosses, denoted by x, to explore maternal expression, featuring *Semigamy* heterozygotes as the female parents, crossed with a *Se*-homozygous pollen parent (tester).

	Male ♂	
Female ♀	*SeSe R1R1 V7V7* (red)	*SeSe r1r1 v7v7* (yellow-green)
Sese r1r1 V7v7 (segregating *v7*)	X	
Sese R1r1 V7V7 (segregating *R1*)		X
Sese R1r1 V7v7 (segregating *R1 and v7*)		X

Sese r1r1 V7v7 X *SeSe R1R1 V7V7*: The female used in the cross above was a *Sese r1r1 V7v7* BC1F1 plant, i.e., heterozygous for *Semigamy* and *v7*, but homozygous for *r1*. In this cross, the female parent had normal green leaves. The male parent was a homozygous *Se* line with deep-red leaves.

Sese R1r1 V7V7 X *SeSe r1r1 v7v7*: The female used in the cross above was a *Sese R1r1 V7V7* BC1F1 plant with medium-red leaves and the male was a homozygous *Se* line that had light-green leaves.

Sese R1r1 V7v7 X *SeSe r1r1 v7v7*: The female to be used in the cross above, *Sese R1r1 V7v7*, was of the F1 generation, with medium-red leaves, whereas the male was a homozygous *Se* line that has light-green leaves.

Table 3.2. Second batch of crosses, denoted by x, to explore paternal expression, featured *Semigamy* heterozygotes as the male parent, and the *Semigamy* homozygotes used as female tester parents.

	Male ♂	
Female ♀	*Sese R1r1 V7V7* (segregating *R1*)	*Sese R1r1 V7v7* (segregating *R1* and *v7*)
SeSe r1r1 v7v7 (yellow-green)	X	X

Note that the two crosses shown in Table 3.2 are the reciprocals of the second and third crosses shown in Table 3.1: *SeSe r1r1 v7v7* X *Sese R1r1 V7V7* and *SeSe r1r1 v7v7* X *Sese R1r1 V7v7.*

The crosses were performed in the field and progeny seeds were harvested and processed for planting i.e., ginned, dried down (3 days at 38º C) and cold-treated (5 days at 5.5º C to 8.8º C, 12 to 15% humidity) to break dormancy. Seeds were planted in 10 x 20-inch trays in the greenhouse and seedlings were screened for chimerism at approximately 1- 2 weeks after germination.

Cotyledon tissue for DNA extraction was collected from chimeras with well-defined pigmentation borders. Sampling was targeted to specific haploids sectors, i.e., yellow-green, green and/or red, which originated from *Sese* parents. For collection, a paper punch was used to obtain uniformly sized tissue samples (Figure 3.3). Phenotypic information about the sector and seedling was maintained as part of each sample's identity.

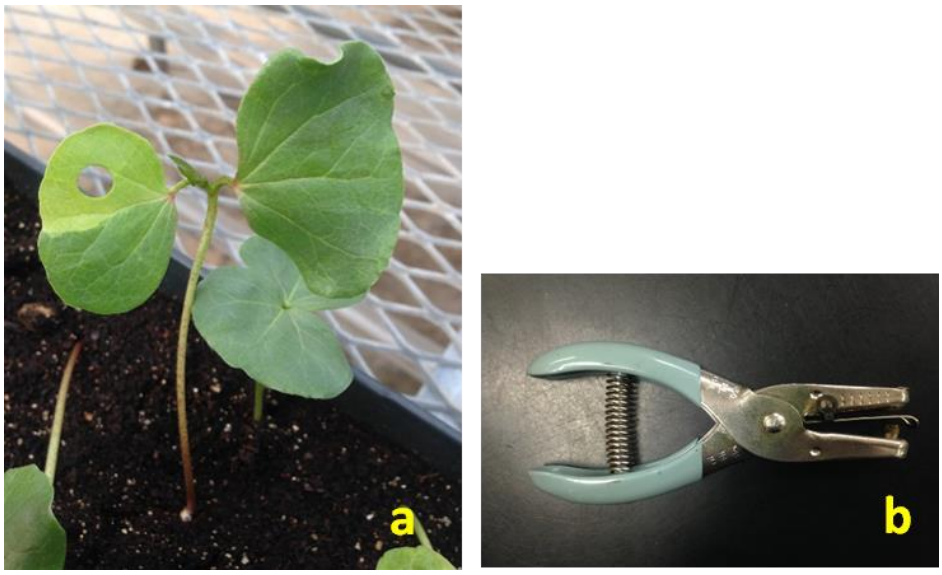

Figure 3.3. (a) Seedling with part of its cotyledonary tissue taken for DNA extraction; (b) Standard hole punch (6mm).

DNA from the sectors was extracted using a high-throughput DNA extraction method developed in our lab (Zheng et al., in press). Later, 2 µL of diluted DNA were loaded into wells of 96-well plates and dried in the oven at 60ºC for 1 hour. After optional storage of plates containing dried DNA samples, 8µL of PCR mixture (containing 4.0µL of Reaction Mix including polymerase, 3.8µL of sterile deionized water, 0.11µL of Primer Mix, and 0.064µL of 50mM MgCl2) was added to each well. After brief centrifugation, the plates were submitted to a thermocycler machine (Eppendorf Mastercycler Ep Gradient S Thermal) set according to KASP assay recommendations, i.e., the PCR cycling program for the KASP assays included a preliminary acclimation step of 94°C for 15 minutes before cycling, followed by 10 cycles at 94°C for 20 seconds, then

65°C for 1 minute decreasing 0.8°C per cycle to an annealing temperature of 57°C for the final cycle. This was followed by 28 cycles of denaturation at 94°C for 20 seconds, and annealing at 57°C for 1 minute. More cycles were added when greater amplification was required (Zheng et al. in press). The plates were briefly centrifuged and then submitted to a Pherastar plate reader (BMG LABTECH). The reads were imaged and interpreted using KlusterCaller software (LGC Genomics) with a set of 7 *G. hirsutum - G. barbadense* SNPs i.e., (1) Gb_010283, (2) Gb_016965, (3) UCcg10220_69, (4) UCcg10680_162, (5) UCcg11310_419, (6) UCcot10015_139, (7) UCcot10322_62.

The genotypic results of the chimeras originated from the F1 and BC1F1 populations testcrossed with the *SeSe* line, for each test (maternal or paternal expression), were submitted to a Chi-square test of homogeneity to check if they can be combined.

Results and Discussion

Inheritance of the *Se* allele was studied in a manner different from the two previous methods: (1) phenotypic classification of segregating progeny according to observation of moderate numbers of seedlings, sufficient enough in number to discern the production of some haploid and chimeric progeny (semigamous), or none at all (normal, non-semigamous), *sensu* Turcotte and Feaster (1967) and (2) the cytological scoring of small populations of syngamous zygotes for their failure (semigamous) versus ability (non-

semigamous) to undergo karyogamy, *sensu* Hodnett (2006) and Biddle (2006). In the seedling-based approach, visually scorable seedling traits that are determined in a cell-autonomous manner can be used to great advantage in detecting the presence of haploid or chimeric seedlings. But in parental combinations where the frequencies of chimeric/haploid seedlings are low for some segregates, it is often challenging to have enough progeny seed and seedling data to know for sure that the seedling haploid/chimerism rate is zero, especially if the female is not fecund and/or inbreeding depression might suppress recovery of haploids and chimeras with haploid sectors. Whereas the seedling-based approach is based on indirect inference from seedling populations that are temporally and developmentally removed from fertilization, the time of *Semigamy* gene action, the cytological method involves direct observation of the recently fertilized ovule for telltale signs of karyogamy (non-semigamous), or lack thereof (semigamous). It is extremely accurate - probably more accurate than seedling-based approaches, but is tedious and requires cytological skills, a microscope, and imaging system of sufficient quality. So that we might obtain complementary information on the inheritance, linkage relationships, and expression of *Semigamy*, we devised a method based on genotyping at the seedling stage, i.e., without the need for a segregating family of flowering plants, without the need for additional seed production, and without the need for extensive cytological analysis. We identified chimeras containing a haploid sector that originated from a heterozygous parent. In the various

crosses, the heterozygotes were used as male or female, so it was possible to separately study the paternal and maternal inheritance. The *Sese* plants used as parents were from interspecific (*G. barbadense* - *G. hirsutum*) F1 and BC1F1 populations from backcrosses with *G. hirsutum* as the recurrent parent.

We facilitated the detection of chimeras and parent-specific haploid sectors by using parental plants that differ in leaf color (Figure 3.4). Two of the *G. hirsutum* parents had the *R1* gene in TM1 or DeRidder Red backgrounds. One of the *G. barbadense* parents had the *R1* gene in Pima background. The *R1* allele is co-dominant and confers a red color to the leaves and bracts of the cotton plant by anthocyanin pigmentation. Homozygous plants (*R1R1*) tend to present a deeper red color than heterozygous plants (*R1r1*). The gene is in chromosome 16 (Fryxell 1984). The other *G. barbadense* parents had the *v7* gene (virescence mutant). The *v7* allele is recessive and it confers a yellow-green color to the leaves and bracts of the cotton plant, especially in low-light situations. The gene is located in chromosome 21 (Fryxell 1984).

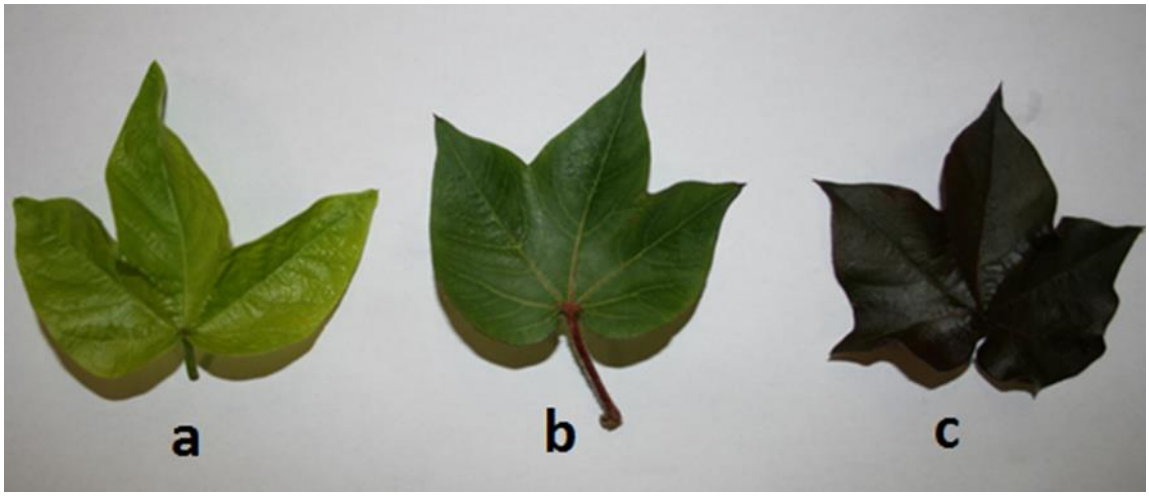

Figure 3.4. Leaves from 3 different parental plants. (a) Virescent-color leaf caused by the recessive allele v_7; (b) Normal green leaf; (c) Red-color leaf caused by the dominant allele R_1.

The use of haploid sectors for analysis of inheritance and expression maximized genetic simplicity of the samples. The simplicity was leveraged by creating parental *Sese* heterozygotes that were also heterozygous for nearby molecular SNP markers; the abundance of nearby SNPs was assured by creating interspecific hybrids. When the highly heterozygous *Sese* parents were used in reciprocal crosses with *SeSe* homozygous lines to produce chimeras, each progeny sector's genotype at the most closely *Se*-linked SNP loci allowed the *Se*-allele genotype to be inferred, and recombinants to be detected. Given knowledge of the direction of the cross and the *Se*-transmission results, it was feasible to deduce the expression of the *Se* gene from maternal versus paternal heterozygotes.

While this new strategy based on the sectors and closely linked DNA markers offers several important advantages, it also suffers from some efficacy-reducing issues. One is that it is relatively difficult to recover chimeras and haploids from interspecific crosses such as *Gossypium hirsutum* x *Gossypium barbadense*, due to inbreeding depression at embryonic and seedling stages. The overall percentage of chimeric seedlings from crosses between *SeSe* females x *Sese* males was 5.78%. When the reciprocal cross was performed, i.e., using *Sese* as females, the percentage is even lower at 1.73%. This difference in frequencies of chimeras from male and female heterozygotic parents might be attributed to possible maternal effects, mitochondrial effect on *Semigamy* gene expression in which the expression of the gene in this case, production of haploids/chimeras, is dependent on the number of *Gb Semigamy* alleles present in the female parent (Biddle 2006). A possible explanation for the relatively low number of chimeras recovered in both crosses is inbreeding depression. When a wide cross is made with the purpose of producing haploids and chimeras, those plants (or sectors) find themselves in an “allelic mess” where alleles from both species are present in a hemizygous state; intergenic incongruities would be expected to cause elevated rates of inviability and poor germination.

Another potential problem is cryptic chimerism. Chimerism of a plant can be represented in plant tissues in a number of ways that depend on the number of mutant cells, their distribution in the meristem(s) and the patterns of cell

division in the meristem(s). Sectoral, periclinal or mericlinal chimeras are commonly occurring types (Tilney-Bassett 1986). Various types of chimerism have been described for cotton leaves and the semigamously formed chimeras, including maternal and/or paternal leaf sectors or layers being different from the other(s) in genotype and/or ploidy (Dolan and Poethig 1998). To contend with that possible complication, which could alter genotypes, all the sectors used for genotyping were also submitted to a ploidy test to confirm haploidy, which is detailed below.

Seven SNPs were used to genotype the haploid sectors: two *Se*-linked (Gb_010283 and Gb_016965) and five *Se*-independent (no linkage) *G. hirsutum* - *G. barbadense* SNPs (UCcg10220_69, UCcg10680_162, UCcg11310_419, UCcot10015_139 and UCcot10322_62), all localized to chromosomes other than chromosome 12 or its homeolog 26, where we believe the *Semigamy* gene is located. The five unlinked markers served two roles, one being that they served as "controls" for linkage analysis, and the second being that they enable a simple test for ploidy: all five control markers are unlinked and mark loci in five different chromosomes. The heterozygous parent could have *G. barbadense* or *G. hirsutum* alleles at its SNP loci. The tester parent is homozygous for the *G. barbadense* allele at all seven SNP loci. The only progeny arising from a cross between the *Sese* parent and the homozygous tester that appeared to be homozygous for the *G. hirsutum* allele at any of the SNPs was a haploid from the *Sese* parent. Each SNP locus of normal zygotes would have received at

least one *G. barbadense* allele from the *SeSe* tester, and would thus be either heterozygous or homozygous for *G. barbadense* allele. Any sample heterozygous for any of the SNP markers would not be a true haploid sector, and therefore was not used in the analysis. Samples used for subsequent analysis were haploids exhibiting homozygous-like TM-1 genotypes at one or more of the SNP loci (preferable) and samples that did not show a heterozygous genotype in any of the SNP loci.

When analyzing the genotypic results from haploid sectors, two types of expression were considered: sporophytic and gametophytic. The plant life cycle is basically represented by an alternation of generations where there is a haploid gametophytic phase and a diploid sporophytic phase (Raven, Evert et al. 2005). When the expression of a gene occurs in the gametophytes, only gametes that have the allele corresponding to that trait will be expressing it. When expression occurs in the sporophyte stage, all gametes formed will equally express the gene independently of which allele they have i.e., the trait will be determined by gene expression before the gametes are formed (Figure 3.5).

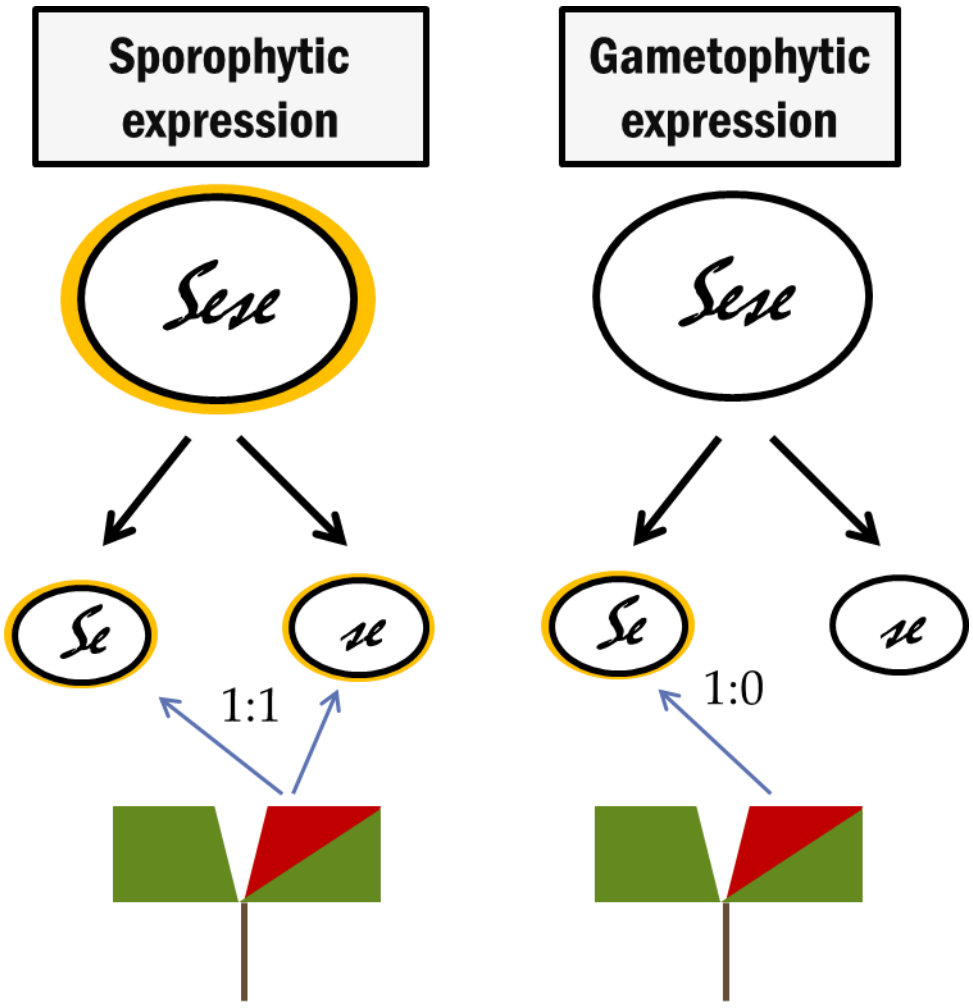

Figure 3.5. Sporophytic versus gametophytic expression illustration. The yellow ring represents gene expression. Red-colored sector in the seedlings represent haploid sectors originated from *Sese* parent. *Gb* (*Se*) and *Gh* (*se*) alleles would be found randomly (1:1) in the sectorial tissues if *Se* expression relies exclusively on sporophytic transcription, whereas only *Gb allele*s (1:0) would be found if expression relies exclusively on gametophytic transcription.

As suggested by Biddle, one of the hypotheses for the *Semigamy* expression was that it occurs in the gametophytes. Moreover, the *Gb allele* (*Se*) must be transmitted through the female gamete if the semigamy trait is to be expressed, *i.e.*, for semigamous reproduction to occur (Biddle 2006). Therefore, all of the maternal haploid cotyledonary sectors from *Sese* females were expected to have the *Gb* allele, *Se* (Figure 3.5). This argument and also cytogenetic tests made by Biddle support our theory that the behavior of the *Semigamy* gene through the ovule parent is determined by expression in the megagametophytes. If this theory proves to be true, then the currently accepted mode of gene action (co-dominant sporophytic action) of *Semigamy* is very possibly, if not likely, incorrect, i.e., it could easily be a recessive gametophtyic gene. In addition, the occurrence of chimeras in *SeSe* x *sese* crosses, but never in *sese* x *SeSe* crosses led our lab to suggest possible differences in *Se* expression in mega- versus micro-gametophytes, and/or the resulting zygotes.

Sese females testcrossed with *SeSe* pollen parents yielded 7531 progeny seed. When germinated under greenhouse conditions in commercial soil mix, 5365 germinated (71.24% germination rate); 93 chimeras were screened in detail, and 54 were considered to have well-defined borders and usefully large sectors. DNA samples from the 54 sectors were submitted to the SNP-based ploidy test, 39 sectors were treated as haploids by presence of a maternal *Gh* in at least one SNP allele or the absence of a hybrid genotype for all the SNP loci and thus considered for further genotyping analysis. For *Sese* as

the male parent testcrossed onto *SeSe* females, 2565 seeds were planted and 2197 germinated (85.65% germination rate); 127 chimeras were screened in detail, and 67 were found to have well-defined borders and usefully large sectors. DNA samples from the 67 sectors were submitted to the SNP-based ploidy test, and 48 were treated as haploids and used for further genotyping analysis.

According to Chi-square test of homogeneity, genotypic results of chimeric haploid sectors originated from populations used as *Sese* parents (F1 and BC1F1) for each test (maternal or paternal expression) did not depart significantly from homogeneity ($p > 0.05$), so the data were combined and analyzed collectively.

Based on the number of seeds recovered from each cross and the fact that approximately the same number of reciprocal crosses were made, we infer that *Semigamy* affects number of seeds per boll. The average number of seeds per boll was 7.3 when *Se* homozygous lines were used as females and 13.0 when the reciprocal cross was performed, i.e., when *Sese*-heterozygous plants were used as females. Moreover, the frequency of boll abortion was higher when homozygotes were the female parent. Although more seeds were obtained on *Sese* female, chimera recovery was much lower (1.73% female vs. 5.78% male). The difference can be attributed partially to the fact that *Semigamy* expression depends on the number of semigamous alleles present on the

female parent, so *SeSe* females will likely produce more chimeras than *Sese* females.

When we combine low levels of recovery and eliminate samples that were not 100% haploids (divergent ploidy on the 3 tissue layers), the population of true haploid sectors was very small. Some ideas for future work are: (1) To enlarge the population, increasing drastically the number of crosses to produce a higher amount of progeny seeds to be screened, (2) Combine a purely *G. barbadense* background to avoid inbreeding depression with a target chromosome segment substitution from *G. hirsutum* to be able to use interspecific SNPs in chimeric sectors.

Analysis of two *Se*-linked SNPs, Gb_010283 and Gb_016965, in the 39 haploid sectors that originated from heterozygous females (*Sese*) revealed that all but one sector carried the *Se*-linked *G. barbadense* allele (*Gb*), and the one exception included a recombinant between the two SNPs (planting # 58.2) (Figure 3.6, Table 3.3). These ratios strongly indicate that expression of the *Se* gene during megagametophyte stage is determining the origin of the haploid sectors, i.e., the semigamous reproduction leading to their formation.

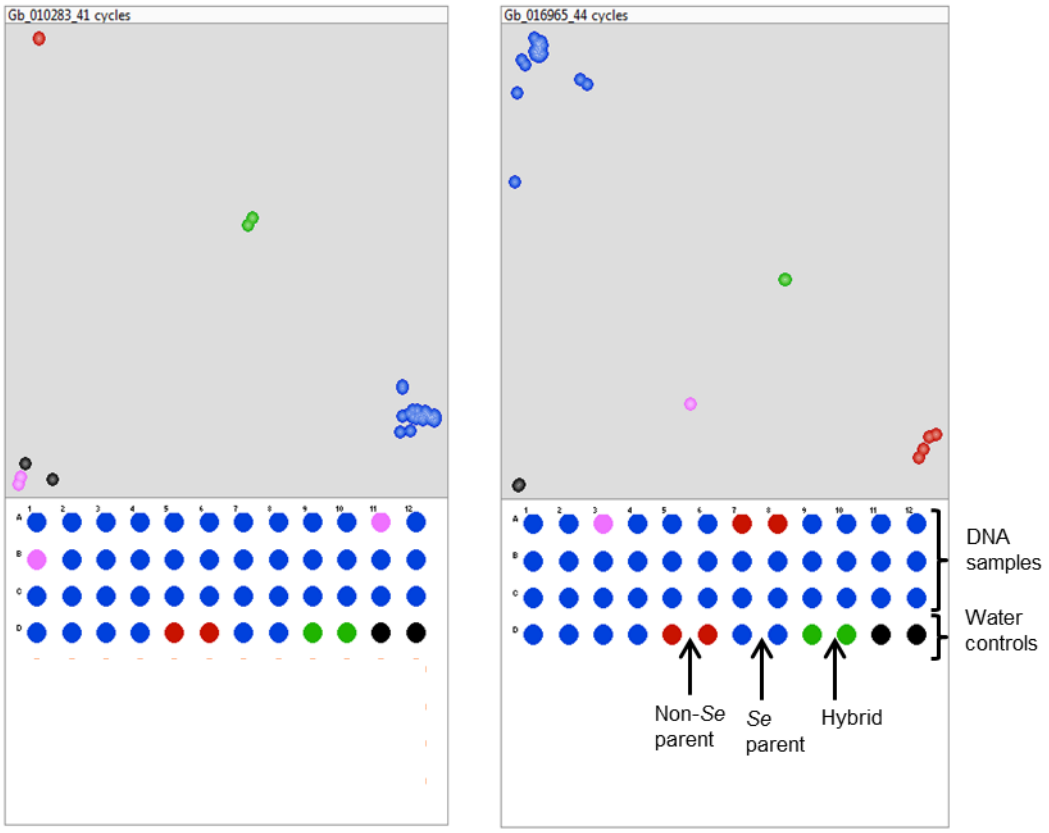

Figure 3.6. KlusterCaller figure. Genotypes of some haploid sectors originated from *Sese* female parents. The figure depicts results from the two *Se*-linked SNPs, Gb_010283 and Gb_016965. The red dots represent the non-semigamous (*Gh*) genotype, the blue dots represent the semigamous (*Gb*) genotype and the green dots represent the hybrid between *Gh* and *Gb*. Each sample was duplicated in the DNA plate. Despite some of the DNA samples not amplifying (pink dots), it was clear that female sector samples clustered almost exclusively in *Gb* position (*Se* parent). There was one recombinant between SNPs (A7/A8 position).

Table 3.3. Genotypes of maternally derived chimeric haploid sectors originating from a *Sese (GbGh)* x *SeSe (GbGb)* cross according to *Se*-linked and *Se*-independent SNP markers.

	Se-linked		*Se*-independent		
Plant_ #	Gb_010283	Gb_016965	UCcg10680_162	UCcg11310_419	UCcot10015_139
22	*Gb*	*Gb*	*Gh*	*Gb*	*Gb*
29	*Gb*	*Gb*	failed	*Gh*	*Gb*
36	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
47	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*
58.1	*Gb*	*Gb*	failed	*Gh*	failed
58.2	*Gb*	*Gh*	failed	*Gh*	failed
59	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
60	*Gb*	*Gb*	failed	failed	*Gb*
117.1	*Gb*	*Gb*	failed	*Gh*	*Gb*
117.2	*Gb*	*Gb*	failed	*Gh*	*Gb*
118	*Gb*	*Gb*	failed	*Gh*	*Gb*
148	*Gb*	*Gb*	failed	failed	*Gb*
149	*Gb*	*Gb*	*Gb*	failed	*Gb*
161	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
170	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
177	*Gb*	*Gb*	failed	failed	*Gb*
199	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
205	*Gb*	*Gb*	failed	*Gb*	*Gb*
238	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
243	*Gb*	*Gb*	*Gh*	failed	*Gh*
255	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*
290	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*
307	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*
308	*Gb*	*Gb*	*Gb*	failed	*Gb*
319	*Gb*	*Gb*	*Gb*	failed	*Gb*
329	*Gb*	*Gb*	failed	*Gh*	*Gh*
341.1	*Gb*	*Gb*	failed	*Gh*	*Gh*
341.2	*Gb*	*Gb*	failed	*Gh*	*Gh*
362	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
364	*Gb*	*Gb*	failed	*Gh*	*Gb*
367	*Gb*	*Gb*	failed	*Gb*	*Gh*

Table 3.3. Continued

	Se-linked		Se-independent		
Plant_ #	Gb_010283	Gb_016965	UCcg10680_162	UCcg11310_419	UCcot10015_139
369	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*
373	*Gb*	*Gb*	failed	*Gb*	*Gb*
375	*Gb*	*Gb*	failed	*Gb*	*Gb*
379.1	*Gb*	*Gb*	failed	*Gb*	*Gh*
379.2	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
386	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
398	*Gb*	*Gb*	*Gb*	failed	*Gb*
419.1	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*

The genotypic results of the 48 haploid sectors that originated from *Sese*-heterozygous males presented both *Gb* and *Gh* alleles (Figure 3.7; Table 3.4). In contrast to the genotypic uniformity observed among haploid sectors from maternal heterozygotes, ~25% of the paternal haploid sectors carried the *G. hirsutum* SNP allele rather than the *G. barbadense* allele. The ratio of sectors was significantly different from 1:1 for *Gh:Gb* alleles ($p < 0.05$), which indicates that microgametophytic expression influences the incidence of semigamous reproduction, and/or that transmission of *Se*- and *se*-linked SNPs via pollen was distorted for other reasons. The frequencies of *Se* recovery between paternal and maternal haploid sectors significantly diverged from each other ($p < 0.05$), which strengthens our hypothesis that the *Semigamy* gene behaves differently between female and male plants, or that its pollen-mediated transmission is considerably influenced by other factors.

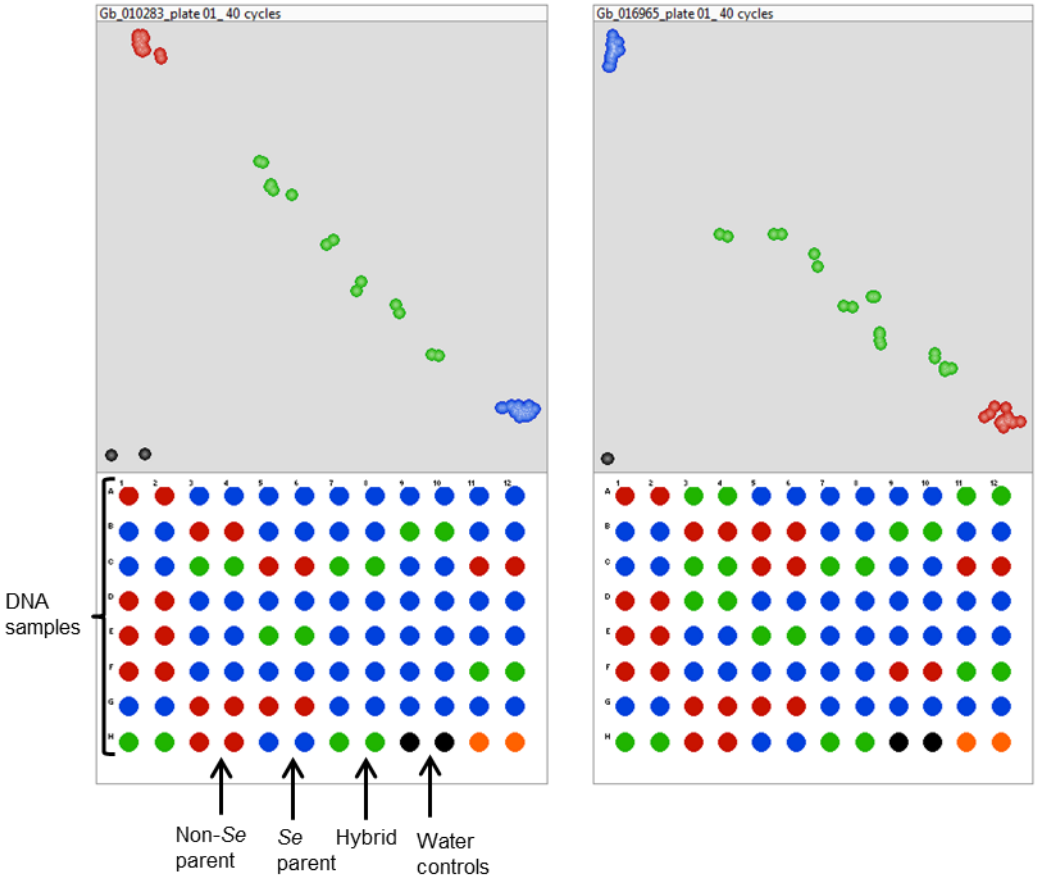

Figure 3.7. KlusterCaller figure. Genotypes of some haploid sectors originating from *Sese* male parents. The figure depicts results from the two *Se*-linked SNPs, Gb_010283 and Gb_016965. The red dots represent the non-semigamous (*Gh*) genotype, the blue dots represent the semigamous (*Gb*) genotype and the green dots represent the hybrid between *Gh* and *Gb*. Each sample was duplicated in the DNA plate. Samples that did not cluster in *Gb* or *Gh* position were considered tetraploids or mixture of tissues (n + 2n) and therefore, were not included in further genotypic analysis. In contrast with *Sese* female originated sectors, *Sese* male originated sectors clustered in *Gb* as well as in *Gh* position.

Table 3.4. Genotypes of paternally derived chimeric haploid sectors originating from a *SeSe (GbGb)* x *Sese (GbGh)* cross according to *Se*-linked and *Se*-independent SNP markers.

	Se-linked		*Se*-independent			
Plant_#	Gb_010283	Gb_016965	UCcg10680_162	UCcg11310_419	UCcot10015_139	UCcot10322_62
3	*Gh*	*Gh*	*Gb*	*Gb*	*Gh*	*Gh*
7	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*
10	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
16	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*
24	*Gh*	*Gh*	*Gb*	*Gh*	*Gh*	*Gh*
32	*Gb*	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*
35	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*
52.2	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*
63	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*
64	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*
67	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*	*Gb*
92	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*
96	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*	*Gh*
104	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*	*Gh*
105	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*	*Gh*
113	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*	*Gb*
114	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*	*Gb*
118	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*	*Gb*
120	*Gh*	*Gh*	*Gh*	*Gb*	*Gh*	*Gb*
121	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*	*Gh*
126	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*	*Gh*
160	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*	*Gb*
165	*Gb*	*Gh*	*Gb*	*Gh*	*Gh*	*Gb*
173.2	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*
174	*Gh*	*Gh*	*Gh*	*Gh*	*Gb*	*Gb*
197	*Gh*	*Gh*	*Gb*	*Gh*	*Gh*	*Gb*
203	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*	*Gb*
207	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*	*Gb*
215	*Gb*	*Gb*	*Gh*	*Gb*	*Gb*	*Gh*
226	*Gb*	*Gb*	*Gb*	failed	*Gh*	*Gb*

Table 3.4. Continued

	Se-linked		*Se*-independent			
Plant_ #	Gb_010283	Gb_016965	UCcg10680_ 162	UCcg11310_ 419	UCcot10015_ 139	UCcot10322_ 62
239	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*	*Gb*
248	*Gh*	*Gh*	*Gh*	failed	*Gb*	*Gh*
252	*Gh*	*Gh*	*Gb*	*Gb*	*Gh*	*Gb*
260	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*	*Gh*
264	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*	failed
286.1	*Gb*	*Gb*	*Gh*	*Gb*	*Gb*	*Gh*
286.2	*Gb*	*Gb*	*Gh*	failed	*Gh*	*Gh*
287.1	*Gb*	*Gb*	*Gb*	failed	*Gh*	*Gb*
293	*Gb*	*Gb*	*Gb*	failed	*Gb*	failed
294	*Gb*	*Gb*	*Gh*	failed	*Gh*	*Gb*
298	*Gb*	*Gb*	*Gh*	*Gh*	*Gh*	*Gb*
299	*Gb*	*Gb*	*Gh*	*Gb*	*Gh*	*Gh*
312	*Gb*	*Gb*	*Gb*	*Gb*	*Gh*	*Gb*
319	*Gb*	*Gb*	*Gb*	failed	*Gb*	*Gb*
323.1	*Gb*	*Gb*	*Gb*	failed	*Gh*	*Gh*
323.2	*Gb*	*Gh*	*Gb*	failed	*Gb*	*Gb*
329	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*
331	*Gb*	*Gb*	*Gb*	*Gh*	*Gh*	*Gb*

SNP marker UCcg10220_69 did not give good amplification and/or clustering results in the KlusterCaller software across all plate samples and it was not used for calling genotypes. For unknown reasons, SNP marker UCcot10322_62 did not amplify and/or efficiently cluster several samples originated from *Sese* female parents and it was not used for calling genotypes of maternally derived sectors. SNP UCcg10680_162 also did not amplify/cluster efficiently some samples originated from *Sese* females but it could be partially used for ploidy test. Therefore, five markers were used for the ploidy test of

maternally derived haploid sectors (two *Se*-linked and three *Se*-independent) and six markers were used for the ploidy test of paternally derived haploid sectors (two *Se*-linked and four *Se*-independent). The SNP data on maternal haploid sectors that originated from *Sese* females indicates that *Gb Se* is transmitted to all or virtually all semigamous megagametophytes and egg cells, confirming by independent methods the deduction by Biddle (2006) that the *Semigamy* gene is expressed in the female gametophytes. Conversely, our results from haploid sectors that originated from *Sese* males raise questions about paternal expression of *Semigamy*. The data indicate that paternal inheritance does not follow the patterns expected based on sporophytic *or* gametophytic expression, i.e., it follows neither 1:1 ($p < 0.05$) nor 1:0 ($p = 0$), respectively, for *Se:se* (*Gb*:*Gh*).

Several hypotheses that might explain the unusual behavior of the gene when derived from a heterozygous pollen parent were made. (1) The genotype of the egg cell, megagametophyte or functional megaspore affects zygotic behavior such that it largely but incompletely determines the critical karyogamy-determining event(s) and incompletely over-rides paternal and/or microgametophytic genotypic effects. (2) Paternal determination involves a mixture of microgametophytic and paternal sporophytic gene action effects, such that the distribution of *Gb* versus *Gh allele*s in chimeric haploid sectors would be neither sporophytic-like (1:1) nor gametophytic-like (1:0), but a combination of both. (3) *Semigamy* is expressed by a different locus in male parents i.e., the

semigamous apparatus of heterokaryonic zygotes is composed by more than one gene, one responsible for transcripts in the female parent and the other (s) determining the paternal effects on rates of semigamous reproduction, e.g., a *Se*-homeolog. In this case, SNP markers used would be linked to the “female” locus and far from the “male” locus (loci). The reason why more than one locus is being considered for the male parent is that there could be two independent recessive factors to cause semigamous development, in which case 25% transmission of *se* allele would be expected. My data seem to exclude such an effect by just one gene, since that would lead to a 1:1 distribution. (4) Epistatic interactions affect *Semigamy* in the male parent before/during gamete formation, e.g. by one or more other loci. (5) Epigenetics effects on the male genome could alter the way that the gene is being expressed on the male parent. Gutierrez-Marcos et al. (2006) indicated some methylation asymmetry between male and female parental alleles during gametogenesis in maize (Gutierrez-Marcos, Costa et al. 2006). (6) Recombination rates differ in male and female parents, which together present altered relationships between the *Se* gene and *Se*-linked SNPs. Jessup et al. (2003) observed variable chromosomal behavior between male and female gametes in buffelgrass (*Pennisetum ciliare* L.). They suggested a possible difference in the frequency of crossing-over events along the chromosomes of male versus female parents to explain the differences in repulsion-phase association across loci of maternal and paternal maps. (Jessup, Burson et al. 2003). Thomas and Rothstein (1991) suggested various insights

about region-specific differences in recombination between sexes. Although their subject of study was human meiosis, their explanations can perhaps be expanded to other organisms. They reported that “male and female gametes undergo independent meiotic processes in different tissues, and the differences in recombination observed may reflect the physiologies of the meiotic cells of each sex”. Another possibility was that “recombination may be regulated by the synthesis of sex-specific factors that control a regulatory step in the recombination event”, thus male and female would have different recombination events depending on the sex-specific regulation of necessary enzymes (Thomas and Rothstein 1991). More studies related to the *Semigamy* expression in male and female parents are definitely necessary to answer those questions.

Conclusions

The sector-based SNP data clearly show that the SNPs are linked with the *Se* gene, and that the *Semigamy* gene is inherited at different rates when being transmitted from female versus male *Sese* heterozygotes. All haploid sectors that arose from female heterozygous parents contained only the *Gb allele* (*Se*), whereas the analogous sectors from male heterozygotes contained either the *Gb allele* (*Se*) allele (~75%) or the *Gh* (*se*) allele (~25%). The results indicate that semigamous reproduction depends on megagametophytic expression of the *Semigamy* allele, and/or, perhaps just as likely, on the lack of megagametophytic expression by a *Gh allele*, i.e., only egg cells with a *Gb*

Semigamy allele would undergo akaryogamous fertilization, and therefore only the *Se* allele would be found in maternal haploid sectors of a chimeric progeny. In contrast, the data for haploid sectors from male *Sese* heterozygous parents indicate they transmit both *Gb* (*Se*) and *Gh* (*se*) alleles. Transmission of the *se* allele discounts any possibility that paternal effects on rates of semigamous reproduction are *exclusively* determined by microspore or microgametophytic expression manner, but it does not preclude the possibility of *partial* determination by microgametophytic expression. The ratio of *Se:se* paternally derived haploid sectors (~3:1) also indicated that a single independent locus could not account for observations ($p < 0.05$). Leading hypotheses to explain the paternal transmission of the *Gh allele* are as follows. (1) The genotype of the egg cell, megagametophyte or functional megaspore largely but incompletely determines the critical karyogamy; the paternal effect is limited -- influencing but not determining. (2) Paternal determination involves a mixture of microgametophytic and paternal sporophytic gene action effects. (3) *Semigamy* is expressed by a different locus in male parents i.e., the semigamous apparatus is composed by more than one gene, e.g. an *Se*-homeolog. (4) Epistatic interactions affect *Semigamy* in the male parent before/during gamete formation. (5) Epigenetics effects on the male genome. (6) Recombination rates differ in male and female parents.

It is essential that further studies related to the *Semigamy* expression continue to be carried in order to answer those questions and enable the use of

the full potential of this gene. Hopefully the results of this present work will help future studies on the *Semigamy* gene.

CHAPTER IV

HYPOTHESES ON *SEMIGAMY* MODEL

Based on the characteristics of plants that possess the *Semigamy* gene, i.e., non-fusion of gametes nuclei, we hypothesize that the non-semigamous (*se*) gene encodes a protein(s) essential for nuclear fusion during fertilization. Given the cytological proximity of gametic nuclei, we think the defective process involves failure to initiate the nuclear envelope bridges that are normally observed in nuclear fusions (Stelly, personal communication), i.e., perhaps involving the outer membrane. It would not be surprising to find that the protein is embedded in or attaches to proteins embedded in the nuclear envelope of the egg cell nucleus. The fact that the incidence of akaryogamy is ~100% in *SeSe* homozygotes (Biddle 2006; Hodnett 2006) suggests that the semigamous plants might be entirely lacking the necessary molecule, rather than under-producing it. The fact that the semigamy does not seem to affect central cell triple fusion (two polar nuclei and one sperm cell nucleus) suggests that the *se* gene might be expressed after formation of the egg cell, or that its function is cell-specific. A mutation in this gene could cause any of several problems, e.g., inadequate RNA synthesis, incorrect RNA sequence, improper RNA processing, inadequate protein synthesis, ineffective trafficking, or incorrect protein sequence. The question of gene action is of interest, since the implications for a dominant mutation are quite different than for recessive one. A dominant mutation might

produce a transcript that compromises the processing of RNA from one or more loci, or a protein that fouls a multi-protein complex, while a recessive one might simply be a null mutant, with no transcript, a non-functional transcript, or a non-functional protein. Based on these assumptions, some hypotheses on the Semigamy model are suggested.

Hypotheses

Hypothesis 1. The *se* gene encodes a protein that is expressed in the egg cell and is crucial to nuclear fusion, or at least karyogamous nuclear fusion, while the proteins in the sperm are helpful, but they're not as essential for nuclear fusion as egg proteins. That's why plants must have the *Se* gene in the female parent in order for *Semigamy* to be expressed.

Hypothesis 2. There is more than one gene involved in this bizarre reproductive behavior and they interact epistatically with each other. Those genes would be related to gametogenesis. One of the genes, *Semigamy*, would be responsible for the expression of nuclear membrane proteins in the megagametophyte and also for triggering, secondarily, the activation of sperm nuclear proteins for nuclear fusion. The other gene would be responsible for the expression of nuclear membrane proteins in the microsporophyte. The presence of *Semigamy* would affect the expression of the nuclear membrane proteins in the megagametophyte and the efficient activation of the sperm nuclear proteins for nuclear fusion. If *Semigamy* is present only in the female, the nuclear

membrane proteins will not be expressed correctly and nuclei fusion will be often dismissed. If *Semigamy* is present only in the male, that will not affect megagametophyte membrane proteins or the triggering of sperm nuclear proteins so, fusion will occur normally. Because the occurrence of haploids/chimeras increases if both parents have the *Se* gene, that leads us to believe that *Semigamy* must influence, secondarily, the gene responsible for the expression of sperm nuclear proteins. This hypothesis was based on the indication by Sprunck et al. that the egg cell secrets Egg Cell 1 (EC1) proteins during egg and sperm attachment triggering the redistribution of sperm protein Hapless2 (Hap2) throughout the sperm cell surface to achieve successful gamete fusion (Sprunck, Rademacher et al. 2012). Despite Sprunck's indication being related to cell fusion, perhaps analogous interactions are to nuclei fusion.

Hypothesis 3. Gametic cell cycles do not synchronize completely and nuclear fusion does not occur. Gamete interactions during fertilization in angiosperms consist of a series of complex events involving several proteins and molecules responsible for gamete adhesion, syngamy and karyogamy. One of the steps that sperm and egg cells have to overcome during fertilization is the synchronization of their cell cycle (Sprunck and Dresselhaus 2009). In tobacco, sperm cells have 1C DNA content during pollen tube elongation and increase their DNA content from 1C to about 2C until encountering the egg cell. The egg cell also increases its DNA content upon arrival of pollen tube and cell fusion occurs after completion of S-phase. This signifies that egg and sperm cells

should be synchronized in DNA content and the phase of the cell cycle in order to interact and fuse (Tian, Yuan et al. 2005). In this hypothesis, the *Semigamy* gene alters the protein (s) responsible for signaling for cell cycle synchronization. If this protein is more active in the egg cell than in the sperm cell, the effect of the gene will be much higher on the female parent, causing chimeras and haploids only when the gene is present in the megagametophytes.

Mechanism of Nuclear Fusion and Possible Candidates for the *Semigamy* Gene

After syngamy and nuclear migration, the egg and one sperm nuclei go through successive nuclear membrane fusion events, first of the outer-membranes and then of the inner membranes. Both fusions during double-fertilization i.e., sperm and egg nuclei, and sperm and central cell nuclei, follow that same overall pattern (Jensen 1964). The endoplasmic reticulum (ER) is contiguous with the nuclear envelope's outer membrane, and has been found to play an important role in nuclear envelope fusion, in that the outer-nuclear membranes from both nuclei come in contact through the ER and then, fusion takes place (Ohnishi, Hoshino et al. 2014). Recently, Maruyama et al. (2014) reported that ER-resident proteins from the well-conserved J-domain interact with other proteins to regulate polar nuclear membrane fusion in *Arabidopsis thaliana*. The immunoglobulin binding protein (BiP) functions as an ER-resident chaperone of the heat shock protein 70 (Hsp70) family and it interacts with ER-

resident J-proteins resulting in nuclear fusion. According to the authors, *bip1bip2* double-mutant female gametophytes failed to undergo fusion and instead contained two unfused nuclei lying beside each other (Maruyama, Yamamoto et al. 2014).

After I screened the mRNA sequences for proteins found to be related to nuclei fusion in *A. thaliana* against the *Gossypium raimondii* reference genome (D5), sequences similar to bip1 and bip2-like proteins were found on a genomic region similar to chromosome 12, but in a significant physical distance away from where the SNPs used in this work are located (about 8Mb, proximally). Also, a HAP2-like protein (related to sperm cell surface in *A. thaliana*) was found on a genomic region similar to chromosome 12, about 10Mb distal from the location of the SNPs. A list of annotated genes present on the target region from the D5 reference genome was analyzed. Based on the function of those genes, some were selected as being possible candidates for the *Semigamy* gene or related to it. Among them are a J-domain related chaperone protein, a microtubule-associated protein 70, a heat shock transcription factor and a male gametophyte defective 3 protein.

CHAPTER V

FINAL CONCLUSIONS

As an “apomictic-like” gene, *Semigamy* has been used successfully as an efficient tool for its relatively high monoembryonic haploid frequency in several cotton research programs with emphasis in mapping studies utilizing *Se*-derived haploid and doubled haploid populations and layer-specific gene expression in cotton leaves from chimeric plants. Moreover, it’s been proven that *Semigamy* is much more effective as a haploid producer than the polyembryony system. Classical genetic segregation analysis of testcross populations based on chimeric progeny screening identified a mixture of maternal/paternal chimeras, haploids, and zygotic progeny which led to the hypothesis of incompletely dominant genetic control that could cause plants to reproduce semigamously at frequencies related to *Se* gene dosage of both parents.

Recently, our lab developed a cytological "clearing method" for reliable cytological analysis of fertilization events in cotton ovules resulting from several intercrosses among homozygous *Semigamy* (*Sese*), heterozygotes (*Sese*) and non-semigamous (*sese*) plants. That method provided cytological proof of semigamous reproduction in cotton (Hodnett, Crane et al. 1997, Biddle 2006, Hodnett 2006). Moreover, the overall results from Biddle's (2006) research showed that the actual frequencies of semigamous ovule production differ markedly from the rates that had been previously reported, because those had

been deduced from progeny scoring rather than by direct cytological observation. One of the hypotheses made based on her results was that the *Se* locus is expressed at the gametophytic stages, in which case the mode of gene action cannot be inferred from available data. Combining cytological evaluation and molecular analysis using SSR markers, close linkage was detected between the *Semigamy* locus with a couple of SSRs in chromosome 12 of *Gossypium barbadense.*

In order to verify past results from our lab, objectives of the present study included linkage mapping of the *Se* gene using SNP markers, and analyzing *Se* expression in male and female parents. SNPs localized in the same genomic region where *Semigamy* is putatively located according to the reference D5 genome were used to genotype chimeric haploid sectors and for MAS in a backcross population. Genotypic results of the haploid sectors and backcross population confirmed that those SNPs are linked to the *Semigamy* gene. There is some incongruity about the chromosomal location of those SNPs in *Gossypium barbadense*, whether they are in chromosome 12 and/or in its homeolog, chromosome 26 and this issue must be addressed shortly. A few genes listed on our D5 target region also were pointed as possible candidates to be related to or actually be the *Semigamy* gene. A future high-resolution density map combined with population phenotypic data (capability of producing haploid/chimeric progeny) or cytological data (non-fusion of gametes nuclei) is

essential to improve results and narrow down the *Semigamy* genomic location in *Gossypium barbadense*.

Genotypic results of chimeric haploid sectors originating from female and male heterozygous (*Sese*) parents indicate that semigamous reproduction depends on megagametophytic expression of the *Semigamy* allele. Given megagametophytic determination and the haploid state of the megagametophyte, there has been no legitimate means to deduce whether or not the "gene action" of the *Gb allele Semigamy* is dominant, recessive or something in between. With this new information, there is no basis for considering *Se* gene action in the megagametophytes to be incompletely dominant. The creation of a bi-allelic megagametophytic genotype that is viable and transmissible could be a possible way to determine the mode of gene action in the megagametophytes. In contrast, the data for haploid sectors from male *Sese* heterozygous parents indicate that they transmit both *Gb* (*Se*) and *Gh* (se) alleles. Transmission of the *se* allele discounts any possibility that paternal effects on rates of semigamous reproduction are *exclusively* determined by expression in the microspore or microgametophyte, but it does not preclude the possibility of *partial* determination by microgametophytic expression. However, the observed ~3:1 ratio could be due to other factors that significantly biased the recovery of Se-linked SNPs. Some other possibilities are (1) *Se* is expressed on the paternal side but is not the sole determinant; (2) Epistatic interactions affect *Semigamy* in the male parent before/during gamete formation, e.g. by one or

more other loci; (3) Recombination rates differ in male and female parents, which presents altered relationships between the *Se* gene and *Se*-linked SNPs; (4) A non-*Se* locus (not linked to the *Se*-linked SNPs), e.g., a *Se*-homeolog, is determining the paternal effects on rates of semigamous reproduction; (5) Epigenetics effects on the male genome.

Right now, our understanding of the *Semigamy* gene is still not comprehensive, which limits its manipulation, but our efforts to localize the gene and to analyze its reproductive behavior can be valuable to future research. A deeper comprehension of the mode of gene action, expression and molecular biology of *Semigamy* would likely help explain why there are differences in frequencies of haploids and chimeras in reciprocal crosses among semigamous and non-semigamous plants. In addition, the localization of the gene to a well-characterized genetic marker region in the cotton genome would enable marker-assisted selection within a segregating population during research with the gene and introgression to other cotton species as well as to make one step further toward the possibility of cloning the gene to use it in different agronomic crops.

REFERENCES

Agarwal, M., N. Shrivastava and H. Padh (2008). "Advances in molecular marker techniques and their applications in plant sciences." Plant Cell Reports **27**(4): 617-631.

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

Asker, S. and L. Jerling (1992). Apomixis in Plants, CRC press.

Battaglia, E. (1946). "Ricerche cariologiche ed embriologiche sul genere Rudbeckia (Asteraceae) VI. Apomissia in *Rudbeckia speciosa* Wender." Plant Biosystem **53**(1-2): 27-69.

Battaglia, E. (1955). "Unusual cytological features in the apomictic *Rudbeckia* Sullivantii Boynton et Beadle: (with 56 figures)." Caryologia **8**(1): 1-32.

Battaglia, E. (1963). “Apomixis.” In: Recent Advances in Embryology of Angiosperms: 221–264. International Society of Plant Morphology, University of Delhi.

Battaglia, E. (1980). "81 Embryological questions: 3. Semigamy, hemigamy and gynandroembryony." Annali Bot **39**: 173-175.

Batygina, T. and G. Y. Vinogradova (2007). "Phenomenon of polyembryony. Genetic heterogeneity of seeds." Russian Journal of Developmental Biology **38**(3): 126-151.

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

Biddle, K. D. (2006). An Analysis of Alternative Forms of Plant Reproduction Using *Gossypium barbadense* and *Arabidopsis thaliana*, Texas A&M University.

Blakeslee, A. F. and J. Belling (1924). "Chromosomal mutations in the Jimson weed, *Datura stramonium*." Journal of Heredity **15**(5): 195-206.

Blakeslee, A. F., J. Belling, M. Farnham and A. D. Bergner (1922). "A haploid mutant in the jimson weed," *Datura stramonium*"." Science (New York, NY) **55**(1433): 646.

Blenda, A., J. Scheffler, B. Scheffler, M. Palmer, J.-M. Lacape, Z. Y. John, C. Jesudurai, S. Jung, S. Muthukumar and P. Yellambalase (2006). "CMD: a cotton microsatellite database resource for *Gossypium* genomics." BMC Genomics **7**(1): 132.

Calzada, J.-P. V., C. F. Crane and D. M. Stelly (1996). "Apomixis--The Asexual Revolution." Science **274**(5291): 1322-1323.

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.

Chalyk, S. (1994). "Properties of maternal haploid maize plants and potential application to maize breeding." Euphytica **79**(1-2): 13-18.

Chaudhari, H. (1978). "Use of semigamy in the production of cotton haploids." Bulletin of the Torrey Botanical Club: 98-103.

Chaudhari, H. (1979). "The production and performance of doubled haploids of cotton." Bulletin of the Torrey Botanical Club: 123-130.

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

Coe, G. E. (1953). "Cytology of reproduction in *Cooperia pedunculata*." American Journal of Botany: 335-343.

Dolan, L. and R. S. Poethig (1998). "The Okra leaf shape mutation in cotton is active in all cell layers of the leaf." American Journal of Botany **85**(3): 322-327.

Eder, J. and S. Chalyk (2002). "In vivo haploid induction in maize." Theoretical and Applied Genetics **104**(4): 703-708.

Feaster, C. V. and E. Turcotte (1973). "Yield stability in doubled haploids of American Pima cotton." Crop Science **13**(2): 232-233.

Forster, B.P.,E. Heberle-Bors, K.J. Kasha and A.Touraev (2007). “The resurgence of haploids in higher plants.” Trends in Plant Science **12**(8): 368-375.

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

Fryxell, P. A. (1984). "Taxonomy and germplasm resources." Cotton(cotton): 27-57.

Griffing, B. (1975). "Efficiency changes due to use of doubled-haploids in recurrent selection methods." Theoretical and Applied Genetics **46**(8): 367-385.

Gutierrez-Marcos, J.F., L.M. Costa, M. Dal Pra, S. Scholten, E. Kranz, P. Perez and H.G. Dickinson (2006). “Epigenetic asymmetry of imprinted genes in plant gametes.” Nature Genetics **38**(8): 876-878.

Gwyn, J. J. and D. M. Stelly (1990). "Time and mode of gene expression of the semigamy mutant of Pima cotton." p. 70. In J. Brown and D. Richter (eds.) Proc. 1990. Beltwide Cotton Production Research Conference, Las Vegas, NV.

Han, Z.-G., W.-Z. Guo, X.-L. Song and T.-Z. Zhang (2004). "Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton." Molecular Genetics and Genomics **272**(3): 308-327.

Hanna, W. and E. Bashaw (1987). "Apomixis: its identification and use in plant breeding." Crop Science **27**(6): 1136-1139.

Hermsen, J. T. and J. Verdenius (1973). "Selection from *Solanum tuberosum* group phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot." Euphytica **22**(2): 244-259.

Hodnett, G. L. (2006). The Effect of the *Semigamy* (*Se*) Mutant on the Early Development of Cotton (*Gossypium barbadense* L.), Texas A&M University.

Hodnett, G. L., C. F. Crane and D. M. Stelly (1997). "A rapid stain-clearing method for video based cytological analysis of cotton megagametophytes." Biotechnic & Histochemistry **72**(1): 16-21.

Hulse-Kemp, A.M., H. Ashrafi, X. Zheng, F. Wang, K.A. Hoegenauer, A.B.V. Maeda, S.S. Yang, K. Stoffel, M. Matvienko, K. Clemons (2014). “Development and bin mapping of gene-associated interspecific SNPs for cotton (*Gossypium hirsutum* L.) introgression breeding efforts.” BMC Genomics **15**: 945.

Jacobs, S. W. (2000). Grasses: Systematics and Evolution, CSIRO Publishing.

Jensen, W. A. (1964). "Observations on the fusion of nuclei in plants." The Journal of Cell Biology **23**(3): 669-672.

Jessup, R. W., B. L. Burson, G. B. Burow, Y. W. Wang, C. Chang, Z. Li, A. H. Paterson and M. A. Hussey (2002). “Disomic inheritance, suppressed recombination, and allelic interactions govern apospory

in buffelgrass as revealed by genome mapping.” Crop Sciences **42**(5): 1688 – 1694.

Jessup, R. W., B. L. Burson, G. B. Burow, Y. W. Wang, C. Chang, Z. Li, A. H. Paterson and M. A. Hussey (2003). “Segmental allotetraploidy and allelic interactions in buffelgrass (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* (L.)) as revealed by genome mapping.” Genome **46**(2): 304-313.

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

Koltunow, A. M., R. A. Bicknell and A. M. Chaudhury (1995). "Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization." Plant Physiology **108**(4): 1345.

Lacape, J.-M., T.-B. Nguyen, B. Courtois, J.-L. Belot, M. Giband, J.-P. Gourlot, G. Gawryziak, S. Roques and B. Hau (2005). "QTL analysis of cotton fiber quality using multiple× backcross generations." Crop Science **45**(1): 123-140.

Lanaud, C. (1988). "Origin of haploids and semigamy in *Theobroma cacao* L." Euphytica **38**(3): 221-228.

Lashermes, P. and M. Beckert (1988). "Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines." Theoretical and Applied Genetics **76**(3): 405-410.

Li H. and R. Durbin (2009). “Fast and accurate short read alignment with Burrows-Wheeler Transform.” Bioinformatics **25**:1754-60.

Liu, W., M. Y. Zheng, E. A. Polle and C. F. Konzak (2002). "Highly Efficient Doubled-Haploid Production in Wheat (L.) via Induced Microspore Embryogenesis." Crop Science **42**(3): 686-692.

Maheshwari, P. (1950). An Introduction to the Embryology of Angiosperms, New York: McGraw-Hill.

Mahill, J. F., J. N. Jenkins, J. McCarty and W. Parrott (1984). "Performance and stability of doubled haploid lines of upland cotton derived via semigamy." Crop Science **24**(2): 271-277.

Maluszynski, M., K. Kasha and I. Szarejko (2003). “Published doubled haploid protocols in plant species.” Doubled Haploid Production in Crop Plants, Springer: 309-335.

Maruyama, D., M. Yamamoto, T. Endo and S.-i. Nishikawa (2014). "Different Sets of ER-Resident J-Proteins Regulate Distinct Polar Nuclear-Membrane Fusion Events in *Arabidopsis thaliana*." Plant and Cell Physiology **55**(11): 1937-1944.

Melchinger, A. E., W. Schipprack, H. Friedrich Utz and V. Mirdita (2014). "In Vivo Haploid Induction in Maize: Identification of Haploid Seeds by Their Oil Content." Crop Science **54**(4): 1497-1504.

Nogler, G. A. (1984). Gametophytic Apomixis, Springer.

Ohnishi, Y., R. Hoshino and T. Okamoto (2014). "Dynamics of male and female chromatin during karyogamy in rice zygotes." Plant Physiology **165**(4): 1533-1543.

Paterson, A. H., J. F. Wendel, H. Gundlach, H. Guo, J. Jenkins, D. Jin, D. Llewellyn, K. C. Showmaker, S. Shu and J. Udall (2012). "Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres." Nature **492**(7429): 423-427.

Poehlman, J. and D. Sleper (1995). Breeding Field Crops, Iowa State University Press, Ames, Iowa.

Prasanna, B., V. Chaikam and G. Mahuku (2012). Doubled Haploid Technology in Maize Breeding: Theory and Practice, CIMMYT.

Pupilli, F. and G. Barcaccia (2012). "Cloning plants by seeds: inheritance models and candidate genes to increase fundamental knowledge for engineering apomixis in sexual crops." Journal of Biotechnology **159**(4): 291-311.

Ramulu, K., V. Sharma, T. Naumova, P. Dijkhuis and M. van Lookeren Campagne (1999). "Apomixis for crop improvement." Protoplasma **208**(1-4): 196-205.

Rao, P. and D. Narayana (1980). "Occurrence and identification of semigamy in *Coix aquatica* (tribe Maydeae)." Journal of Heredity **71**(2): 117-120.

Raven, P. H., R. F. Evert and S. E. Eichhorn (2005). Biology of Plants, Macmillan.

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

Reinisch, A. J., J.-M. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendel and A. H. Paterson (1994). "A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome." Genetics **138**(3): 829-847.

Richards, A. (2003). "Apomixis in flowering plants: an overview." Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **358**(1434): 1085-1093.

Richards, A. J. (1997). Plant Breeding Systems, Garland Science.

Rober, F., G. Gordillo and H. Geiger (2005). "In vivo haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding." Maydica **50**(3/4): 275.

Snape, J. (1989). "Doubled haploid breeding: theoretical basis and practical applications." Review of Advances in Plant Biotechnology **2**: 19-31.

Spillane, C., M. D. Curtis and U. Grossniklaus (2004). "Apomixis technology development—virgin births in farmers' fields?" Nature Biotechnology **22**(6): 687-691.

Sprunck, S. and T. Dresselhaus (2009). "Gamete interactions during double fertilization in flowering plants." Zellbiologie Aktuell **35**: 22-27.

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

Stebbins, G. L. (1941). "Apomixis in the angiosperms." The Botanical Review **7**(10): 507-542.

Stelly, D. M., J. A. Lee and W. L. Rooney (1988). "Proposed Schemes for Mass-Extraction of Doubled Haploids of Cotton." Crop Science **28**(6): 885-890.

Thomas, B. J. and R. Rothstein (1991). "Sex, maps, and imprinting." Cell **64**(1): 1-3.

Tian, H. Q., T. Yuan and S. D. Russell (2005). "Relationship between double fertilization and the cell cycle in male and female gametes of tobacco." Sexual Plant Reproduction **17**(5): 243-252.

Tilney-Bassett, R.A.E. (1986). Plant Chimeras, Edward Arnold.

Turcotte, E. and C. Feaster (1974). "Methods of producing haploids: semigametic production of cotton haploids." Haploids in Higher Plants: Advances and Potential. Ontario Agric. College, Univ. of Guelph, ON, Canada: 53-64.

Turcotte, E. and C. Feaster (1975). “Inheritance of semigamy in American Pima cotton (*Gossypium barbadense* L.).” Agronomy Abstracts. ASA, Madison, WI.

Turcotte, E. and C. V. Feaster (1963). "Haploids: High-frequency production from single-embryo seeds in a line of Pima cotton." Science **140**(3574): 1407-1408.

Turcotte, E. and C. V. Feaster (1967). "Semigamy in pima cotton." Journal of Heredity **58**(2): 55-57.

Turcotte, E. and C. V. Feaster (1969). "Semigametic production of haploids in Pima cotton." Crop Science **9**(5): 653-655.

Wakana, A. and S. Uemoto (1987). "Adventive embryogenesis in *Citrus*. I. The occurrence of adventive embryos without pollination or fertilization." American Journal of Botany: 517-530.

Wendel, J. F., J. McD. Stewart and J. Rettig (1991). "Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*." Evolution: 694-711.

Wilms, H., J. Van Went, M. Cresti and F. Ciampolini (1983). "Adventive embryogenesis in *Citrus*." Caryologia **36**(1): 65-78.

Winkler, H. K. A. (1908). Parthenogenesis und Apogamie im Pflanzenreiche. G. Fischer.

Zhang, J., W. Guo and T. Zhang (2002). "Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L.× *Gossypium barbadense* L.) with a haploid population." Theoretical and Applied Genetics **105**(8): 1166-1174.

Zhang, J., J. Stewart and R. Turley (1999). "Genetic and molecular characterization of *Semigamy* expression in cotton (*Gossypium barbadense*)." Special Reports-University of Arkansas Agricultural Experiment Station **193**: 275-277.

Zhu, Q. H., A. Spriggs, J.M. Taylor, D. Llewellyn and I. Wilson (2014). "Transcriptome and complexity-reduced, DNA-based identification of intraspecies single-nucleotide polymorphisms in the polyploid *Gossypium hirsutum* L." G3: Genes|Genomes|Genetics **4**:1893-1905.