DISSECTING KEY BIOMASS TRAITS THROUGH TRADITIONAL QTL MAPPING, EXPLORING NOVEL FUNCTIONAL GENOMICS APPROACHES, AND INVESTIGATING SELF-INCOMPATIBILITY IN *MISCANTHUS*

BY

JUSTIN MICHAEL GIFFORD

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor in Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

    Professor John A. Juvik, Chair
    Assistant Professor Patrick J. Brown
    Professor Stephen P. Moose
    Assistant Professor Erik J. Sacks

ABSTRACT

It has long been recognized that relience on fossil fuels for the world's energy needs is unsustainable, yet few attractive alternatives have been developed. Establishing an alternative energy source for the transportation sector has been particularly challenging as the vast majority of the current alternative energy technologies result in production of electricity as opposed to liquid fuel. The one major exception to this is biofuels. While corn ethanol has grown rapidly over the last decade, the dedication of such a high percentage of corn production to fuel has put a strain on food prices. For this reason, dedicated lignocellulosic crops and the technology to efficiently produce ethanol from lignocellulosic feedstocks have moved into focus. *Miscanthus*, a genus of perennial grasses that can produce high yields in temperate environments, has been identified as a potential substrate for ethanol production.

Until recently, most of the effort put into *Miscanthus* breeding over the last century has focused on developing it as an ornamental crop. As breeding goals shift to increasing yield and adapting *Miscanthus* to local environments, breeders are in need of information regarding the genetic architecture of the traits integral to yield and adaptability. To address this, a four year quantitative trait locus (QTL) study was conducted on a full-sib, F_1 mapping population of *M. sinensis* segregating for flowering time, height, leaf width, and yield using a genetic map consisting of 846 segregating SNP and SSR markers. In total, 78 QTLs with LOD scores above the genome-wide, permuted threshold equivalent to a P-value of 0.05 were identified across 13 traits. Forty of the 78 QTLs were detected in multiple years, and the power to detect QTLs appeared to peak in the third year of growth. Both the use of spring emergence and vigor rating as a covariate to account for variation related to differences in establishment increased the power

to detect QTLs in the two year establishment period. Finally, a dry period in the middle of the 2012 growing season suggested that yield declines may be due to a decrease in tiller diameter.

Despite the rapid expansion of genetic information pertaining to the *Miscanthus* genus over the last decade, a rapid, highly efficient functional genomics tool remains elusive. Limited success has been achieved with traditional bombardment and *Agrobacterium*-mediated transformation, but the low efficiencies of these protocols in combination with the genotype specificity and lengthy time requirements make these techniques imperfect. Virus induced gene silencing (VIGS) may present an ideal alternative. This technique is rapid and robust to polyploidy, but it is untested in *Miscanthus*. Here we attempted to adapt an existing VIGS vector with a large monocotyledonous host range for use in *Miscanthus*. *Barley stripe mosaic virus* (BSMV) has been reported to infect over 240 species within the Poaceae family. This vector was previously modified to include a ligation independent cloning (LIC) site to aid in cloning small plant gene fragments into the virus. In theory, during viral replication, double-stranded RNA will trigger the post-transcriptional gene silencing response, resulting in down regulation of viral transcripts and the cloned plant gene. In order for this process to work, the virus must be able to infect *Miscanthus*. Five *Miscanthus*, two maize, and two sorghum genotypes were screened for susceptibility to BSMV to no prevail. Simultaneously, a number of gene fragments were attempted to clone into the LIC site. After extensive trouble shooting, it was determined that a deletion in the original LIC BSMV clone was preventing successful integration of gene fragments. Two strategies are weighed moving forward: test additional genotypes of *Miscanthus* or acquire one or more additional VIGS vectors to screen *Miscanthus* for susceptibility. Neither method is guaranteed to work. The decision comes down to the amount of resources and effort available to dedicate towards developing a VIGS system for *Miscanthus*. Screening additional

genotypes with the current BSMV VIGS vector will require the fewest resources but may have a smaller chance of success.

Beyond the need for a functional genomics tool, the adoption of *Miscanthus* as a dedicated bioenergy crop is impeded by, among other things, the necessity to establish fields via clonal propagation. Plants grown from seed segregate widely, resulting in a decrease in plot yield. Further, seed established plots increase the invasive potential of this non-native genus of grasses. Both of these concerns could be alleviated by developing hybrid seed from inbred lines. In order to accomplish this goal, the strong self-incompatibility (SI) mechanism acting in *Miscanthus* would need to be manipulated. To better understand this mechanism, SI relationships between full siblings of two biparental populations were utilized to map the locus/loci responsible for self-recognition. A single locus was found to be segregating in both mapping populations, and conserved synteny among the grasses does not suggest that this locus corresponds to the S- or Z-loci found to operate in self-recognition of the Pooideae subfamily of Poaceae. Using rice and sorghum as bridge species, this region appears to be homologous to the T-locus, previously believed to only have two alleles: a functional allele and a non-functional allele. Increasing evidence suggests that sorghum has a deletion in the region homologous to the mapped SI locus in *Miscanthus*, explaining its lack of SI. Despite only detecting a single locus in each mapping population, an investigation of pollen tube growth in a more diverse set of crosses revealed two separate crosses in which 75% of the pollen was compatible. This can only occur in a two locus SI system, suggesting that at least one more locus is responsible for SI in *Miscanthus*. In light of these new revelations, the model in which the common ancestor of all grasses had a four locus SI system is supported. After the split of the Pooideae and Panicoideae subfamilies, these four SI loci underwent distinct evolutionary pathways, resulting in two related but distinct SI systems.

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor Dr. Jack Juvik for his professional and personal guidance throughout my years in his lab. Your patience and passion for molding young scientists is unrivaled. Thank you Dr. Darin Eastburn for the opportunity to TA for PLPA 200. I think there would be a much deeper appreciation for teachers and all of the hard work that it takes to develop and teach a meaningful class if every student tried their hand on the other side of the classroom. Additionally, I would like to extend my thanks to my graduate committee members Dr. Patrick Brown, Dr. Stephen Moose, and Dr. Erik Sacks for their consideration and helpful feedback. I am also grateful to my colleagues Kang Mo Ku, Talon Becker, Alicia Gardner, Michael Paulsmeyer, and Laura Chatham for their constructive criticism along the way as well as the occasional shared suffering. Avery Shikanai deserves special thanks for his hard work, drive to learn, and creativity in association with the virus induced gene silencing work. My sincerest apologies go out to those individuals that I have failed to name directly. That is one of the great things about the University of Illinois—you have so much support throughout your tenure that you cannot possibly thank everybody by name.

Of course, I would be remiss to forget those individuals that came before U of I or those individuals that were instrumental to my life outside of school. First and foremost, I need to thank my beautiful and wonderfully understanding wife Heather. I understand that it is not easy to be married to somebody in graduate school, but you certainly made it look easy! There is no motivation like knowing what waits beyond the finish line. I would have never made it here without my family. I benefited from arguably the best parents a person could ask for. I know I did not always make it easy, but somebody had to put your parenting skills to the test! Christine, I could not have asked for a better big sister and role model. Without you paving the way, I

v

certainly would not have even attempted half of what I have accomplished already. Dylan, regardless of how many times you ate couch, you have always been the stronger more courageous of the two of us. Lastly, Alex, I know we sometimes acted like it was a pain having you tag along, but it never was. Do not let anyone set limits on your achievements, and do not underestimate what a powerful tool hard work is. Lastly, special thanks go out to my friends. Some friends come and go as time passes, but I have been fortunate to have so many stick around and remain such a positive influence in my life. For everyone who has helped me along the way, thank you. I am truly blessed.

TABLE OF CONTENTS

CHAPTER 1: LITERATURE REVIEW

1.1 Energy outlook

Despite concerted efforts to move towards energy efficiency, the global energy demand is predicted to increase by approximately 33% from 2011 to 2035 as a result of both population growth and increases in the standard of living (Figure 1.1). The overwhelming majority of this increase in energy demand is expected to come from developing countries, particularly those in Asia. While fossil fuels will undoubtedly continue to provide the bulk of our energy needs over the coming decades, it is expected that their contribution will decrease from 82% to 76% of the total worldwide energy consumption by 2035 (WEO Factsheet, 2013). Renewable energy sources are expected to continue to increase in importance in spite of the heavy subsidies needed to make them competitive today; in 2012, 101 billion dollars was spent on renewable fuel subsidies worldwide. The necessary advances in renewable energy development do not come without a considerable cost to the world economy either as an estimated 6.5 trillion dollars worth of investment in renewable energy technology will be necessary to meet the International Energy Agency's 2035 predictions (WEO: Renewable Energy Outlook, 2013).

One of the major driving forces behind the push for renewable fuels is the eminent threat of global climate change. Vast quantities of carbon dioxide are released from burning fossil fuels, resulting in an increased concentration of atmospheric carbon. Carbon dioxide, along with other greenhouse gases, prevents longwave radiation from escaping Earth's atmosphere. Ultimately, this can lead to an increase in global mean temperatures if more energy is absorbed from the sun than is dissipated. In the late 1950s the concentration of carbon dioxide in the atmosphere was just below 320 parts per million (ppm) while in 2014 the level had risen to nearly 400 ppm (Tans & Keeling, 2014). Carbon dioxide levels preceding these dates can be calculated from ice cores

collected in Antarctica. It has been shown that there is a correlation between carbon dioxide levels and temperatures during past deglaciation events (Shakun *et al*., 2012). While it is difficult to predict the exact effects of global climate change, it is believed that agriculture will suffer due to increases in the frequency and severity of droughts and pests (Stireman *et al*., 2005; Dai, 2010; Thomson *et al*., 2010). It is believed that we can avoid or limit the negative effects caused by global warming if we can switch to carbon neutral fuels, i.e. fuels that do not result in a net increase in atmospheric carbon dioxide. Further, it can be argued that the price of fossil fuels does not represent the true cost of their use as the environmental pollution they produce is not factored into their cost.

An additional reason to move away from fossil fuels is the fact that there is a finite amount of extractable fossil fuels available. Fossil fuels form over millions of years from the decomposition of prehistoric organic life buried deep underground. Different fossil fuels are formed based on the nature of the organic matter, the length of time that it has been buried, and the heat and pressure it has been exposed to (How fossil fuels were formed, 2013). It is estimated that coal, natural gas, and crude oil reserves will be depleted in 109, 56, and 53 years, respectively, given the current extraction rates (BP, 2013). As scarcity increases the price of these fuel sources, alternative energy will become more and more competitive. Proposed increases in electricity prices in Armenia and South Africa have recently sparked widespread protesting and rioting (Demirjian, 2015; Nicolaides, 2015). The degree to which this price increase is related to scarcity as opposed to politics is difficult to parse out. Regardless, scenes such as these will likely continue to increase in frequency. Advances in fossil fuel extraction technology can delay but cannot prevent this inevitable fate.

Lastly, the uneven distribution of fossil fuels leads to the need for energy security, i.e. uninterrupted access to affordable energy. This is an issue for countries that import much of their energy. In the U.S. crude oil presents the largest concern regarding energy security as we consume 19.8% of the total worldwide consumption yet produce less than half of the crude oil we consume and hold less than 3% of the world's proven reserves (BP, 2013). The U.S. is not the only country that has an unbalanced energy portfolio. To combat the fact that many countries do not control the production of the majority of their petroleum needs, the International Energy Agency requires member countries to store the equivalent of 90 days worth of net oil imports (Energy Act of 2000, 2000); in theory, this should buffer member countries from short term disruptions to the economy, but it does not provide long term energy security.

Advances need to be made in all carbon neutral energy fields, including wind, solar, geothermal, nuclear, water, and biomass derived energy in order to meet our growing energy demands while avoiding environmental pollution. The vast majority of the imported crude oil in the United States is used to satisfy our transportation needs (BP, 2013), so much of the policy in the United States has focused on energy from biomass due to its potential as a drop-in liquid fuel for our transportation system. The initial renewable fuel standard required blending 7.5 billion gallons of renewable fuel with gasoline by 2012 (Energy Policy Act of 2005, 2005). This was modified two years later by the Energy Independence and Security Act of 2007. The key modifications included raising the required amount of renewable fuel to 36 billion gallons by 2022, creating new categories of renewable fuels, and setting volume goals for each (Figure 1.2), and ensuring that the blended fuels reduced carbon emissions over the entire lifecycle of the renewable fuel (Energy Independence and Security Act of 2007, 2007).

Despite these government mandates, many of the goals set forth by the Energy Independence and Security Act of 2007 have not been realized (Annual Energy Outlook 2013, 2013). In particular, cellulosic ethanol production has lagged behind. 20,000 gallons of cellulosic ethanol were produced in 2012, accounting for a paltry 0.004% of the 500 million gallon goal according to information provided by the United States Energy Information Administration (US EIA). The US EIA suggests three main reasons for this discrepancy: difficulties involved with start-up companies obtaining financing, difficulties scaling up technology, and corporate shifts in policy due to the increased availability of low-cost natural gas ("Cellulosic biofuels," 2013).

In addition to these concerns, cellulosic ethanol faces the same challenges that corn ethanol faces, e.g. the ability of the transportation sector to consume higher volumes of ethanol. Nearly all gasoline in the U.S. already contains 10% ethanol, and not all cars are capable of running on higher blends. To account for these limitations, the EPA recently decreased the required volumes for both advanced biofuels and total renewable fuels. The EPA used the actual volume of renewable fuels consumed in 2014 to set the goals for 2015 and 2016. Thirty-three million gallons of cellulosic biofuels were incorporated into the transportation sector in 2014, accounting for 0.019% of the total fuel used by the transportation sector. The modified goals would see these numbers reach 206 million gallons (Figure 1.2), 0.114% of the total fuel used by the transportation sector, by 2016 (EPA, 2015).

As more flex fuel cars, cars that can run on either an 85% ethanol blend or the regular 10% ethanol blend, are produced, it is expected that the capacity to consume ethanol in the United States will increase. Additionally, as oil prices rise, ethanol is expected to become more economically competitive. Combined with the fact that corn ethanol is capped at 15 billion gallons to prevent spikes in food prices (Energy Independence and Security Act of 2007, 2007),

there is still reason for optimism that cellulosic ethanol can make a major contribution to the

U.S.'s energy portfolio. KiOR Inc. became the first company to open a cellulosic ethanol plant in

the U.S. in the fall of 2012 followed shortly by INEOS Bio in 2013, and at least five additional

facilities were expected to come online by the end of 2014, although many of these facilities

have missed their original opening date. Unfortunately, many of these pioneering plants have run

into trouble. KiOR Inc. has since filed for chapter 11 bankruptcy, and INEOS Bio's Vero Beach,

FL facility has spent the majority of its time off-line. Issues with new technologies were expected,

and the increase in cellulosic ethanol production from 20,000 gallons in 2012 to 33 million

gallons in 2014 is reason enough for optimism.

1.2 Lignocellulosic ethanol

The United States current ethanol industry is almost exclusively based on the

fermentation of sugars from starch found in corn grain. The most common concern cited with

this process is the competition it creates with food, i.e. the food vs. fuel debate. Although much

of the corn produced in the United States is not directly consumed by humans, corn is one of the

main components of livestock feed. For this reason, the amount of corn ethanol that can be

produced in a given year is 15 billion gallons; corn ethanol production has surpassed 13 billion

gallons each year since 2010, reaching a maximum of 13.9 billion gallons in 2011 (Annual

Energy Outlook 2013, 2013). According to the United States Department of Agriculture (USDA),

45.3% of the total domestic corn produced in 2013 was used in livestock feed while 42.7% was

used for ethanol production (Figure 1.3). For comparison, these numbers were 69.4% and 14.0%,

respectively, just ten years prior. With the increased demand for corn resulting from ethanol

production, we also see a rise in the average price of corn from $2.42 per bushel to $4.60 per

bushel in 2003 and 2013, respectively ("Corn Background," 2014). While many other factors

affect the price of corn, it is apparent that ethanol production is putting a strain on the price of corn, resulting in higher input costs for livestock and the ethanol industry alike.

In order to avoid competition between food and fuel, alternative substrates for ethanol production should be utilized. In fact, the renewable fuel standard specifies that 16 billion gallons of biofuel originate from cellulosic sources by 2022. The law defines cellulosic biofuel as any fuel derived from cellulose, hemicellulose, or lignin—the main structural components of plant cell walls—produced from a renewable source. It further stipulates that cellulosic biofuels must reduce greenhouse gas emissions by 60% as compared to the baseline greenhouse gas emissions established from fossil fuels (Energy Independence and Security Act of 2007, 2007). Theoretically, cellulosic ethanol provides many of the benefits of corn ethanol but does not come with the same competition with food as it can be produced from a variety of non-food plant products.

Cellulose is a straight-chain, homopolymer polysaccharide consisting of D-glucose molecules connected by β-1,4 glycosidic bonds. The glucose chains aggregate to form cellulosic microfibrils through hydrogen bonding. The chains of glucose can be up to 10,000 units in length, so the length and shape of the molecules lead to a large number of hydrogen bonds, resulting in high tensile strength (Klemm *et al*., 2005). As the most common organic polymer on Earth, cellulose has nearly limitless potential as a source of sugars for fermentation. In the cell wall, cellulose is closely associated with hemicellulose and lignin. The abundance of cellulose makes it a great candidate feedstock for biofuels, but the chemical nature that makes it so useful as a structural component of the plant cell wall also makes it difficult to degrade into fermentable sugars.

The major differences between readily degradable starch and cellulose stem from the fact that starch—such as the starch found in corn grain—is formed by α-1,4 glycosidic bonds instead of β-1,4 glycosidic bonds (Figure 1.4). This seemingly small difference results in rather contrasting molecular properties. Instead of straight chains, starch forms tightly coiled helices (Buleon *et al*., 1998), which increases the number of glucose-glucose bonds accessible to hydrolytic enzymes. In contrast to starch, cellulose is unbranched and much longer. Both of these characteristics reduce the number of ends available for exocellulosic hydrolysis (Zhang & Lynd, 2004), and the number of available ends is believed to be a major limiting factor to the hydrolysis of cellulose (Schulein, 2000). Cellodextrins, short glucose chains resulting from the breakdown of cellulose, are mostly insoluble at lengths greater than six glucose molecules (Miller, 1963); oligosaccharides resulting from starch breakdown, on the other hand, become soluble at much longer lengths, approximately 60 glucose molecules (John *et al*., 1982). Secondary hydrolysis of soluble cellodextrins progresses at a much quicker rate than primary hydrolysis of insoluble cellulose (Zhang & Lynd, 2004); therefore, the difference in solubility can greatly decrease the rate of cellulose hydrolysis in comparison to starch hydrolysis. Finally, starch loses it crystalline structure upon heating to 60-70$^{\rm o}$C (Atwell *et al*., 1988) while cellulose maintains its crystalline structure up to temperatures near 320$^{\rm o}$C (Deguchi *et al*., 2006). The crystalline structure affects the speed of hydrolysis indirectly through changes in surface area available for hydrolytic enzymes to bind. All of these factors increase the difficulty involved with using cellulosic feedstocks as compared to starch; the combination of the aforementioned differences between cellulose and starch result in starch hydrolysis proceeding 100 times faster than cellulose hydrolysis (Zhang & Lynd, 2004).

Enzymatic hydrolysis followed by fermentation is the preferred method of cellulosic ethanol production. The most common source of cellulase for this process is from the fungal species *Trichoderma reesei* because they produce relatively large quantities of cellulases (Juhasz *et al*., 2005). Supernatant collected from cultures of *T. reesei* contains at least two exocellulases (cellobiohydrolases 1 & 2, abbreviated CBH1 & 2), five endoglucanases (EG1-5), β-glucosidases, and hemicellulases (Vinzant *et al*., 2001). Three enzymes in particular account for the majority of the cellulasic protein; CHB1, CHB2, and EG2 constitute 60, 20, and 12 percent of the total cellulasic protein present in the supernatant (Gritzali & Brown, 1978; Goyal *et al*., 1991).

As suggested previously when comparing starch and cellulose hydrolysis, the rate of cellulase hydrolysis is dependent on a number of factors: the surface area available for enzymes to bind, the length of the polymers, the number of chain ends, the solubility of cellodextrins, and the amount and distribution of lignin. Hydrolysis proceeds more quickly when surface area increases (Chang *et al*., 1981); when the solubility of cellodextrins increase, such as at higher temperatures or shorter polymer lengths (Klemm *et al*., 1998); and when the number of chain ends increase (Schulein, 2000). Contrarily, as polymerization decreases, the rate of hydrolysis increases. Finally, less lignin leads to faster degradation of cellulose because lignin competitively absorbs cellulases (Bernardez *et al*., 1993); therefore, removing lignin or decreasing the surface area of it increases the efficiency of cellulose hydrolysis. It has also been suggested that the activity of cellulases may decrease over time (Valjamae *et al*., 1998), but the actual kinetics of the cellulases themselves are not rate limiting. Cellulases tend to have comparable catalytic rates to enzymes involved in starch hydrolysis (Zhang & Lynd, 2004).

The modification of the aforementioned attributes of cellulosic material through various pretreatment methods has been a primary focus of research. Progress has been achieved in

increasing the cost efficiency of ethanol production from cellulosic material through

pretreatments, but an additional means of increasing the cost efficiency is to decrease the cost of

the cellulasic enzyme mixtures or by tweaking their levels within the mixtures. The known gene

sequences of glycosyl hydrolases has surpassed 10,000 as of 2003 from an initial 300 in 1991

(Zhang & Lynd, 2004). It has been known for a while that adding β-glucosidases to *T. reesei*

supernatant can increase the yield of glucose from cellulose (Spindler *et al*., 1989; Schell *et al*.,

1990). Hence, cheap sources of β-glucosidases have been explored. Genetic modification of

plants allowing them to produce β-glucosidases has been achieved. Wei *et al*. (2004) achieved β-

glucosidase expression levels of up to 2.3% of the total soluble protein while Jung *et al*. (2010)

achieved 5.8% six years later. Most recently, chloroplast transformation was utilized to achieve

β-glucosidase yields in excess of 10% of the total soluble protein in planta, which is believed to

be the economic threshold (Gray *et al*., 2011). Continuing efforts in cellulose degradation

technology should increase the yield of fermentable sugars from cellulosic plant material and

decrease the production cost of cellulosic ethanol, making it more competitive every year, as

evidenced by the advances in β-glucosidase production.

1.3 Cellulosic feedstocks

The goal of the Billion-Ton Study (Perlack *et al*., 2005) was to assess the feasibility of

producing one billion tons of dry biomass—the estimated amount of biomass required to replace

30% of the U.S.'s current annual petroleum use—under different scenarios by the middle of the

century. The study concluded that under high yield scenarios, the U.S. could indeed produce this

amount of dry biomass in a sustainable manner, but under the baseline scenario, the goal would

likely not be obtained. While many assumptions went into this work, one of the most pertinent

assumptions to the work herein is that the yield of energy crops would increase by two, three, or

four percent annually as compared to the base line increase of one percent annual yield improvement. These improvements would be through gains in both breeding and agronomy, with the higher annual increases attributed mostly to breeding for increased biomass production (Perlack *et al.*, 2005).

The Billion-Ton Study received a lot of criticism for utilizing unrealistic assumptions. For example, the cost of the various feedstocks was not included in the analysis. Not restricting the potential feedstock sources by price would likely result in an overestimation of the available feedstock. An update to this study was released in 2011 to address the original study's shortcoming as well as provide information on smaller scales and over a time period that coincides with the Renewable Fuel Standard in order to make the report more useful (Downing *et al.*, 2011). The overall result of the update was similar to that of the 2005 Billion-Ton Study although available biomass from both crop residue and forest residue was lower due to soil carbon management and a decrease in paper production. These decreases in estimated biomass were offset by an increase in the biomass coming from energy crops, producing as much as 800 million dry tons a year by 2030 under the assumption of an annual increase in yield of four percent and a price of \$60 ton^{-1} (Downing *et al.*, 2011). While Downing and colleagues (2011) did note that the major driver of these estimates was price, the rate of yield increases also played a major role in determining the tons of dry biomass available from energy crops.

Regardless of the predicted scenario, it is clear that dedicated energy crops will play a pivotal role in reaching the goals set forth by the Renewable Fuel Standards. Further, no single energy crop will be viable across the entire spectrum of lands dedicated to energy production. Due to size alone, the U.S. presents too many growing environments for a single species of plant to produce the maximum amount of biomass in each. When looking at the scale of a single farm,

one energy crop may not even be adapted to grow throughout the property due to differences in water drainage or soil type. It is clear that no single species will satisfy the entirety of the U.S.'s biomass requirement, but there are some general characteristics that many dedicated energy crops will share. The ideal crop would be high yielding even under low nutrient and water regimes. It would be a perennial species with high genetic diversity, likely utilize the C_4 photosynthetic pathway, be native or noninvasive to the region it is grown in yet be cheap to establish and maintain, and be environmentally beneficial.

Many different plant species have been proposed as candidate bioenergy crops in the U.S. These include many herbaceous grasses such as *Panicum virgatum* (switchgrass), *Miscanthus* spp., and more recently *Spartina pectinata* (prairie cordgrass) as well as woody species such as *Populus deltoides* x *Populus nigra* (hybrid poplar), *Robinia pseudoacacia* (black locust), and *Salix* spp. (willows). In general, the woody species would be more suited to cooler growing environments as the grasses do not tend to be as cold tolerant, particularly in establishment years. Although sugarcane has proven to be one of the more productive plant species worldwide and provides sustainable ethanol that supplies much of Brazil's transportation needs, it is not suitable for all but the most southern regions of the contiguous U.S. as it is not adapted to temperate environments. Each of the aforementioned species have their own benefits and drawbacks, and it would not be out of the question if all of them are exploited in order to meet the energy demands of the U.S. in a sustainable fashion. That being said, the focus of the subsequent research will be on *Miscanthus* spp. as they are particularly promising bioenergy feedstock sources.

1.4 *Miscanthus*

Miscanthus is a genus of perennial grasses native to southeastern Asia that can be found growing naturally throughout eastern Asia and the Pacific Islands (Hodkinson *et al*., 2002a;

Sacks *et al*., 2013). The genus is of particular interest as an energy crop due to its high yields and longevity, C_4 photosynthetic pathway, and its adaptation to temperate environments (Schwarz *et al*., 1994b; Clifton-Brown & Jones, 2001; Clifton-Brown *et al*., 2004; Heaton *et al*. 2004b). Among other things, *Miscanthus* has been traditionally used as thatch for roofs, forage for cattle, and fiber for paper production (Xi, 2000; Hodkinson *et al*., 2002b). It is closely related to *Saccharum officinarum* (sugarcane) and more distantly related to *Sorghum bicolor* (sorghum) and *Zea mays* (maize), all members of the Andropogoneae tribe (Clayton & Renvoize, 1986; Amalraj & Balasundaram, 2006).

The number of distinct species within the *Miscanthus* genus is debatable. Clayton and Renvoize (1986) proposed the existence of as many as 20 species, but more recent studies suggest that 11-12 species would be more accurate (Clifton-Brown *et al*., 2008). Part of the reason for the large discrepancy between these two numbers is the fact that the Himalayan and southern African species once thought to belong in the *Miscanthus* genus are no longer characterized as such (Hodkinson *et al*., 1997; Amalraj & Balasundaram, 2006). A major piece of evidence for this split comes from differences in basic chromosome number: x=5 or 10 for Himalayan accessions (Mehra & Sharma, 1975), x=15 for southern African accessions (Hodkinson *et al*., 2002a), and x=19 for *Miscanthus* species (Adati & Shiotani, 1962). Additionally, as compared to *Miscanthus*, Himalayan accessions have a higher degree of relatedness to sugarcane and southern African accessions have a higher degree of relatedness to sorghum (Hodkinson *et al*., 1997; Amalraj & Balasundaram, 2006). Despite this confusion regarding the number of species within the genus *Miscanthus*, two species, along with their interspecific hybrid, have been identified as the most promising candidates as dedicated energy crops: *M. sinensis*, *M. sacchariflorus*, and *M.* ×*giganteus*.

M. sinensis has the largest distribution of all of the species in the genus and is commonly found throughout Japan, Korea, and China, oftentimes on hillsides, open grasslands, or other well drained areas (Lee, 1993; Shouling & Renvoize, 2006). Prior to any interest in energy crops arose, it was widely grown as an ornamental in both Europe and the U.S. (Darke, 1994). As a member of the section *miscanthus*, *M. sinensis* exhibits a tufted growth habit due to its short rhizomes. Other distinguishing characteristics of this species are the lack of axillary bud growth, the pilose nature of the abaxial leaf surface, and the presence of awned spikelets (Figure 1.5; Sacks *et al*., 2013). Due to its broad latitudinal range, a great amount of flowering diversity can be seen in *M. sinensis*. The more northerly adapted genotypes tend to be day neutral while the southern adapted genotypes tend to be induced to flower under short day length, although other factors such as growing degree days most certainly play a role (Deuter, 2000; Jensen *et al*., 2011; Arnoult *et al*., 2014). The extended growing period of southern adapted genotypes often results in these genotypes obtaining greater stature as further increases in height are restricted by the terminal inflorescence. In general, genotypes within this species range in height from 0.7 to 3 meters (Lee, 1993; Renvoize, 2003; Sun *et al*., 2010) and can obtain yields in excess of 20 tons of dry matter ha[-1] (Gauder *et al*., 2012).

The increased application of genomic markers to diverse sets of germplasm within the *Miscanthus* genus is continuing to refine our understanding of the evolution of this genus. Clark *et al*. (2014) investigated a diverse set of *M. sinensis* cultivars with over 21,000 SNP markers. The six populations of *M. sinensis* identified were found to be mostly influenced by patterns of past glaciations and geographical barriers. *M. sinensis* radiated outward from SE China (Zhang *et al*., 2013; Clark *et al*., 2014). Interestingly, the genetic evidence suggests that the first migration of *M. sinensis* out of SE China was to Japan about 14,000 years ago. Subsequently, it migrated

north and west throughout China and the Korean Peninsula from SE China (Figure 1.6; Clark *et al*., 2014). A more in depth investigation of the population structure of *M. sinensis* in Japan revealed a seventh distinct population (Clark *et al*., 2015).

The status of *M. floridulus* as a species distinct from *M. sinensis* has been questioned many times (Hodkinson *et al*., 2002b; Chae *et al*., 2014; Clark *et al*., 2014). Admittedly, these authors recognized the limitations of their small *M. floridulus* sample sizes and stressed the need for additional testing. More extensive work suggests that while gene flow continues between the two species the two species split 1.59 Mya (Huang *et al*., 2014). The distinction between population structure within one species and the beginnings of a divergence of two species in this case may be blured. Morphologically Sun *et al*. (2010) cites differences in spikelet length, axis length, and number of racemes although overlap in all of these traits makes them poor distinguishers. Ultimately, the question of whether *M. sinensis* or *M. floridulus* are distinct species has little effect on their use as biofuel crops.

M. sacchariflorus is the most northerly adapted species in the *Miscanthus* genus; it can be found growing as far north as southern Siberia. As a member of the section *triarrhena*, it opposes *M. sinensis* in many of its defining characteristics. For example, it exhibits a spreading growth habit due to its long rhizomes, generally has axillary buds, has glabrous abaxial leaf surfaces, and has spikelets lacking awns (Figure 1.5; Sacks *et al*., 2013). It is also the only *Miscanthus* species that is commonly found growing in poorly drained soils (Lee, 1993). In China it is nearly exclusively diploid, yet it is often tetraploid in Japan. Due to its growth habit and affinity for riparian environments, it can be found on some invasive species lists (MA PPL, 2013; TIS, 2014). *M. sacchariflorus* is a quantitative short day plant (Jenson *et al*., 2013). It can obtain comparable heights and yield to that of *M. sinensis*, but it has a tendency to be shorter and flower earlier due

to the prevalence of northern adapted genotypes. Further, this species displays a rather strong

dormancy response. In the greenhouse, this dormancy response can be overcome without any

cold treatment by removing the above ground vegetation prior to the completion of flowering or

by repotting.

There has been some debate over the years regarding the relationship between *M. sinensis*

and the tetraploid *M. sacchariflorus* in Japan and whether the tetraploid Japanese genotypes are

true *M. sacchariflorus*. This debate has been mostly resolved. Clark *et al*. (2015) genotyped 78

M. sacchariflorus plants native to Japan with over 20,000 SNPs. On average, seven percent of

the tetraploid *M. sacchariflorus* ancestry can be traced back to *M. sinensis*. Thus, tetraploid *M.

sacchariflorus* in Japan appears to have originated via an autopolyploidy event followed by

introgression of *M. sinensis* DNA (Clark *et al*., 2015). The percentage of *M. sinensis* heritage is

variable and appears to be correlated with latitude. Clark *et al*. (2015) speculated that this may be

a result of *M. sinensis* offering a selective advantage to the warmer temperatures of southern

Japan. This gene flow between species and across ploidy levels continues to shape the evolution

of the two species in Japan. In contrast, no instances of introgression were noted in diploid *M.

sinensis* and *M. sacchariflorus* found in China. Diploid hybrids that were identified by Jiang *et al*.

(2013) had equal contributions from each parent.

The most widely studied and highest yielding genotype within the genus belongs to the

nothospecies *M. ×giganteus*, which is defined as any hybrid between *M. sinensis* and *M.

sacchariflorus*. Therefore, by definition the ploidy level of a genotype does not affect its

classification as *M. ×giganteus* despite the fact that most people incorrectly declare this species

to be a sterile triploid. Recently, Głowacka *et al*. (2014b) showed through molecular marker data

that nearly every triploid genotype, regardless of its given name, originated from a single

collection acquired in Japan in the 1930s. Of the 33 genotypes used in the study, only the plants

that were known to come from seed showed variation higher than that expected by the typical

error rate of the genotyping platform. The reason that this genotype has been so highly prized is

in part due to its high yields, its late flowering, and its broad adaptability (Clifton-Brown *et al*.

2001; Heaton *et al*. 2004b). Although the parentage and exact collection site of this particular

clone is unknown, it has been shown that two copies of its genome are from *M. sacchariflorus*

and the single, paternal dose donated by *M. sinensis* (Hodkinson *et al*. 2002b; Swaminathan *et al*.

2010).

As would be reasonable to expect based on its designation as a nothospecies, *M.*

×*giganteus* shows an intermediate growth habit (Figure 1.5), which plays a large role in its high

yields as the number of tillers per unit area and ability to spread within a field are optimized.

Because an inadequate amount of triploid genotypes have been collected or bred, it is difficult to

make broad generalizations regarding its nature, but the 'Illinois' clone does have thick stems

and rhizomes, flowers in late fall in central Illinois, if at all, and can reach heights near four

meters (Pyter *et al*., 2009). Another desirable characteristic that this genotype possesses is its

ability to shed its leaves after senescence. Although this does reduce the total dry mass produced

per unit area, this trait increases the nutrient use efficiency and decreases the ash content of the

feedstock (Beale *et al*., 1996). Presumably the selection pressure for the evolution of this trait

was provided by moving water in lowland adapted cultivars. The loss of leaves would reduce the

drag on the plant, resulting in decreased lodging (Sacks *et al*., 2013).

Diploid *M.* ×*giganteus* varieties also show vigorous growth although no clones have

been identified that out yield the widely studied triploid. Similar to the 'Illinois' clone, diploid *M.*

×*giganteus* genotypes tend to have an intermediate growth habit. Since the exact parents of the

'Illinois' clone are unknown, it is difficult to ascertain whether its superior performance is due to the ploidy level or the combination of superior parental genomes. In order to answer this question, a ploidy set would need to be generated by crossing diploid by diploid, diploid by artificiality induced tetraploid, and artificiality induced tetraploid by artificiality induced tetraploid varieties of *M. sinensis* and *M. sacchariflorus* in order to obtain adequate numbers of diploid, triploid, and tetraploid progeny to test the effect of ploidy level on yield characteristics. Further, this would have to be done for multiple sets of *M. sinensis* and *M. sacchariflorus* genotypes to make the inference space appropriate. An experiment of this magnitude requires a considerable time and resource commitment; therefore, it has yet to be initiated.

Protocols for artificial genome doubling have been optimized for all three of the *Miscanthus* species of interest (Peterson *et al*., 2002; Yu *et al*., 2009). This allows us to compare the effects of diploid vs. tetraploid genomes and triploid vs. hexaploid genomes. In an initial greenhouse experiment, both diploid and triploid progeny fared better than their otherwise genetically identical counterparts (Chae *et al*., 2013). Two points of interest can be taken from this experiment. First, while yields were lower for the polyploid individuals, certain yield characteristics, stem diameter and flowering, were improved. Second, not all of the artificially induced polyploids were identical. Thus, variation can be created via ploidy manipulation. The number of genotypes tested remains too low to confidently conclude that competitive cultivars cannot be created through ploidy manipulation, but as of now it does not appear to be the most promising avenue to pursue.

1.5 *Miscanthus* evolution

Understanding the evolutionary relationship between the model grass species and *Miscanthus* is of critical importance if one wishes to fully utilize these key resources.

Polyploidization is believed to have played a major role in the evolution of the Poaceae family. A common ancestor of all species belonging to the Poaceae family underwent a genome duplication event approximately 70 million years ago (Mya), and genetic evidence suggests that the Poaceae subfamilies did not diverge from one another until approximately 50 Mya (Paterson *et al*., 2004). Therefore, the initial diploidization events following this genomic duplication are common across all members of the grass family. For some species, such as sorghum, this was the last whole genome duplication event to have occurred in its lineage (Paterson *et al*., 2009). In many other species additional duplication events have taken place.

It is believed that sorghum and maize diverged from one another ~12 Mya (Swigonova *et al*., 2004a). By extension, this represents the divergence of maize and *Miscanthus*. Approximately one million years after this divergence maize underwent a genome duplication event (Gaut & Doebley, 1997). Sorghum and the Saccharinae clade, which contains *Miscanthus* and sugarcane among others, diverged approximately 5 Mya (Al-Janabi *et al*., 1994; Kim *et al*., 2014), although some estimates of this divergence are higher (Jannoo *et al*., 2007). Shortly after their divergence, a shared *Miscanthus-Saccharum* allopolyploid event occurred (~3.8 Mya). Thereafter, the *Miscanthus* and *Saccharum* genera diverged from each other (~3.1 Mya; Kim *et al*., 2014). A phylogeny of the aforementioned grasses including approximate times and duplication events has been provided for clarity (Figure 1.7).

Whole genome duplication has consequences on the size of the genomes of each of grass species, but it by itself cannot account for the differences in the genome sizes of rice, sorghum, maize, and *Miscanthus*. *Miscanthus* and maize have both underwent genome duplications after their split from rice and sorghum. Of the four grasses, *Miscanthus* and maize have the largest genomes. Depending on the species, the *Miscanthus* genome ranges from 2.1-2.8 Gb (Rayburn *et*

al., 2009; Chae *et al*., 2014) while the maize genome is 2.3 Gb (Rayburn *et al*., 1993). In comparison, rice has a genome size of 0.389 Gb (International Rice Genome Sequencing Project, 2005), and sorghum has a genome size of 0.76 Gb (Arumuganathan & Earle, 1991). Much of the additional genome expansion can be attributed to transposable elements. 35, 61, 85, and 95 percent of the rice (International Rice Genome Sequencing Project, 2005), sorghum (Paterson *et al*., 2009), maize (Schnable *et al*., 2009), and *Miscanthus* (Swaminathan *et al*., 2010) genomes, respectively, are repetitive elements. Chae and colleagues (2014) suggested that the difference in genome sizes of *M. sinensis* and *M. sacchariflorus* is due to differential loss of genome components. In addition to this, increased action of transposable elements in *M. sinensis* could account for some of this difference in genome size.

Despite these differences, there is extensive conserved synteny between the grasses. There are a few patterns of note. First, synteny is more highly conserved in euchromatin than heterochromatin (Bowers *et al*., 2005). Bowers *et al*. (2005) went on to suggest that the reason for this differential conservation of synteny is because rearrangements in genic regions are generally deleterious. Further, a disproportionally large amount of the genome expansion of sorghum in comparison to rice has occurred in heterochromatic regions. In contrast, Wei *et al*. (2007) found evidence that differential gene loss and insertion of retrotransposable elements was the major cause behind the three fold difference in the size of maize syntenic regions in comparison to rice. In combination with the genome duplication event in maize, this explains the six fold difference in genome sizes. It is important to note that the expansion is variable across the genome. The conserved synteny between all of the grass species has been and will continue to be a valuable tool.

1.6 Miscanthus as a biofeedstock

The perennial nature of *Miscanthus* is an advantageous trait for a biofuel crop to possess for a number of reasons. First, it reduces the input costs. Planting is a resource intense process; with a perennial like *Miscanthus*, cultivation is a onetime event. Initial estimates claimed that *Miscanthus* could remain productive for 20 years (Lewandowski *et al*., 2000). Additional long-term studies have suggested that this estimate was likely overly optimistic. Yields generally decline after a period of 5-10 years (Clifton-Brown & Jones, 2007; Angelini *et al*., 2009; Arundale *et al*., 2014a). It is worth noting though that the economic threshold for replanting would be well beyond this initial decline. Further, the perennial nature of *Miscanthus* allows it to recycle nitrogen from one growing season to the next using underground rhizomes, which increases its nitrogen use efficiency. Lastly, perenniality leads to environmental benefits compared to annual crops, e.g. increased erosion control and carbon sequestration.

In addition to its perennial nature, *Miscanthus* utilizes the C_4 pathway. The C_4 photosynthetic pathway results in many desirable attributes for energy crops in comparison to the C_3 photosynthetic pathway. C_3 plants utilize ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) in the initial fixation of CO_2 in the mesophyll cells. Issues arise with this process because Rubisco has affinity for both CO_2 and O_2, as the name suggests. Despite its 75-85 times higher affinity for CO_2 in C_3 plants (Pierce, 1998), up to 40% of the reactions can occur with O_2 under high temperatures due to the relative abundance of O_2 compared to CO_2 (Sharkey, 1988; Erhlinger *et al*., 1991). This process is known as photorespiration. Instead of producing two 3-phosphoglycerate (three carbon) molecules, the reaction with O_2 results in one 3-phosphoglycerate and one 2-phosphoglycolate (two carbon) molecule, which has to undergo a

complex series of reactions involving the chloroplast, mitochondria, and peroxisomes to be recovered (Buchanan *et al.*, 2000).

C_4 plants, on the other hand, utilize phosphoenolpyruvate (PEP) carboxylase in the initial CO_2 fixation step, which occurs in the leaf mesophyll cells. This eliminates photorespiration because PEP carboxylase cannot use oxygen as a substrate (Brown, 1978). The four carbon product, oxaloacetate, is converted to malate or aspartate, depending on the particular subgroup that the C_4 plant belongs to, and then transported into bundle sheath cells surrounding the vascular bundle. Once in the bundle sheath cell, CO_2 is cleaved from malate or aspartate, resulting in CO_2 and pyruvate. This CO_2 concentrating mechanism employed by C_4 plants results in concentrations of CO_2 that are 10-30 times higher than that found in the mesophyll cells (Kanai & Edwards, 1999). Rubisco then can fix CO_2 in the bundle sheath cells without the interference of oxygen. The process mirrors that of the C_3 pathway thereafter.

These differences in carbon fixation between C_3 and C_4 plants have far reaching consequences. While the extra steps involved in carbon fixation do require more energy (5 ATP/CO_2 for C_4 plants and 3 ATP/CO_2 for C_3 plants, theoretically), the lack of photorespiration more than makes up for the sacrificed energy (Bassham & Calvin, 1957). C_4 plants have a higher rate of maximum net photosynthesis, 15-40 mg $CO_2/dm^2/hr$ in C_3 vs. 40-80 mg $CO_2/dm^2/hr$ in C_4 plants (Hatch *et al.*, 1971; Björkman *et al.*, 1972), and a higher maximum growth rate, 0.5-2 g DW/dm^2 leaf area/day in C_3 vs. 4-5 g DW/dm^2 leaf area/day in C_4 plants (Cooper & Tainton, 1968; Cooper, 1970). Additionally, the CO_2 compensation point, i.e. the amount CO_2 necessary to exactly balance carbon assimilation and loss through respiration, for C_3 plants is significantly higher than C_4 plants: 30-70 ppm CO_2 and 0-10 ppm CO_2, respectively (Moss, 1962; Jones &

Mansfield, 1972). Finally, the level of Rubisco in the leaves of C_4 plants is as much as 60-80%

lower than plants that utilize the C_3 carbon fixation pathway (Long, 1999).

One consequence of these differences is that plants utilizing the C_4 photosynthetic

pathway have a tendency to have higher nitrogen use efficiencies. This has been seen in many

comparisons over the years (Brown, 1978; Wong, 1979; Schmitt & Edwards, 1981; Osmond *et

al*., 1982; Brown & Wilson, 1983). One of the hypotheses put forth to explain this phenomenon

is that C_4 plants have a higher nitrogen use efficiency due to the differences in the content of

Rubisco in the leaf. As mentioned, Rubisco is much more prevalent in C_3 species, accounting for

nearly half of the soluble leaf protein (Brown, 1978). Further, PEP carboxylase is a smaller

protein than Rubisco: 440 kDa vs. 536 kDa (Vidal & Chollet, 1997; Gerhardt *et al*., 1999).

Because nitrogen is an integral part of the amino acids that constitute plant proteins and C_4

species have a reduced amount of Rubisco due to their CO_2 concentrating mechanism, less

nitrogen is needed for the same amount of photosynthetic output. Supporting this theory is the

fact that leaf nitrogen content is often lower in C_4 species (Sage & Pearcy, 1987).

In addition to being a C_4 genus, *Miscanthus* also benefits from being a perennial in terms

of nitrogen utilization. Upon fall senescence, cold adapted genotypes remobilize the nitrogen in

their leaves to underground rhizomes (Beale & Long, 1997). Lewandowski and Schmidt (2006)

concluded that *M.* ×*giganteus* had the highest nitrogen use efficiency when compared to triticale

and reed canary grass. Initial reports suggested that the yield of *Miscanthus* was not dependant

on nitrogen fertilizer although it was recognized that due to the high yields some fertilization

would eventually be required (Schwarz *et al*., 1994a; Himken *et al*., 1997); it was even suggested

that nitrogen fixing bacteria associated with *Miscanthus* were responsible for the lack of

response to nitrogen fertilization in some environments (Eckert *et al*., 2001). It is now commonly

accepted that some low level of nitrogen fertilization will be necessary to maintain high yields, but soil factors and the stage of development interact to make broad fertilization recommendations inappropriate (Lewandowski *et al*., 2000; Miguez *et al*., 2008; Arundale *et al*., 2014b; Shield *et al*., 2014). Regardless of the soil type, nitrogen fertilization requirements for *Miscanthus* will be well below the levels necessary to produce food crops.

A second general benefit that C_4 plants hold over their counterparts is higher water use efficiencies. This is due to the fact that PEP carboxylase has no affinity for O_2 and, therefore, a lower CO_2 compensation point. This allows the plant to keep its stomata closed for longer periods of time as a depletion of CO_2 in the internal leaf air spaces does not significantly reduce photosynthesis. As CO_2 is assimilated, the concentration gradient between air and leaf CO_2 increases, so once stomata do open, CO_2 rushes into the leaf more quickly (Long, 1983). The end effect is reduced water loss via transpiration. Osmond *et al*. (1982) provided evidence for the CO_2 concentration gradient differences between C_3 and C_4 plants. Their experiments showed that C_3 legumes lost twice as much water through transpiration per CO_2 molecule up-taken compared to C_4 grasses. Further, Downes (1969) compared the water use efficiency of several C_3 and C_4 grasses and found that C_4 grasses were up to ten times more efficient in their water use than C_3 grasses, but the cooler the temperatures the smaller the gap.

As expected from its C_4 nature, *Miscanthus* has relatively high water use efficiency values. Beale *et al*. (1999) showed that *M. ×giganteus* actually has higher water use efficiency than maize when environmental corrections were made. A similar conclusion was reached when calculating water use efficiency through simulations of *M. ×giganteus* and maize in the Midwest (VanLoocke *et al*., 2012). In a controlled experiment involving *M. ×giganteus*, *M. sinensis*, and *M. sacchariflorus*, it was found that the three species did not differ significantly in their water

use efficiency when the whole plant was considered, but there was a significant genotype response when the partitioning of dry matter was considered (Clifton-Brown & Lewandowski, 2000b). *M. sacchariflorus* partitioned more resources into the stem compared to the other two species, resulting in significantly higher water use efficiency on a harvestable yield basis than the other two species. Considering the small number of genotypes in the study and the differences between greenhouse drought and field drought, it would be prudent to take caution when extrapolating the results of this greenhouse study.

All the advantages that C_4 species hold are magnified at warmer temperatures and high light intensities (Long, 1999); therefore, many C_4 species are adapted to tropical environments and are particularly prevalent in recently disturbed areas (Sage & Monson, 1999). In many C_4 species there is an irreversible loss of photosynthetic capacity at low temperatures that is proportional to the light intensity throughout the chilling event. It was also observed that C_4 species adapted to cooler environments often avoided chilling by shortening their growing season as opposed to having a mechanism to reduce photoinhibition under cool temperatures (Long, 1983). This reduction in the growing season would lead to a decrease in yield due to the decrease in light interception. Although C_4 species have a higher light saturation point, they also have a higher light compensation point (Forbes & Watson, 1992). Since energy crops will be planted in full sunlight, this should not present any disadvantages. While the advantages of C_4 species are numerous and desirable for biomass feedstock crops, the irreversible loss of photosynthetic capacity under cool temperatures presents a major limitation to C_4 species in temperate environments.

Unlike many C_4 species, *Miscanthus* has the ability to maintain high photosynthetic rates at temperatures below $12^{o}C$ (Beale *et al*., 1996; Naidu *et al*., 2003). Naidu and her colleagues

(2003) directly compared maize adapted to the Midwestern U.S. and *M. ×giganteus*. To do this, they measured the photosynthetic capacity of plants grown in cool environments (14°C/11°C day/night temperatures) and plants grown in an optimal environment (25°C/20°C day/night temperatures). The photosynthetic capacity of these plants was compared across temperatures ranging from 5-38°C. While there were no significant differences between *M. ×giganteus* raised under contrasting temperature regimes, maize plants raised under the cooler regime had an 80% decrease in photosynthetic capacity as compared to those maize plants raised under optimal conditions. This difference in response to chilling was attributed to the stability of the Rubisco large subunit and pyruvate orthophosphate dikinase (PPDK), the enzyme responsible for converting pyruvate into phosphoenolpyruvate. Recent work comparing the chilling tolerance of *M. sinensis* and *M. sacchariflorus* suggested that *M. sacchariflorus* had better chilling tolerance of the two species. Further, by comparing multiple genotypes of triploid *M. ×giganteus,* Głowacka *et al*. (2014a) were able to conclude that the superior chilling tolerance often cited in regards to *M. ×giganteus* 'Illinois' is not simply the result of interspecific hybridization or ploidy level. This fact becomes critical when attempting to breed improved *Miscanthus* cultivars. Subsequently, Głowacka and colleagues (2015) were able to identify *M. sacchariflorus* lines with even greater chilling tolerance. While the photosynthetic efficiency between maize and *Miscanthus* is minimal, the increased ability to tolerate cool temperatures leads to longer leaf duration, resulting in higher productivity (Dohleman & Long, 2009). In summary, temperate adapted *Miscanthus* holds the many advantages attributed to C_4 crops, yet photosynthesis is not inhibited at cool temperatures.

1.7 *Miscanthus* breeding

Regardless of the ideal ploidy level, it appears that hybrids between *M. sinensis* or *M. floridulus* and *M. sacchariflorus* provide a growth habit that may optimize the number of stems per unit area for the maximum yield. In addition, *M. ×giganteus* progeny also appear to benefit from interspecific hybrid vigor. This information, in combination with the fact that sterile triploids circumvent concerns regarding invasiveness, leads to the realization that the immediate future of *Miscanthus* breeding may revolve around producing additional genotypes of triploid *M. ×giganteus*. As previously mentioned, there is a lack of diversity within the triploid *M. ×giganteus* germplasm available to farmers. Recently, Aloterra Energy (http://aloterraenergy.com/) planted 18,000 acres of one genotype of *M. ×giganteus* through the USDA's Biomass Crop Assistance Program across four sites: two in Missouri, one in Arkansas, and another in Ohio. Disregarding the fact that these sites constitute rather different growing environments, planting a single genotype introduces the issue of uniform disease and pest susceptibility.

One potential approach to increasing the diversity of triploid *M. ×giganteus* would be to find existing triploid plants or seeds in areas of sympatric populations. Nishiwaki *et al*. (2011) found three triploid *M. ×giganteus* progeny from seed collected on tetraploid *M. sacchariflorus*. Unfortunately, this approach was not recommended due to the extensive amount of time and effort necessary to collect and confirm the limited number of triploid hybrids. A more practical approach to creating diversity would be to cross tetraploid and diploid individuals in a controlled fashion. Up until recently, this approach was limited in the United States due to the limited availability of *M. sacchariflorus* genotypes. While a few diploid genotypes of *M. sacchariflorus* were available through commercial nurseries, subsequent genotyping showed no diversity within

these genotypes (Erik Sacks, personal communication). By chance, one mislabeled tetraploid *M. sacchariflorus* cultivar was identified in the University of Illinois' collection (Chae, 2012). Initial crossing efforts with this plant yielded >100 small seeds with aborted endosperms similar to those described by Martinez-Reyna and Vogel (2002), but only five seeds germinated and only four individuals survived from these seeds (Chae *et al*., 2013). These genotypes, along with the triploids collected from Japan, are currently being tested in small scale plots in five locations around the U.S. Additional collection trips throughout Southeastern Asia by Erik Sacks have since provided 67 additional tetraploid *M. sacchariflorus* genotypes that can be utilized in crosses at the University of Illinois. Subsequent crossing has in fact contributed more than 30 triploid progenies that were planted in small scale field trials in the spring of 2014 (Clark *et al*., 2015). Clearly, controlled crosses are the more efficient means to create diversity within triploid *M. ×giganteus*.

In addition to collection trips, breeding within *M. sacchariflorus* and *M. sinensis* must be carried out to create improved parents. Three traits of particular importance for regional adaptation will be flowering time, cold tolerance, and cold hardiness. As discussed previously, *Miscanthus* is able to out yield other productive C_4 grasses by extending the growing season. Adapting flowering time to latitudinal regions will extend the vegetative growing season as long as possible while still allowing for senescence prior to the first hard frost. This will enable the maximum growth period while maintaining nutrient translocation to the rhizomes. Lines with higher levels of cold tolerance will allow for earlier spring regrowth while avoiding damage from late frosts. Additionally, cold hardiness should be increased in order to reduce winter kill that can be particularly problematic after the establishment season.

Traits of significance regardless of locale will be height, stem diameter, and the number of stems per unit area as these are the most highly correlated with yield on a per plot basis (Clifton-Brown & Lewandowski, 2002; Jeżowski, 2008; Gauder *et al*., 2012). On the field scale, the number of stems per unit area was most closely associated with *M.* ×*giganteus* yield (Lesur-Dumoulin *et al*., 2015). Lesur-Dumoulin and colleagues (2015) noted a 20% discrepancy between yield measured on small plots and commercial fields on average. Understanding how the basic traits measured on individual plants and small plots relates to yield at field scales will be crucial to accelerating breeding progress. This discrepancy between plot and field yields may be in part due to patchiness within fields. Zimmermann *et al*. (2014) observed six commercial *M.* ×*giganteus* fields and found that growth was limited to 87% of the field area on average. Understanding the causes underlying patchiness and addressing it either through breeding or agronomy will also play a key role in maximizing yield and breeding progress.

As has been noted in numerous papers, yield tends to peak in late summer and decline as leaves are lost throughout the winter months (Clifton-Brown *et al*., 2001). Although we do not tend to think in terms of harvest index when the harvestable portion is the entire above ground biomass, it may be useful to think in these terms as stems have higher energy density and lower ash content than leaves (Jørgensen, 1997). Therefore, putting the maximum amount of biomass into culms should reduce the decline in yield from late summer to winter. To accomplish this alteration of assimilate partitioning, the leaf area ratio, i.e. the ratio of leaf area to plant weight, of genotypes may be useful for selection. Further, the concept of harvest index can be applied to the composition of *Miscanthus* genotypes. Zhang *et al*. (2012) provided evidence that both theoretical and actual sugar yields were related to the glucan and xylan content of the feedstock and that *Miscanthus* shows a wide range for these two traits. Of the 80 diverse genotypes tested,

glucan concentration ranged from 28.7% to 46.4% and xylan concentration ranged from 19.6% to 27.1% while total glucan and xylan concentration in any one genotype ranged from 49.6% to 72.0%. Despite the fact that breeding for biological yield is quite different than breeding more traditional grain crops, the same principles can be applied in the breeding process.

Another trait that has been fundamental in increasing the yield of a traditional grain crop (maize) is leaf angle as this affects canopy photosynthesis. Higher leaf angles, i.e. more erect leaves, lead to the dispersal of sunlight over a larger number of leaves. This increases the efficiency of canopy photosynthesis as leaves tend to photosynthesize more efficiently at low light levels, and it also leads to a higher optimum number of stems per unit area through changes in leaf area index (Forbes & Watson, 1992). This was accomplished by increasing planting density in maize, but since *Miscanthus* is a perennial, rhizotomous species, this can be accomplished by altering the length of the rhizome. Along the same lines, erectness of stems may be important in terms of providing the proper leaf angle. Stem erectness likely correlates with rhizome length too as short rhizomes necessitate stems to grow outwards prior to growing upwards.

A theoretical discussion of plant breeding always proves to be much easier than the implementation itself. Regardless of the funding source, breeding programs are always financially restricted. This necessitates compromise between what a breeder would like to do and what they can do with the allotted budget. Therefore, plant breeders themselves need to think in terms of monetary efficiency when organizing their programs. Along the same lines, basic crop research should be, in part, focused around improving the efficiency of breeding. This often can come in the form of providing new tools or methods that increase the speed or decrease the cost of breeding.

One of the major limitations in breeding *Miscanthus* as a biomass feedstock for renewable energy production is the length of the breeding cycle. Although *Miscanthus* sexually matures in the first growing season, the true potential of a genotype is often not revealed until after the third year. This issue is further discussed and partially addressed in chapter two. Another limitation in *Miscanthus* breeding is the lack of a consistent and efficient means to genetically engineer plants. While not addressed in the traditional sense of genetic engineering, chapter three introduces a tool that would allow breeders to rapidly and transiently alter gene expression. Finally, due to the strong self-incompatibility mechanism functioning in *Miscanthus*, large segregating populations must be screened each breeding cycle to identify genotypes possessing the desired traits. Also, hybrid vigor cannot be maximized without being able to produce inbred lines. The genetics underlying self-incompatibility in *Miscanthus* is investigated in the final chapter.

1.8 Figures

Figure 1.1 Global energy consumption from 1988 to 2013 broken down by source (A) and the world population growth trends over the same period of time (B).
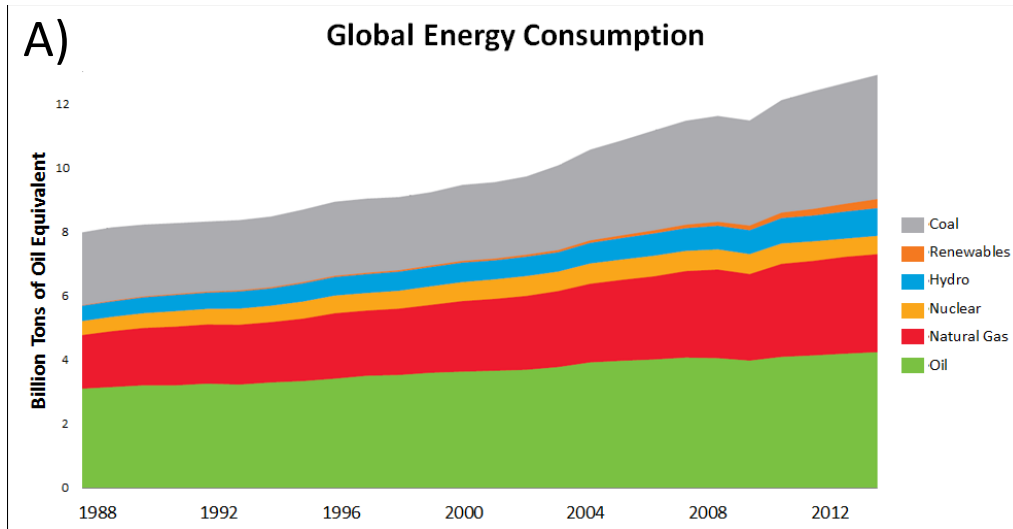
*Source: Slightly modified from the BP Statistical Review of World Energy, 2014
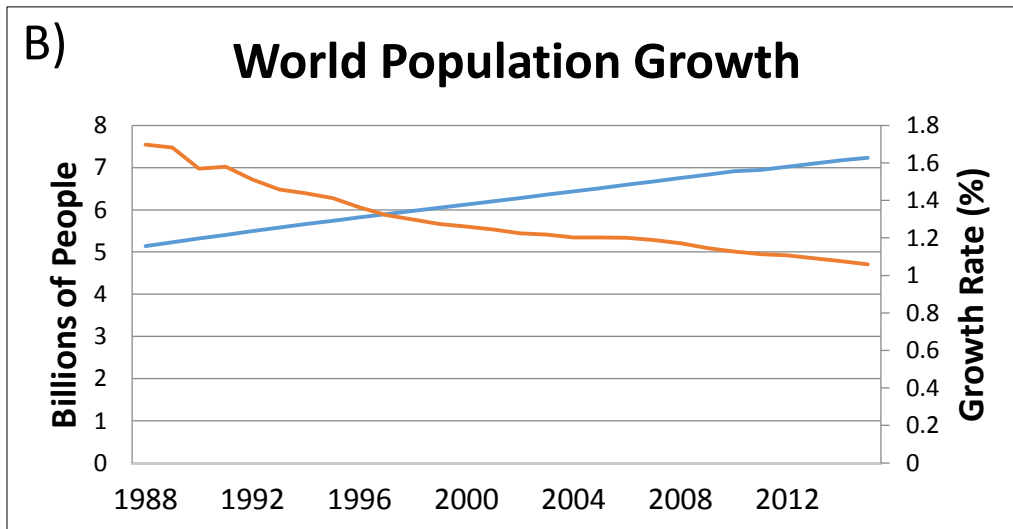
*Source: United States Census Bureau, 2015

Figure 1.2 The government mandated levels of cellulosic ethanol production as set forth by the Energy Independence and Security Act of 2007 (blue) alongside the recently modified goals set by the EPA for 2014-2016 (red).
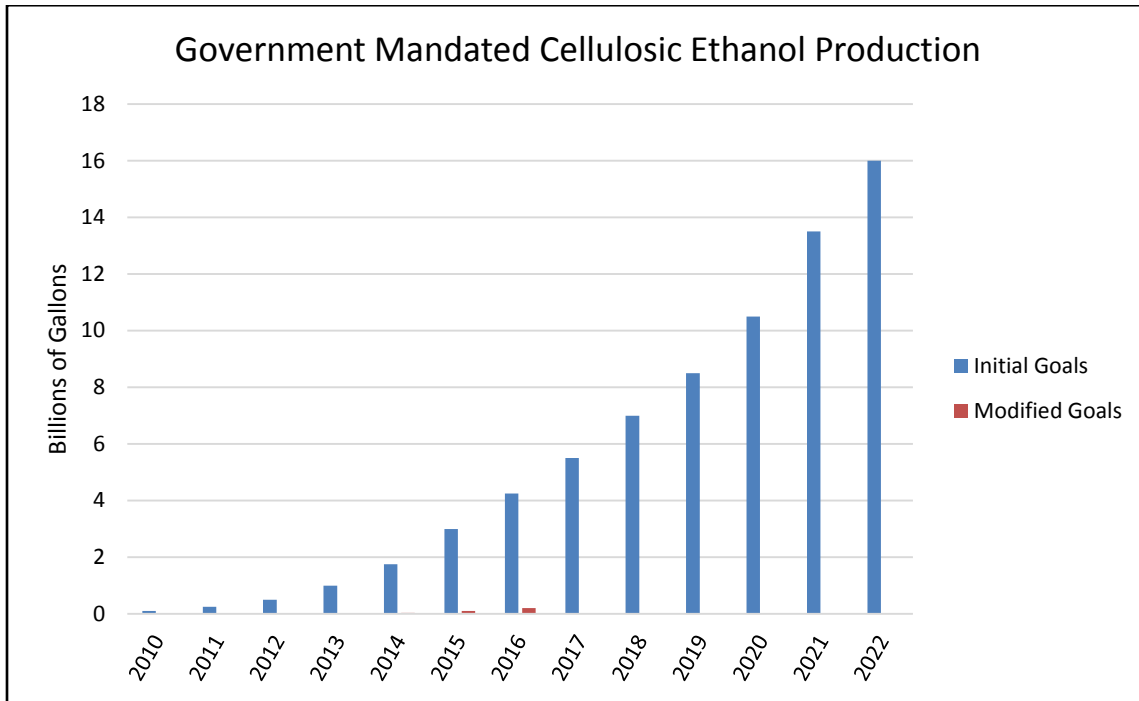
Figure 1.3 The fraction of the total corn harvest that went towards feed (orange) and ethanol production[1] (blue). The secondary axis shows the average price per bushel of corn in US dollars[2] (grey).
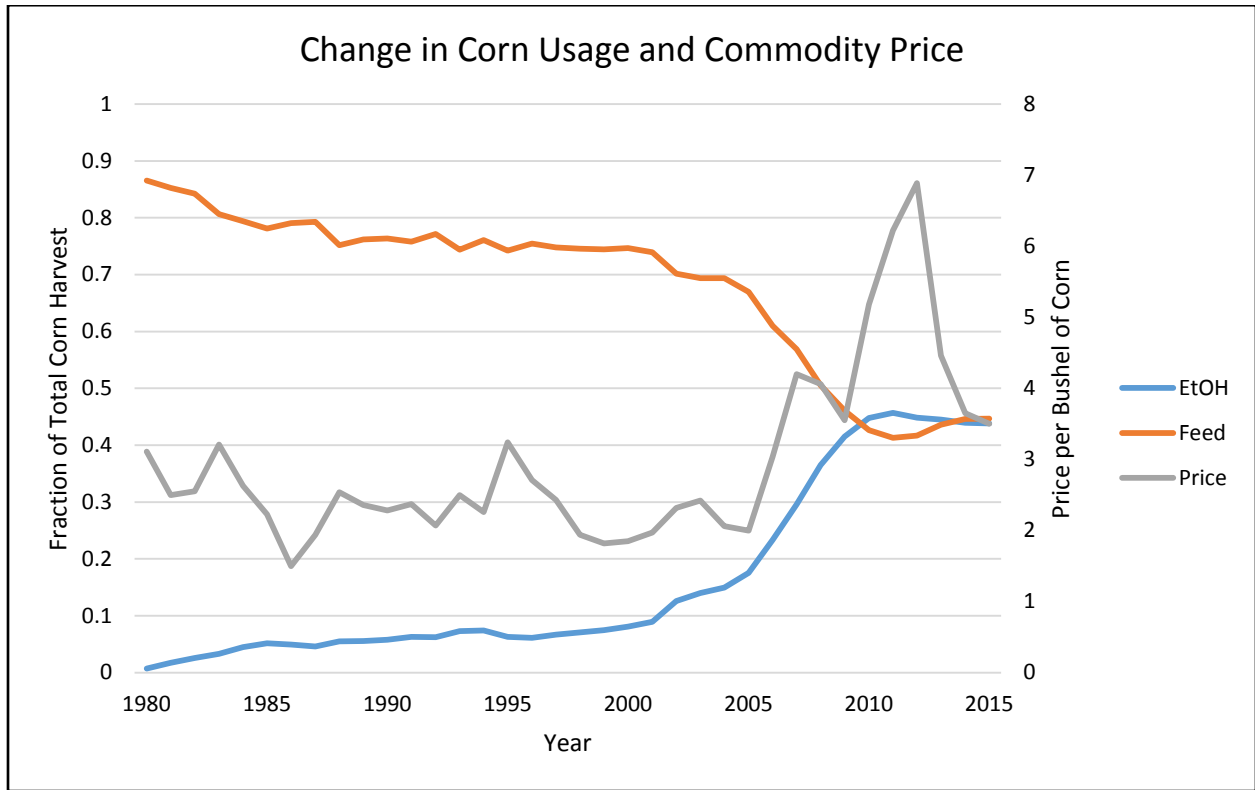
1 USDA, Economic Research Service
2 USDA, World Agriculture Outlook Board, World Agriculture Supply and Demand Estimates

Figure 1.4 Chemical structure of cellulose showing glucose molecules linked by β-1,4 glycosidic bonds (A) and starch showing glucose molecules linked by α-1,4 glycosidic bonds (B). The β-1,4 glycosidic bonds result in long, linear chains of glucose molecules that stack together through hydrogen bonds (C). The α-1,4 glycosidic bonds result in a helical secondary structure with branching occurring at the sixth carbon (D). Electron micrographs showing cellulose in a plant cell wall (E) and starch granules in corn grain (G).

A)
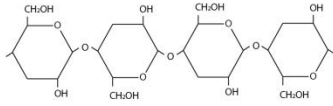
B)
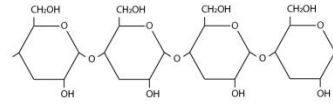
C)
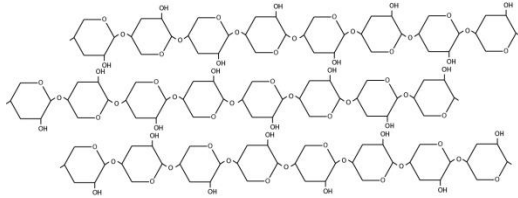
D)
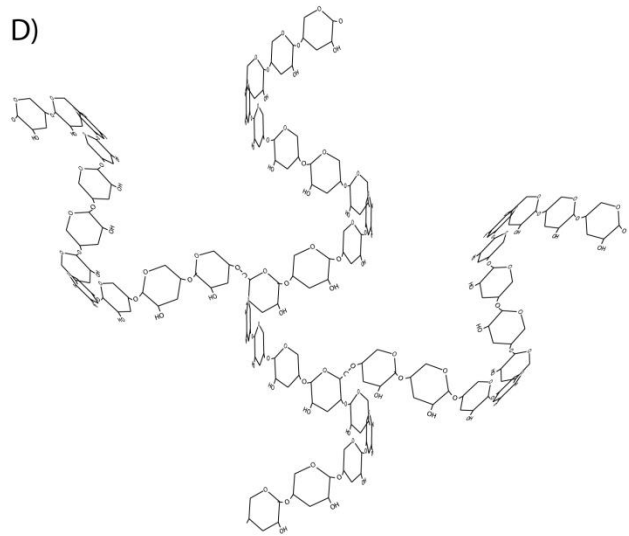
E)
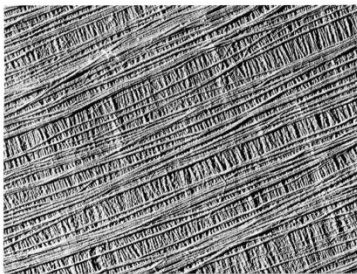
F)
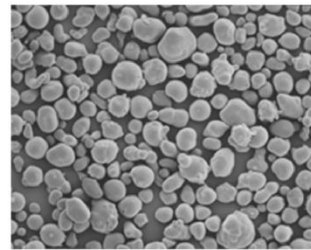
Figure 1.5 Key traits for morphology-based taxonomy of the *Miscanthus* genus. a) a pilose abaxial leaf of *M. sinensis*; glabrous abaxial leaves of b) *M. sacchariflorus* and c) 'Purpurascens'; awns in spikelets of d) *M. sinensis* and e) 'Purpurascens'; f) awnless spikelets of *M. sacchariflorus*; buds at the node of g) *M. ×giganteus* and h) *M. sacchariflorus*; rhizomes of i) *M. sinensis*, j) 'Purpurascens' and k) *M. sacchariflorus*; l) tufted (clumped) growth habit of *M. sinensis*; m) rhizomatous (spreading) growth habit of *M. sacchariflorus*; intermediate growth habits of n) 'Purpurascens'and o) *M. ×giganteus*. Bars represent 5mm. Pictures were taken from May to October 2010 in the Energy Bioscience Institute Farm at UIUC. [As seen in Chae *et al*. (2014)]
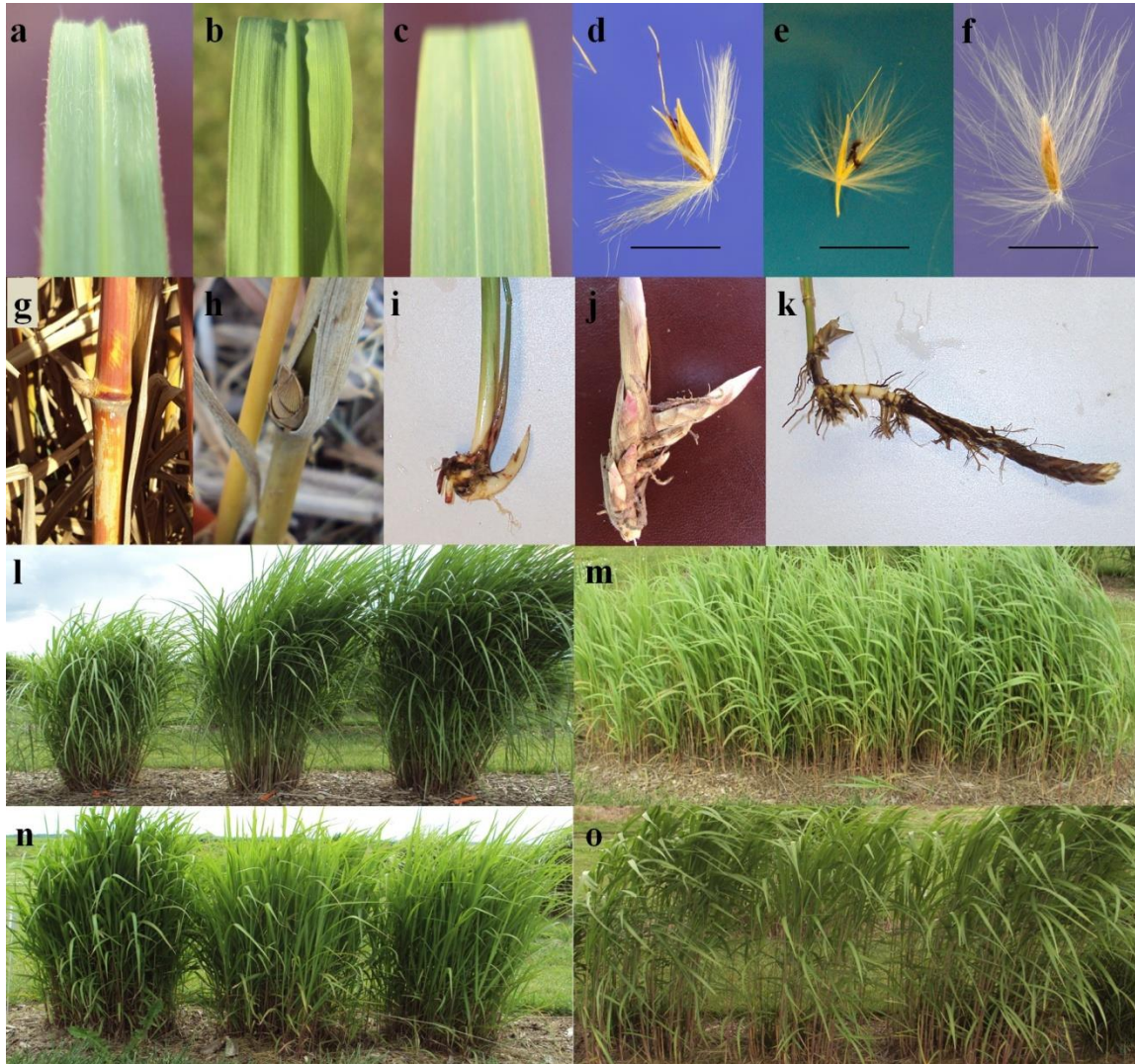
Figure 1.6 Population structure of *M. sinensis* in SE Asia as reported in Clark *et al*. (2014; 2015). The arrows represent the movement of *Miscanthus* following the last glaciation event. The black arrow represents 21,000 years ago, the intermediate grey arrow 14,000 years ago, and the light grey arrows 10,000 years ago.
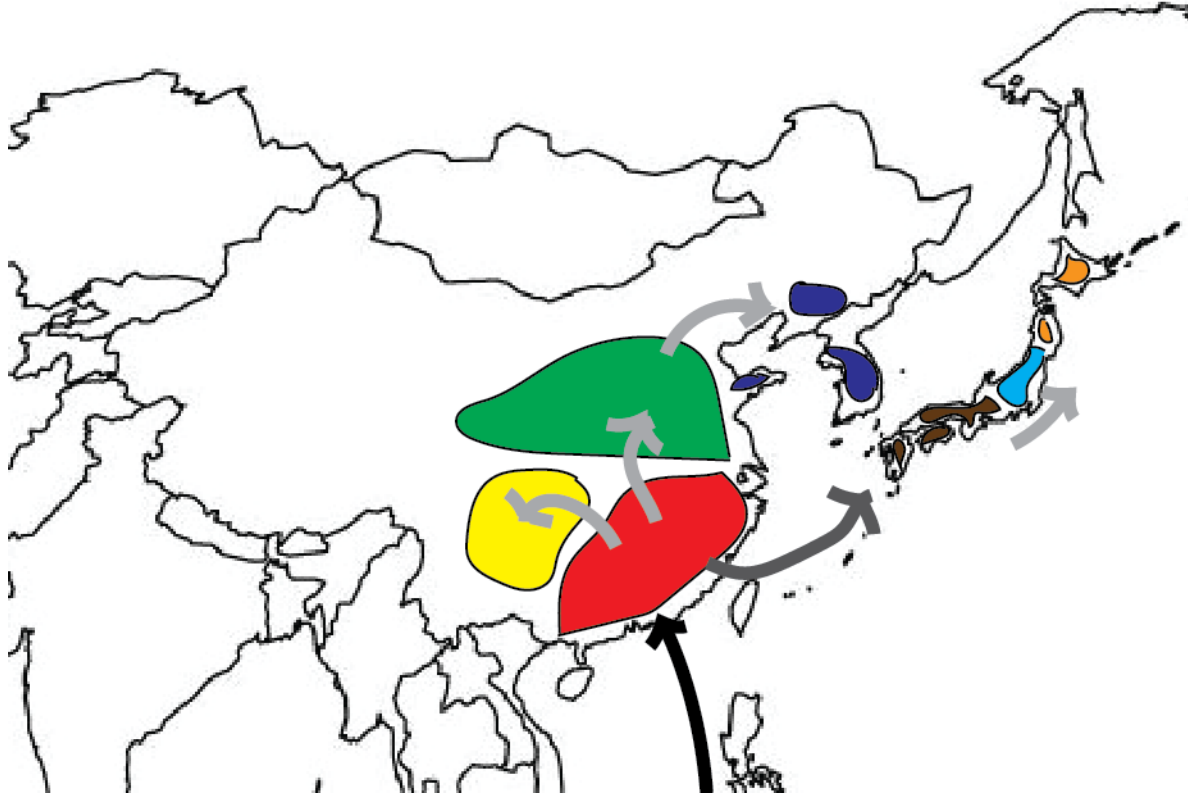
Figure 1.7 A phylogenetic tree of *Miscanthus* and its key relatives (A) along with a close up to better show detail (B). Red circles indicate genome duplication events
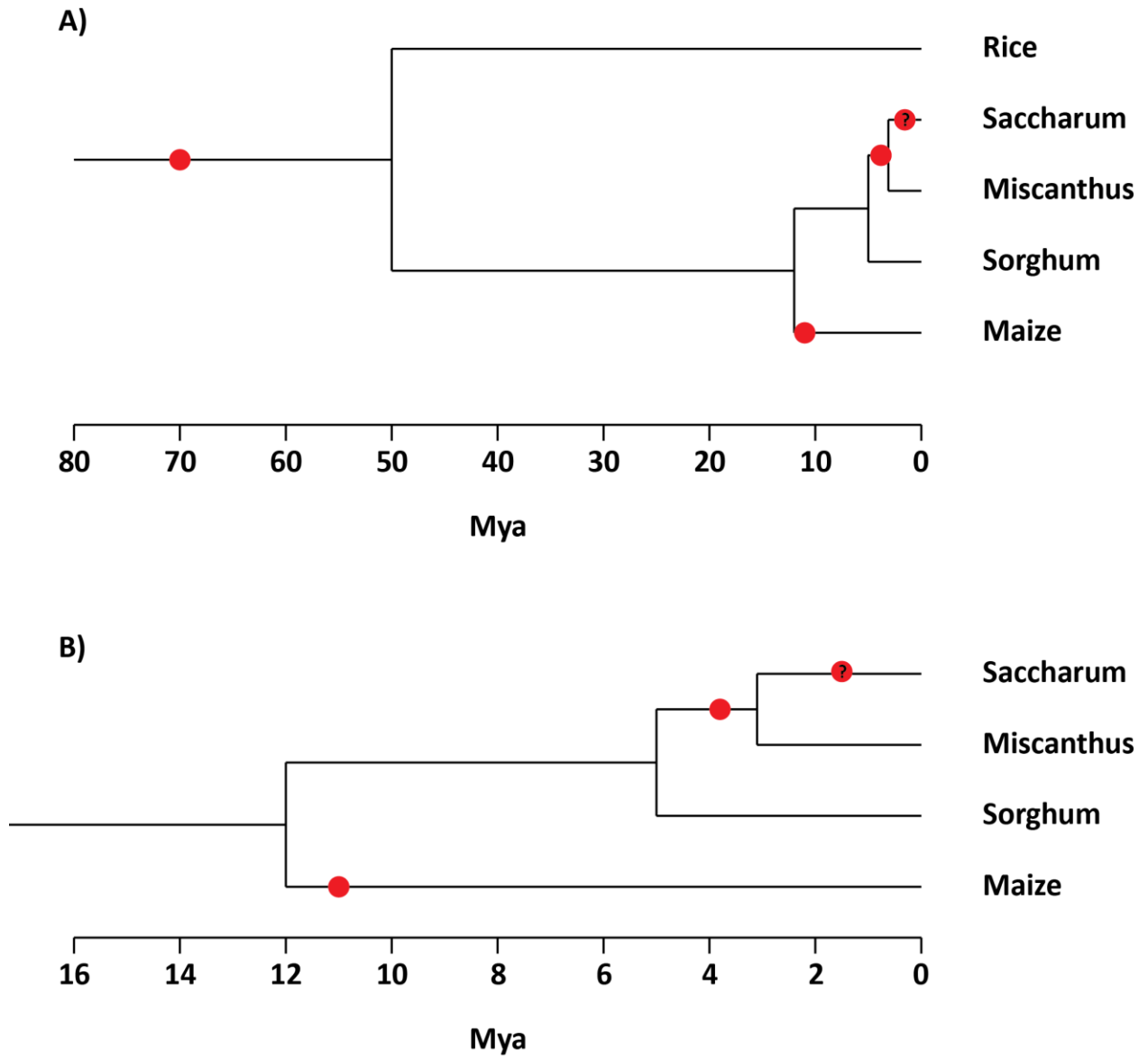
CHAPTER 2: MAPPING THE GENOME OF *MISCANTHUS SINENSIS* FOR QTL ASSOCIATED WITH BIOMASS PRODUCTIVITY[1]

2.1 Abstract

In light of rising energy costs, lignocellulosic ethanol has been identified as a renewable alternative to petroleum based transportation fuels. In an attempt to reach government mandated ethanol production levels, potential plant biofeedstock candidates have been investigated, and cold-tolerant, perennial accessions within the C_4 grass genus *Miscanthus* have been identified as leading contenders in the Midwestern U.S. To facilitate the development of improved cultivars through marker-assisted breeding, a quantitative trait locus (QTL) study was conducted on a full-sib, F_1 mapping population segregating for flowering time, height, leaf width, and yield using a genetic map consisting of 846 segregating SNP and SSR markers. This was a four year study investigating the genetic architecture underlying traits important to biomass production in a population of 221 progeny from a cross between *M. sinensis* 'Grosse Fountaine' and *M. sinensis* 'Undine' established in the spring of 2010; 78 QTLs with LOD scores above the genome-wide, permuted threshold equivalent to a P-value of 0.05 were identified across 13 traits. Forty of the 78 QTLs were detected in multiple years, and the power to detect QTLs appeared to peak in the third year of growth. Both the use of spring emergence and vigor rating as a covariate to account for variation related to differences in establishment increased the power to detect QTLs in the two year establishment period. Finally, a dry period in the middle of the 2012 growing season suggested that yield declines may be due to a decrease in tiller diameter.

---

[1] The 2010−2012 data and much of the text was first published in Gifford *et al.* (2014) and is reprinted here with permission.

2.2 Introduction

Miscanthus is a self-incompatible, perennial, C_4 grass genus capable of producing high biomass yields in temperate environments (Lewandowski *et al*., 2000; Heaton *et al*., 2008). As such, it has been identified as a potential source of biomass for lignocellulosic ethanol production. Improvement through breeding of the high yielding and frequently studied triploid *M.* ×*giganteus* is impossible due to its sterility (Linde-Laursen, 1993). One approach to increase yield and improve agronomic characteristics of *Miscanthus* while maintaining sterility—a desirable trait in terms of reducing the invasive potential of the nonnative *Miscanthus* in the United States—would be to breed the fertile progenitor species of *M.* ×*giganteus* (*M. sinensis* and *M. sacchariflorus*) and use the improved lines to recreate the *M.* ×*giganteus* cross (Sacks *et al*., 2013). Alternatively, *M. sinensis and M. sacchariflorus* may be used as biofeedstocks in their own right as they have been shown to be high yielding (Gauder *et al*., 2012).

While *Miscanthus* plots maintain productivity beyond a decade (Clifton-Brown *et al*., 2001; Gauder *et al*., 2012), a three year establishment phase after initial transplanting exists where phenotypic data is unstable and can be unreliable across clonal replicates (Clifton-Brown & Lewandowski, 2000a). This can be contributed to differences associated with uneven splitting of the rhizome, differences in greenhouse conditions prior to transplanting (e.g., shading and over/under watering), and differences in the field microenvironment at the time of transplanting, leading to variable health of the plant plugs and variable plug establishment. Due to the lengthy establishment phase and the difficulties in phenotyping a large number of individuals, a marker-assisted selection (MAS) program would be a valuable tool in a *Miscanthus* breeding program. Identifying traits that can be reliably measured in the two year establishment period would further enhance selection efficiency.

Prior to the development of a MAS program, quantitative trait locus (QTL) analyses must be performed to establish associations between traits of interest and genetic markers. Generally, this is done using regression analysis or a maximum likelihood model to find a statistical correlation between a marker and a phenotype. At the time of this research, only four QTL studies have been published on *Miscanthus*, and all four of them were conducted on the same population (Atienza *et al*., 2003a, b, c, d). Furthermore, the utility of these studies is limited by the small population size (N=89), use of an incomplete genetic map (28 linkage groups detected whereas *Miscanthus sinensis* has 19 chromosomes), and the use of RAPD markers with limited reproducibility. The low population size limits the power to detect QTLs and leads to drastically overestimated QTL effect sizes (Beavis, 1994). To counter the lack of power resulting from their experimental design, Atienza and his colleagues used multiple methods to set the LOD threshold values used in their experiments, which increases the type one error rate. For example, one of the methods employed, the same used herein, was the permutation method (Churchhill & Doerge, 1994), yet 18 of the 23 QTLs declared in Atenzia *et al*. (2003a) failed to surpass this threshold. Although four *Miscanthus* genetic maps have been recently developed (Kim *et al*., 2012; Ma *et al*., 2012; Swaminathan *et al*., 2012; Liu *et al*., 2015), no QTL studies based on these maps have yet been published. The objective of this study is to identify QTLs for traits related to biomass and to explore methods of removing variation associated with differences in establishment.

2.3 Materials and methods

Plant material and genetic map construction

Two ornamental varieties of *M. sinensis*, 'Grosse Fountaine (GF)' and 'Undine (UN),' originally purchased from local nurseries, were reciprocally crossed in isolation in the Plant Science Laboratory greenhouses at the University of Illinois at Urbana-Champaign in the fall of

2009. Recently, Clark *et al*. (2014) determined through marker data that both accessions ancestry can be traced to southern Japan. Although ornamental accessions were not the preferred choice of parental material, these accessions were among a limited set of materials available at the time the population was developed. Seeds resulting from this cross were planted the ensuing winter in the aforementioned greenhouse (Figure 2.1). The rhizome of the seedlings was split in order to generate three clonal replicates of each progeny. One week prior to transplanting, the plugs were moved to outdoor ground beds to harden off. Just prior to transplanting the plugs, leaf tissue was cut back to 2-4 cm to minimize water loss via transpiration while the root system developed. In total 221 replicated progenies along with each of the parents were transplanted on 1.52 meter centers in May of 2010 at the Energy Biosciences Institute farm in Urbana, IL in a randomized block design. The plot was not watered nor fertilized in 2011 or 2012 although it was watered immediately after transplanting in 2010 and again in the fall of 2013 to reduce stress during flowering. 2,4-D was applied to control broadleaf weeds and atrazine and Medal$^®$ II AT were applied as pre-emergents at the beginning of each growing season. All three were applied at 0.18 kg ha^{-1}.

The genotyping of the population and genetic map used in this analysis were described previously (Swaminathan *et al*., 2012). Briefly, genomic DNA was extracted from young leaf tissue following the Puregene protocol (Qiagen, Valencia, California, USA). GF and UN were screened for polymorphic markers using sugarcane SSR primers developed from expressed sequence tags and intergenic sequences (James *et al*., 2011). Primers that resulted in polymorphic markers were then used to genotype the mapping population; 193 SSR markers were used in the construction of the genetic map. For the purpose of SNP development, total RNA was extracted from young leaf tissue of GF and UN using the CTAB RNA extraction method (Chang *et al*.,

1993). Paired-end RNAseq libraries were sequenced from the two parental accessions at the

Keck Center for Functional Genomics at the University of Illinois on an Illumina GA II platform.

Using this sequence data, a GoldenGate^{TM} genotyping array of 1,536 single nucleotide variants

was developed and used to screen the mapping population. 653 informative SNP markers from

the array were identified and used in the QTL analysis. On average each line had 651 markers

scored.

Using the population and markers described above, a genetic map was constructed using

JoinMap 4.1 (Van Ooijen, 2011). The multipoint maximum likelihood model allows the

integration of parental maps developed from an F_1 population resulting from a cross between two

outbred parents as opposed to developing two separate maps utilizing the pseudo-testcross

approach. The markers mapped to 19 linkage groups, the base chromosome number of the

Miscanthus genus, covering 1890 cM with an average intermarker spacing of 2.7 cM. Although

the map distances between adjacent markers ranged from 0-26.7 cM, only 3.1% of the inter-

marker distances were greater than 10 cM (Swaminathan *et al*., 2012).

Phenotypic data

Phenotypic data for 13 traits were collected in 2011, 2012, and 2013 (year two, three, and

four after establishment; Table 2.1 and Figure 2.2). Limited data were collected on the first year

growth (2010) due to the heavy influence of the establishment effect. Three traits measured in

the first year of growth or at the beginning of the second year of growth (2011); number of

tillers, vigor rating, and spring emergence date; were evaluated for suitability as a covariate to

remove variation associated with the establishment effect. The number of tillers were counted

and a vigor rating score assigned (based on a one to five scale with five being the most vigorous

growth) to each plant one month after planting. Spring emergence was recorded as the date that

photosynthetic leaf tissue first reached a height of 25 cm in the spring following the

establishment year. Least squares means were calculated using the following model $y_{ij} = \mu +$

$B_i + G_j + e_{ij}$ where y equals the response variable, B equals the block, G equals the genotype,

and e equals the error. The block term was treated as a random variable. Least squares means,

Student's t-tests, and Pearson correlation coefficients were calculated using SAS$^®$ 9.3 software

(SAS Institute Inc., Cary, NC). The covariate analysis adjustment was performed as described in

Kuehl (2000) using the following model $y_{ij} = \mu + G_i + B_j + \beta(x_{ij} - \bar{x}..) + e_{ij}$ with β being

the regression coefficient of the covariate on the genotype, x_{ij} is the value of the covariate for a

particular plant, and $\bar{x}.$ is the overall mean of the variable used as the covariate. Covariate

analysis and variance parameter estimates used to calculate the broad-sense heritability and

genetic correlations were also done in SAS 9.3 using the Mixed procedure. The broad-sense

heritabilities were calculated by dividing the variance associated with genotype by the sum of all

variance components. The genetic correlations were calculated as described by Howe *et al*.

(2000) using equation nine. Correlations mentioned henceforth should be assumed to be

phenotypic correlations unless otherwise specified.

Except for leaf morphology, drought response, and flowering time traits, data were

collected at or just prior to harvest in December of 2011, 2012, and 2013. In September of 2011,

2012, and 2013 three leaves, taken at mid-height of each plant, were measured using a LI-3000C

Portable Leaf Area Meter (LI-COR, Inc., Lincoln, NE; Figure 2.2). This instrument

simultaneously records the average leaf width, leaf length, and leaf area. Therefore, leaf area is

not the sum of the total leaf area of each plant but the leaf area of a single leaf. In 2012, an

extended drought in the middle of summer (Figure 2.3) provided the opportunity to take a

drought response measure. The Champaign-Urbana area did not receive rainfall exceeding 25.4

mm from May 30th through August 16th; June precipitation in the area was down 47.5%, and July precipitation was down 87.0% from average according the Illinois State Water Survey. Since no standard measurement has been regularly employed in *Miscanthus*, a drought sensitivity scale for leaf rolling designed for rice was used (International Rice Research Institute, 2009). Flowering time was recorded by two measures: the date when at least one tiller of a plant was heading and when half of a plant's tillers reached anthesis. At harvest, plant heights were recorded in meters. This measurement was taken from the base of the plant to the tip of the tallest inflorescence. In addition, the tiller diameter of three randomly selected stems at mid-plant height (2011 and 2012), the circumference taken at mid-plant height after maximally compressing the tillers together with a zip tie (compressed circumference), and the circumference at the base of each plant (basal circumference) were recorded. In 2013 tiller diameter was recorded near the base of the plant as opposed to at mid-plant height. The ratio of the compressed circumference to the basal circumference was also calculated. Due to the large number of tillers in the second, third, and fourth years of growth, a regression function was calculated to estimate tiller number using the actual tiller counts of 60 plants in 2012. This was done using forward, reverse, and stepwise selection on half of the data for purposes of cross validation within the Reg procedure of SAS 9.3. Although a full second order model was tested, a single order model was selected based on the adjusted r-square, the predicted residual sums of squares, and the mean square prediction error of each model. The model included basal circumference, leaf width, tiller diameter, and yield. The same model was then applied to the 2011 and 2013 data to estimate the number of tillers. Finally, the above ground biomass yield was measured after cutting the plants at a height of ten cm. A subsample of each plant was oven dried for four days at 64° C to calculate percent moisture content and the dry yield; these values are those reported and used in this study.

QTL analysis

QTL analysis was performed on the least squares means described previously with MapQTL 5.0™ (Van Ooijen, 2004) using the cross pollinated (CP) population type and a maximum likelihood mixture model. Under the CP population type, four distributions are fitted, one for each QTL genotype. The expectation-maximization (EM) algorithm is used to predict the most likely genotype of each individual between markers and estimate the mean and variance of each component distribution at each locus (Dempster *et al*., 1977). Because many of the markers in the CP population type are not fully informative, surrounding markers are utilized to determine the probabilities of each QTL genotype. The calculated maximum likelihood is then compared to the likelihood under the null hypothesis that no QTL is segregating at that locus. All QTLs were called based on a permuted, genome-wide significance LOD threshold values based on 1,000 permutations (Churchhill & Doerge, 1994). The target false discovery rate was set to five percent. Initially, an interval mapping approach was taken using a step size of 1 cM; this approach assumes a single QTL is segregating when searching for QTLs. The QTLs found in the initial search were then used in an approximate multiple-QTL model (MQM) approach (Jansen 1993, 1994; Jansen & Stam, 1994). This approach makes no assumption on the number of QTLs segregating. By taking into account variation explained by QTLs already in the model, a more powerful search can be conducted. The MQM approach was reiterated until the LOD scores of no additional loci exceeded the permuted threshold. Confidence intervals were calculated based on the 2-LOD dropoff method (Lander & Botstein, 1989). The difference in the effect of the alleles inherited from GF, UN, and the interaction effects were calculated as described by Sewell *et al*. (2000).

2.4 Results

Phenotypic values and inter-trait correlations

GF and UN, the parents of the population, differed significantly based on a Student's t-test for all traits measured with the exception of spring emergence (not shown), compressed circumference to basal circumference ratio in 2011, leaf rolling in 2012, and compressed circumference in 2013 with a majority of the P-values being < 0.0001 (Table 2.1). In addition, the parents had unequal variances based on the folded F test for both flowering time traits, leaf width, and leaf area in 2011 but only for the ratio between compressed and basal circumferences and heading date in 2012. In 2013 the parents had unequal variances for heading date, leaf length, and leaf area (Table 2.1). The maternal parent had a significant effect in all three years for plant height and leaf length while the maternal parent affected yield in both 2011 and 2012. The number of tillers, the date that half of a plant's tillers reached anthesis, and leaf rolling were affected by maternal parent in 2012 (Table 2.1). When the population was analyzed separately by maternal parent, the variance for the basal circumference of the two populations was unequal in all three years and unequal for leaf length in 2013. Transgressive segregation was apparent in all traits measured (Table 2.1).

The majority of the traits measured herein either increased with age of the stand or stayed roughly the same (Figure 2.4). Plant height, yield, and basal circumference increased with age while the compressed circumference, ratio between the compressed and basal circumferences, flowering time traits, and leaf measurements remained steady with age. The average tiller diameter of a plant was the only trait that decreased from an initial maximum in 2011. After 2011, the tiller diameter remained steady. In some cases, the difference between the individual with the highest average phenotypic value and that with the lowest average phenotypic value

changed with stand age. The disparity between the genotypes with the highest and lowest phenotypic values increased with stand age for yield, basal circumference, and the number of tillers while it decreased with stand age for both flowering time traits. The difference between the individual with the highest average phenotypic value and that with the lowest average phenotypic value was one and a half times larger for tiller diameter in 2011 compared to both 2012 and 2013.

Many significant correlations exist between the traits investigated in this study (Table 2.2). Particularly significant and positive correlations exist between plant compressed circumference, basal circumference, height, and the number of tillers with yield within years: 0.83, 0.68, 0.52, and 0.73 on average, respectively. Interestingly, a significant but weak negative correlation existed between the two flowering time traits and yield in 2012: -0.15 and -0.13 for heading and 50% anthesis date. Similar correlations between the two flowering time traits and yield were not observed in 2013: 019 and 0.18. Genetic correlations, on the other hand, of any trait compared with itself across years are generally very high (0.7-1 on average), and the genetic correlations between tiller diameter and height with the rest of the traits in table 2.2 are relatively low while many of the other traits exhibiting high phenotypic correlations share high genotypic correlations. Due to the number of traits investigated over multiple years, only selected correlations are shown in table 2.2; a complete correlation table can be found in Table A.1.

Covariate analysis

In general, when the phenotypic means for each trait were adjusted for either spring emergence or the vigor rating, the LOD scores of putative QTL loci increased, resulting in an increase in the number of QTLs. In contrast, adjusting the phenotypic means for the number of tillers one month after planting did not improve the power to detect QTLs. When analyzed for

QTLs themselves, the latter was also the only covariate for which any QTLs were detected. An additional three QTLs were identified across all phenotypic variables when using spring emergence or the vigor rating adjusted means, but two previously identified QTLs failed to meet the significance threshold. Of the three additional QTLs detected using this method, two were detected in subsequent years. Likewise, of the two that were not detected using the covariate adjusted means, one was not detected in subsequent years. The utility of this technique was not equal across all traits; it worked well for yield and basal circumference (Figure 2.5) but was ineffective for flowering time traits and leaf traits.

QTL results

78 QTLs, QTLs with LOD values above the permuted genome-wide significance threshold, have been identified across 13 traits (Table 2.3). Forty of the 75 QTLs that could have been identified in multiple years (leaf rolling was only taken in 2012) were indeed detected in at least two years (Figure 2.6). Two, eight, fifteen, and ten QTLs were unique to 2010, 2011, 2012, and 2013, respectively. In 2011 the number of QTLs per trait ranged from one for plant height, the number of tillers per plant, and the ratio between the compressed and basal circumferences to seven for the date that half of a plant's tillers reached anthesis. A larger or equal number of QTLs were found for each trait in 2012 compared to 2011 except for tiller diameter. Interestingly, the number of QTLs per trait declined or stayed the same from 2012 to 2013 in all traits except for yield and heading date. In 2012 the number of QTLs per trait ranged from two for tiller diameter and leaf length to seven for both flowering time traits while in 2013 the number of QTLs per trait ranged from zero for leaf length to eight for heading date.

The percent variation explained (PVE) by a QTL ranged from 3.1% (LA) to 79.0% (LW) in 2011, 3.0% (LW) to 75.2% (LW) in 2012, and 2.6% (HD) to 66.0% (LW) in 2013. The

summation of the PVE described by all QTLs for an individual trait—excluding traits for which

no QTLs were found—ranged from 9.5% (CCBC) to 91.7% (LA) in 2011, 21.3% (LL) to 88.5%

(LW) in 2012, and 21.0% (CCBC) to 85.1% (A) in 2013. As with all QTL studies, it is important

to keep in mind that these estimates are biased upwards due to sampling QTL effects from a

truncated distribution. Further, this overestimation is more severe with small effect QTLs, which

would be the expectation for many of these complex traits. The sum of the PVE for certain traits

increased dramatically from 2011 to later years while others remained relatively constant (Table

2.1).

Many of the QTLs that we detected colocalize with QTLs for other traits. For example,

QTL clusters exist on chromosome three and six (Figure 2.7). Basal circumference, tiller

diameter, the number of tillers, heading date, 50% anthesis, leaf width, leaf rolling, and leaf area

all cosegregate on linkage group three in 2012 while yield, compressed circumference, basal

circumference, the number of tillers, leaf area, leaf length, and leaf width cosegregate on linkage

group six. A similar phenomenon occurred in 2011 and 2013 although there were fewer traits

with overlapping confidence intervals. A goodness of fit test was conducted to determine

whether or not QTLs were randomly distributed along a linkage group or if the apparent clusters

were statistically significant using the Poisson distribution function described by Isemura *et al*.

(2007). The null hypothesis of QTLs being randomly distributed along linkage groups according

to the Poisson distribution was rejected in 2011-2013 for linkage groups three and six (P<0.001)

(Table 2.4).

2.5 Discussion

Phenotypic inferences

While the parental lines do not represent the extreme ends of variation seen within *M. sinensis*, the differences between the two parents are more similar to the differences that would be expected when crossing two elite cultivars. Furthermore, crosses between extremely small and extremely large genotypes may be less informative as the QTLs underlying these differences are likely to be fixed in elite breeding material. For comparison, *M.* ×*giganteus* 'Illinois' established in the same year and grown on an adjacent plot yielded 4.9 kg of biomass, had an average height of 2.7 m, and average tiller diameter of 6.7 mm in 2012. These values are 49%, 52%, and 308% larger than the mean of this mapping population in the same year. It should be noted though that *M. sinensis* 'Cabaret,' an ornamental cultivar with similar single plant yield as *M. sinensis* 'Grosse Fountaine,' was used in a cross with a tetraploid *M. sacchariflorus* to produce new triploid genotypes of *M.* ×*giganteus*. Two of the four progenies had comparable yields to *M.* ×*giganteus* 'Illinois' in initial side by side trials (Erik Sacks, personal communication).

Transgressive segregation is a common phenomenon in plants, particularly in intraspecific crosses, and was reported to be more frequent within domesticated, inbreeding species (Rieseberg *et al*., 1999). Many studies have attributed this phenomenon to the complementary action of additive genes (Xu *et al*., 1998; Rieseberg *et al*., 2003; Mao *et al*., 2011). While *M. sinensis* is neither inbreeding nor domesticated per se, the large number of traits showing transgressive segregation is a promising sign in terms of breeding for improved agronomic traits and lends support for the use of MAS, which increases the efficiency of combining positive alleles to exploit additive gene action. As this is a cross between ornamental

cultivars not selected for biomass yield, it is yet to be determined if a similar phenomenon is observed in crosses resulting from elite breeding material.

In addition to transgressive segregation, the small yet consistent detection of significant differences between reciprocal populations for height, yield, and leaf length may have implications in the way *Miscanthus* breeding programs design breeding populations and heterotic pools. Traits that show the trend of maternal inheritance should be bred strongly in the maternal lineage. In addition, this could be important in the breeding design using interspecific crosses. *M. sacchariflorus* is the maternal parent of *M. ×giganteus* (Hodkinson *et al*., 2002c; Swaminathan *et al*., 2010). In addition, new varieties of *M. ×giganteus* discovered in Japan (Nishiwaki *et al*., 2011), made by M. Deuter (Tinplant; http://www.tinplant-gmbh.de), and created at the University of Illinois (Chae *et al*., 2013; Clark *et al*., 2015) all share *M. sacchariflorus* as the maternal parent. While having multiple copies of one genome may be more influential than any maternal effect, it would be interesting to recreate the *M. ×giganteus* cross with *M. sinensis* as the maternal parent. Only a single incidence has been reported in the literature where *M. sinensis* is the maternal parent of a triploid *M. ×giganteus* progeny (Hirayoshi *et al*., 1957). Reciprocal crosses between diploid and artificially induced tetraploid *M. sinensis* and *M. sacchariflorus* and vice versa should help separate the question of genome dosage from that of maternal inheritance.

As noted in the materials and methods section, a severe drought accompanied by extreme heat in the middle of the 2012 growing season allowed us to measure a drought response variable. Unfortunately, the drought was severe enough that we did not see much variation to drought response in the population. By comparing between the three years, we can infer how this population responds to drought. The average yield did not increase between the second (2011) and third (2012) year after establishment, a result that would not be expected in a favorable

environment (Clifton-Brown & Lewandowski, 2000b). The relationship between drought and biomass in *M.* ×*giganteus* is well established (Heaton *et al*., 2004a; Christian & Haase, 2001), so this result is not surprising. As expected, in 2013 we did detect an increase in yield. The basal circumference and the number of tillers increased with stand age despite the 2012 drought as would be expected because these traits are determined early in the growing season prior to the drought. Christian and Haase (2001) noted a similar phenomenon in *M.* ×*giganteus*. Leaf morphology and flowering time did not seem to be largely affected by the dry conditions in the middle of the 2012 growing season. It should be noted that the timing of the drought may have influenced this observation in regards to flowering as rainfall did resume during peak flowering. In 2013 a dry spell completely stalled flowering progress. The full extent of the affect of this untimely dry spell on flowering could not be explored due to the necessity to irrigate the field to continue data collection on a parallel study. Based on observation though, it would not be surprising if the heat and lack of water during flowering would result in poor seed set. The ability to avoid these conditions could become a factor if commercial *Miscanthus* seed production ever came to fruition. Although total leaf area of *M.* ×*giganteus* decreases under drought conditions, *M. sinensis* leaf area was not found to be affected by drought under controlled conditions (Clifton-Brown & Lewandowski, 2000b; Clifton-Brown *et al*., 2002). Though we cannot say if total leaf area was affected by the drought, individual leaf area was not affected in our experiment. It appears that tiller diameter, which decreased dramatically in 2012, is the trait that was most heavily affected by the midseason drought. To the author's knowledge, this interaction has never been reported previously. While the tiller diameter did not fully recover in 2013, this may be the result of confounding factors associated with a difference in phenotyping protocols.

This suggests that tiller diameter may be useful as a proxy to drought tolerance in *M. sinensis* but further exploration is needed.

Early selection

Establishing correlations between traits in the establishment period with each other and with traits in mature stands has great utility in a perennial plant breeding program. Using traits that can be reliably measured in early years of establishment to model future performance will reduce the time necessary to make selections. Minimally, modeling could be used to remove the genotypes with little potential. To date, many of the published studies on *Miscanthus* deal with correlations between traits and yield measured on a per plot basis (Clifton-Brown & Lewandowski, 2002; Jeżowski, 2008; Gauder *et al*., 2012). In those cases, height, tiller density, and tiller diameter are most highly correlated with yield. While high plot yield is the primary goal in improving *Miscanthus*, it is not feasible to conduct early tests of genotypes during the establishment period for this trait. In this mapping population height and tiller diameter appear to be mainly controlled by different genes than plant yield as judged by the genotypic correlations, suggesting that plant yield may not be the optimum selection criteria for early selection. It is important to remember that the ultimate goal is not to model yield on a per plant basis but to correlate traits measurable on a per plant basis with those measured on an area basis, an objective that is outside of the scope of this study.

Another way to reduce the time per breeding cycle is to minimize the establishment effect. Our attempt to reduce this had mixed results. One of the key assumptions inherent in an analysis of covariance is that the covariant must not result from the treatment. In this case, the treatment is the genotype; therefore, in order for this analytical method to be valid, the timing of spring regrowth following the first year of growth, the number of tillers one month after planting,

and the vigor rating must be due to the establishment effect as opposed to a genetic predisposition towards early or late spring regrowth, a small or large number of tillers, and weak or vigorous growth. Hence, we would not be able to detect QTLs for an appropriate covariate, and the covariate should have a negligible heritability. Spring emergence and the vigor rating, the two covariates that appeared to be successful, did not have any QTLs associated with them and their broad-sense heritabilities were 0.16 and 0.23, respectively, further supporting their usefulness as covariates to remove variation associated with establishment.

Since the vigor rating was taken soon after transplanting, the vigor of the plants likely had more to do with the plant's health at the time of transplanting and less with the genotype. It is possible that taking the vigor rating even earlier after transplanting may be more effective. On the other hand, plants that do not establish well in the first year would have less opportunity to assimilate carbon and inorganic nutrients and translocate them to their rhizomes for overwintering. In turn, this leads to a delayed emergence the following year as fewer nutrients are available to support vigorous spring regrowth. This relationship between carbon reserves in rhizomes and spring regrowth has long been recognized in the literature (White, 1973; Dhont *et al*., 2002; Dhont *et al*., 2004). The strong negative correlation between spring emergence and yield provide further support for this hypothesis. Although this covariate analysis did show utility in accounting for the establishment effect, neither covariates were perfect candidates, and appropriate covariates may vary with the population. For these reasons, additional research needs to be conducted to clarify the broad applicability of these covariates and other measurements should be explored.

QTL analysis

The increased power to detect QTLs in the third year after planting (2012) is likely a consequence of a decrease in variation caused by the establishment effect. The establishment effect can persist for variable lengths of time, depending on the trait of interest, and in some cases it may never fully disappear (Kaiser, 2014). Once the length of time required for the establishment effect to dissipate is established for various traits of interest, breeding programs can more efficiently utilize the resources dedicated to phenotyping plants in the establishment phase. In this case, leaf morphology, stem diameter, and flowering could be measured accurately in year two while height, number of tillers, and yield measurements were less representative of the potential of the genotype based on the broad-sense heritabilities and cumulative PVE (Table 2.1). In some instances, such as yield and basal circumference, a steady increase in heritability was seen with stand age, suggesting that the longer after establishment researchers wait, the more accurately the phenotype will reflect the genotype. A balance must be struck in these cases between accuracy and the time constraints inherent in plant breeding. The fact that the heritability of some traits, such as the two flowering traits, did not appear to be related to stand age suggests that other factors outside of the establishment is producing variation. The most likely candidate for this would be the environment.

It is not unexpected that QTLs would differ between years for a number of reasons: 1) the weather was drastically different between the three years, 2) the plot environment was different between the three years as the plants grew in, and 3) the maturity of the plants was different between years. The close yet non-overlapping confidence intervals for QTLs across years suggests that the confidence intervals calculated based on the 2-LOD dropoff method may be too narrow. Unfortunately, the construction of confidence intervals using the nonparametric

55

bootstrapping method proposed by Visscher *et al.* (1996) is time prohibitive using the software available for this population structure.

Beyond comparing between years, we detected a nonrandom distribution of QTLs, particularly on linkage groups three and six. Differences in local recombination rates can influence the number of genes per centimorgan and dramatically affect the results of a QTL experiment: in simulations the strongest QTLs were found in areas with low recombination rate, such as near centromeres and telomeres (Noor *et al.*, 2001). Additionally, it was found that the decrease in the number of coding sequences near centromeres was insufficient to offset the decrease in recombination rate. Clustering of QTLs around centromeres was indeed found in maize (Khavkin & Coe, 1997) while in rice QTL studies no such result was observed (Cai & Morishima, 2002; Onishi *et al.*, 2007). Rather, the clusters of QTLs appeared to be associated with regions of segregation distortion. QTL clusters are hypothesized to provide a selective advantage as this keeps co-adapted, beneficial traits together (Le Thierry D'ennequin *et al.*, 1999), particularly in outcrossing crops such as *Miscanthus*. While interval mapping has a slight tendency to place QTLs near markers (Darvasi *et al.*, 1993), the marker density in the region of clusters makes it unlikely that the clusters are an artifact of this mapping bias. Whether the clusters of QTLs in this study are a product of natural selection favoring co-adapted traits, pleiotropy, tightly linked genes, or a bias due to differences in recombination rates is yet to be determined.

Although the *Miscanthus* genus has recently undergone a genome duplication event (Kim *et al.*, 2012; Ma *et al.*, 2012; Swaminathan *et al.*, 2012), we see few examples of QTLs occurring on paralogous chromosomes. Of the few we do see, Y1 and Y2 are the only examples where QTLs occur on paralogous chromosomes in all three years. There are a number of possible

56

explanations for why we find few QTLs for a given trait mapping to both paralogous chromosomes. One could be differential gene loss from the two subgenomes such that functional variation is only present on one of the duplicated chromosomes; however, recent evidence from genespace sequencing of *Miscanthus* (Kirkpatrick, 2013) indicates such gene loss is not common. Other explanations include that the GF x UN population is fixed at one of the paralogous loci, neofunctionalization has occurred, paralogous loci are expressed at different times in maturity or under different stresses, or a silencing mechanism acts differentially on paralogs.

M. sacchariflorus is a quantitative short-day plant (Jenson *et al.*, 2013), similar to sorghum and sugarcane, but *M. sinensis* was described as day neutral by Deuter (2000). Recently, Jenson *et al.* (2011) questioned this observation. They observed that flowering time in *M. sinensis* is more complicated than that; it depends on a number of factors, including thermal time, temperature, photoperiod, and precipitation. This coincides with our observation that plants would have a flush of flowering after precipitation events. Despite this divergence in floral initiation in *M. sinensis* from sorghum, using the physical positions of *Miscanthus* markers on the sorghum genome and the meta-QTL analysis performed by Mace and Jordan (2011), we were able to find an overlap in the confidence intervals of four sorghum maturity meta-QTLs and one maturity gene (Ma5) with our maturity QTLs. The diverse interactions between flowering initiation and the environment within the genus in conjunction with the importance of flowering time to adapting cultivars to local environments, make the study of flowering time very interesting.

The largest difference in the effect of an allele inherited from a parent for a maturity QTL in this analysis was 7.26 days. This is moderate in comparison to other species. The genetic

architecture for maize flowering time, for example, consists of many small effect QTLs working in an additive nature. The largest effect QTL found for days to silking (DS) in the NAM population was 1.7 days, and over 98% of the QTLs identified for DS had an additive effect below one day (Buckler *et al*., 2009). It is important though to keep in mind that these comparisons are between family-based and population-based experiments and that the Beavis effect may lead to an unequal overestimation of the QTL effect sizes between the two experiments. GWAS studies of flowering time in *Miscanthus* that are beginning to emerge (Clark, 2015) should help further refine the characterization of the genetic architecture underlying flowering in *Miscanthus*. Other species such as *Arabidopsis*, sorghum, and rice tend to show maturity genes with higher additive effects (Alonso-Blanco *et al*., 1998; Lin *et al*., 1995; Yano *et al*., 1997). One explanation of these differences is in the nature of these crops' reproductive strategies. Buckler *et al*. (2009) proposed that natural selection would favor a large number of small effect maturity QTLs in outcrossing species to ensure some degree of flowering overlap within the population. The same selection pressure would not be present in the self-pollinated species mentioned previously. Since *Miscanthus* is an obligate outcrossing species, a similar flowering architecture to maize would seem plausible. However, if you take into consideration the duration of flowering time within a single plant, the selection pressure for small effect maturity genes for *Miscanthus* would decrease. Due to the large number of tillers present in an individual *Miscanthus* plant, active flowering can persist for weeks. This increased flowering duration has also been observed by Jenson and colleagues (2011) and could explain the difference in genetic architecture between maize, *Miscanthus*, and self-fertilizing species. It is imperative to look at the genetic architecture of flowering over a number of *Miscanthus*

populations prior to drawing any far reaching conclusions as this particular mapping population encompasses a fraction of the flowering time diversity present within the *Miscanthus* genus.

Continuing with the comparison between species, *Miscanthus* does not show very high heritability for plant height (0.52-0.60), unlike many of its closer relatives. The broad sense heritability for plant height for maize and sorghum is generally around 0.9 (Schon *et al*., 2004; Flint-Garcia *et al*., 2005; Eckebil *et al*., 1977; Tomar *et al*., 2012). This could be partially due to the fact that the genotypes in this study were only tested in one environment with three replications or that this particular population has lower genetic variation for height. Alternatively, plant height is less strait forward to measure in *Miscanthus* than in maize and sorghum as both of the latter have a single dominate tiller. In an attempt to measure plant height more accurately, we measured the average tiller height in addition to the tallest tiller in 2012. Fewer QTLs were found for this alternate measure of plant height and the heritability was lower (data not shown). Additional techniques should be explored and genotypes should be tested across multiple environments to improve heritability for this key trait.

Tiller diameter, a trait found to be highly correlated with plot yield in other studies (Jeżowski, 2008; Gauder *et al*., 2012), showed low heritability in all three years as well. In order to keep the effects of plant maturity separate from the effects that changing our phenotyping protocol may cause, we tried not to drastically alter our phenotyping methods from 2011 to 2012. We did, however, still want to improve the accuracy of our phenotyping methods. In two other populations, we compared the accuracy of two different stem diameter measuring techniques as judged by the heritability values: the protocol used herein and measuring the tiller diameter in the lab at the first internode after the leaf sheath was removed. The difference in the heritability between measuring techniques was not drastic in one population (0.51 vs. 0.52) but was

considerably increased by using the new method in the other (0.45 vs. 0.57). While stem diameter is an important trait, the labor associated with the second technique and the inconsistent improvement in accuracy makes it less than ideal. In 2013, stem diameter was taken near the base of the plant as opposed to at mid-height. While the heritability was increased (0.35 to 0.41) from 2012 to 2013, it is difficult to determine if this is a result of the measuring technique or an effect of the environment. Regardless, these techniques leave much to be desired.

This study begins to establish linkage between reproducible genetic markers and a number of key agronomic traits in *M. sinensis*. The reproducibility between years, stringent significance threshold, and syntenic relationship with sorghum all provide strong evidence that these QTLs truly represent important regions of genetic control. As additional QTL and GWAS studies begin to emerge within this genus, we will begin to get a better grasp of which regions of the genome are most important to key traits across a more diverse set of germplasm. We have also shown the utility of spring emergence and vigor rating as a covariate to account for variation related to differences in establishment, which has direct applications to germplasm screening and QTL studies alike. In addition to QTLs, we report highly significant, positive correlations between compressed plant circumference, basal circumference, height, and the number of tillers with yield: 0.83, 0.68, 0.52, and 0.73, respectively. In the ongoing effort to improve *Miscanthus* as a biofuel crop, this work represents a significant stride towards the development of a MAS breeding program.

2.6 Figures and tables

Figure 2.1 Pictures of the mapping population at various times throughout development: (a) Seedlings in the greenhouse prior to rhizome division (b) A Google Earth image of the plot at the EBI farm in the fall of 2010 (c) A picture of the field planting taken in July of 2012 prior to flowering
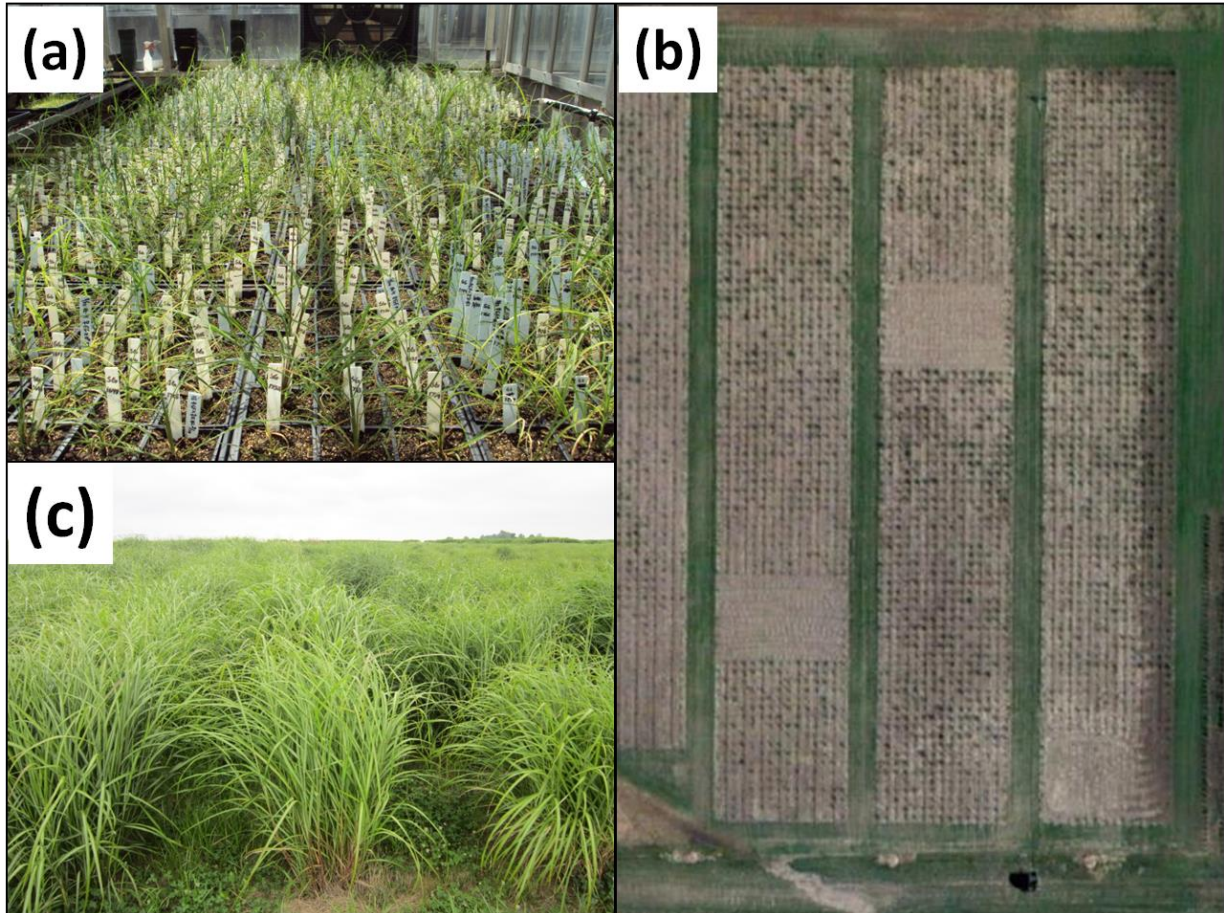
Figure 2.2 Photographs illustrating various phenotypic measurements: (a) An overhead view of a harvested plant around which the basal circumference would be measured (b) Stem diameter being taken on a tiller at mid-plant height (c) A plant tied up prior to harvest; compressed circumference would be taken at the height of the zip-tie (d) A leaf being measured by the LI-3000C Portable Leaf Area Meter (e) A plant just beginning to head (f) A plant at 50% anthesis
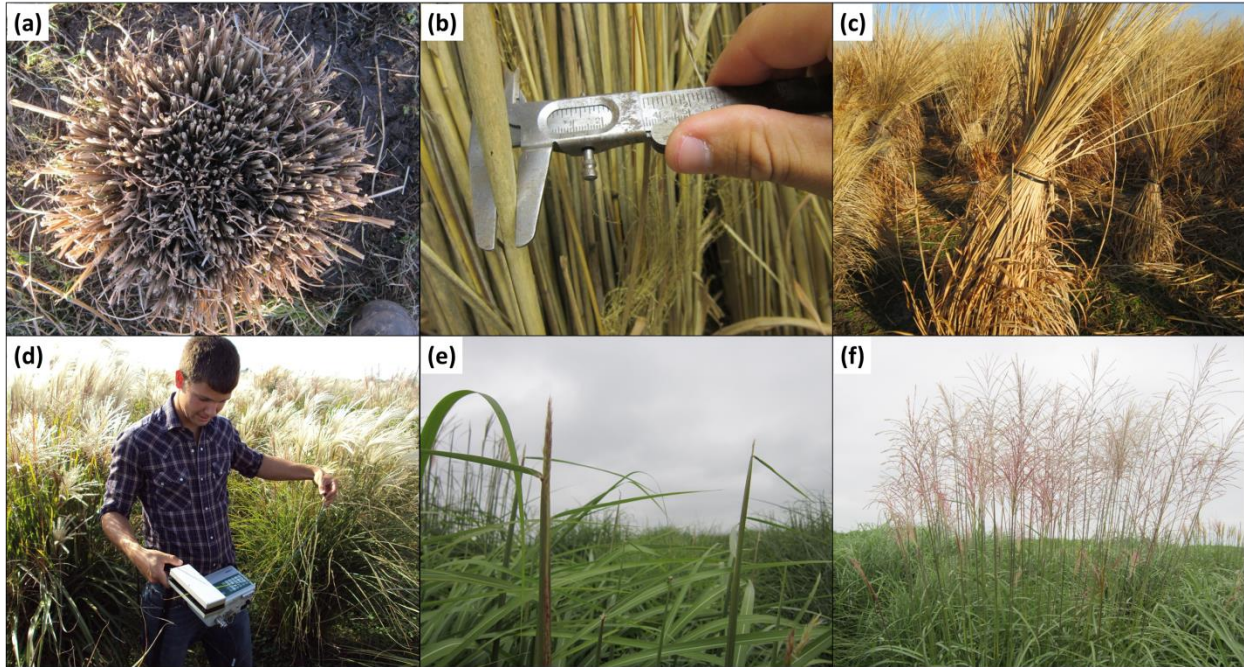
Figure 2.3 The number of growing degree days and precipitation levels depicting the different growing conditions in 2010-2013.

Growing Degree Days 2010-2013

Precipitation 2010-2013
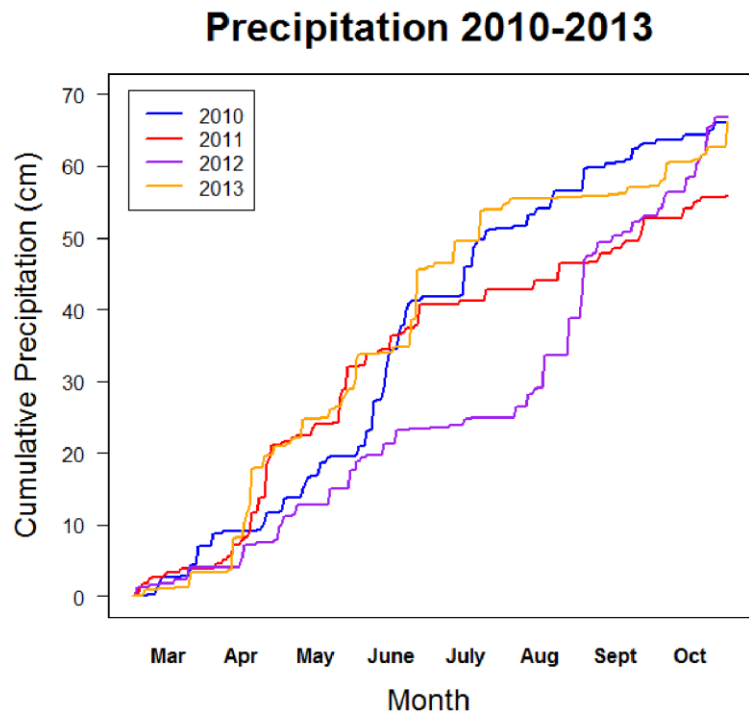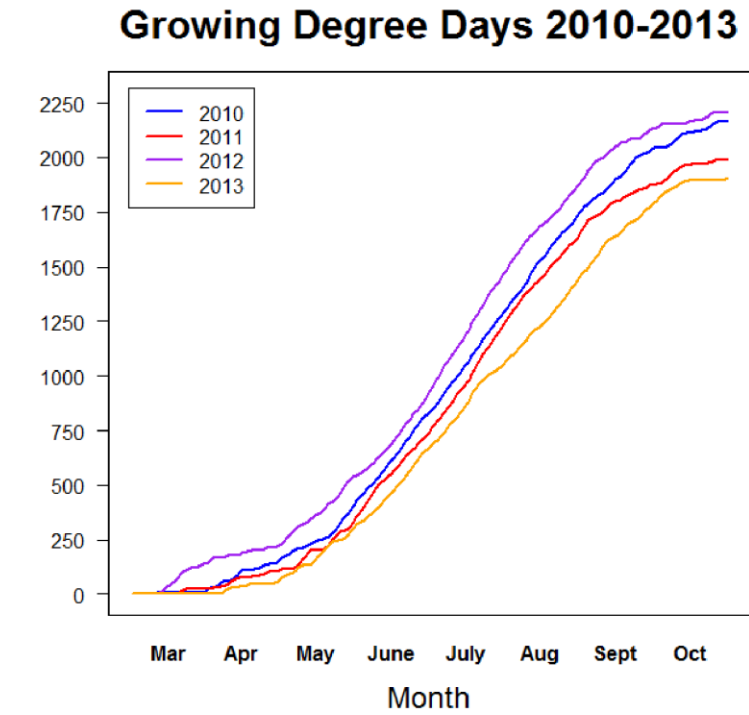
Figure 2.4 The changes in the mean of the population, the two parents, and the range of the population as stand age increased for all of the traits measured excluding leaf rolling.
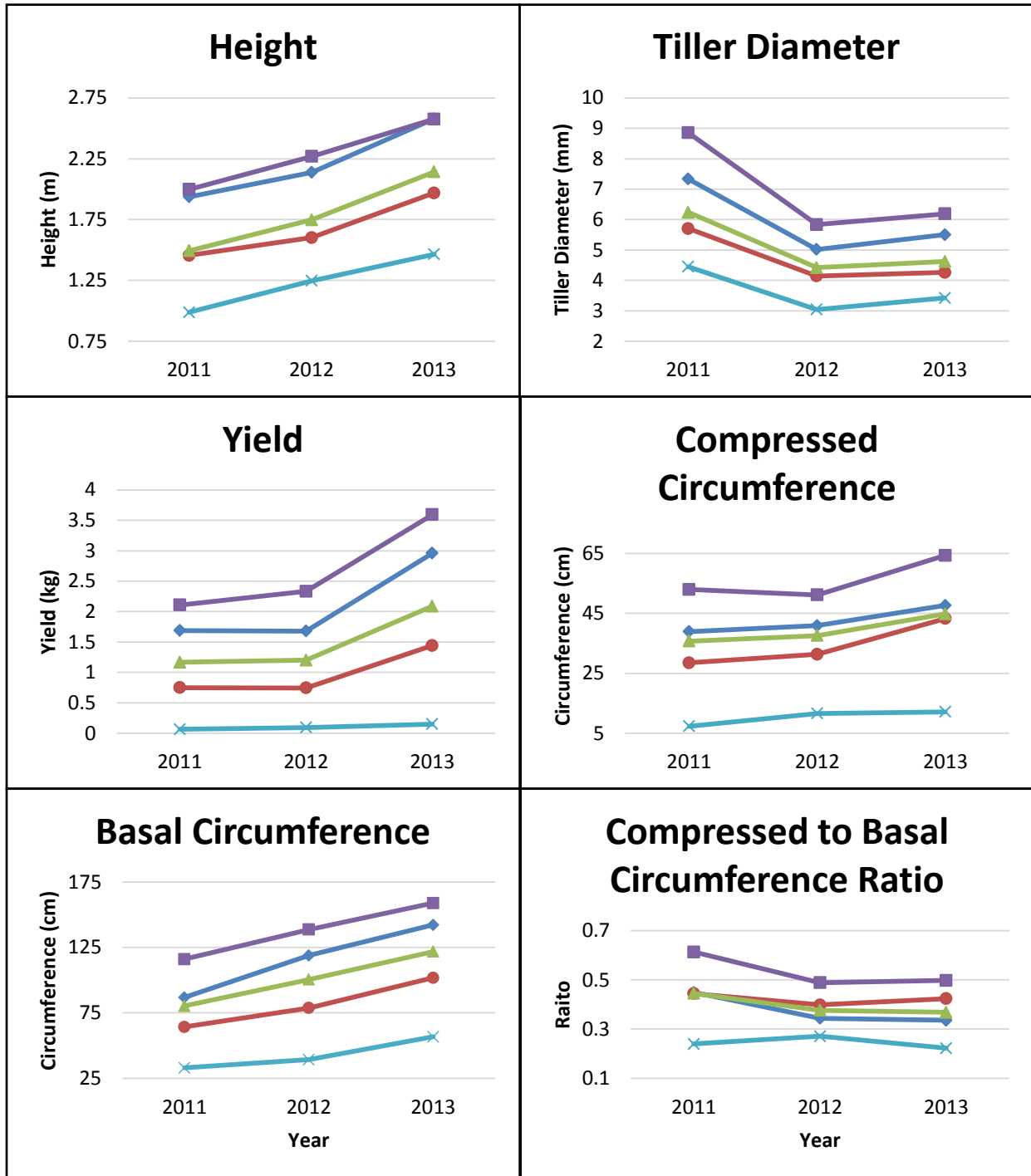
Figure 2.4 (Continued)
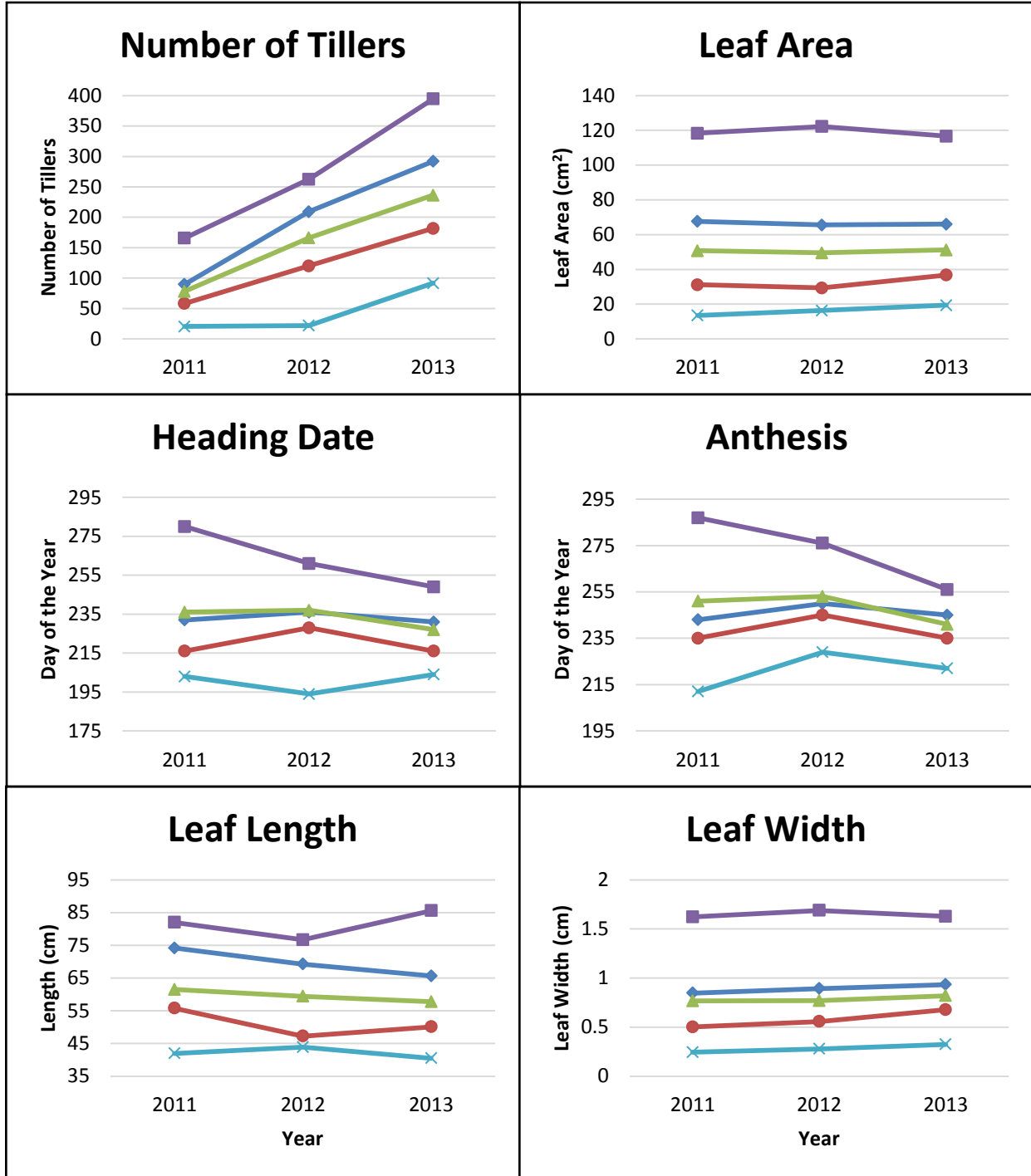
Figure 2.5 P-value plot for basal circumference comparing the results of the QTL analysis run on unadjusted and covariate adjusted values with that of the 2012 QTL analysis results. Only the chromosomes on which QTL are located are depicted. The horizontal line represents the significance threshold. Triangles indicate cases were the covariate adjusted data predicted the 2012 results successfully.
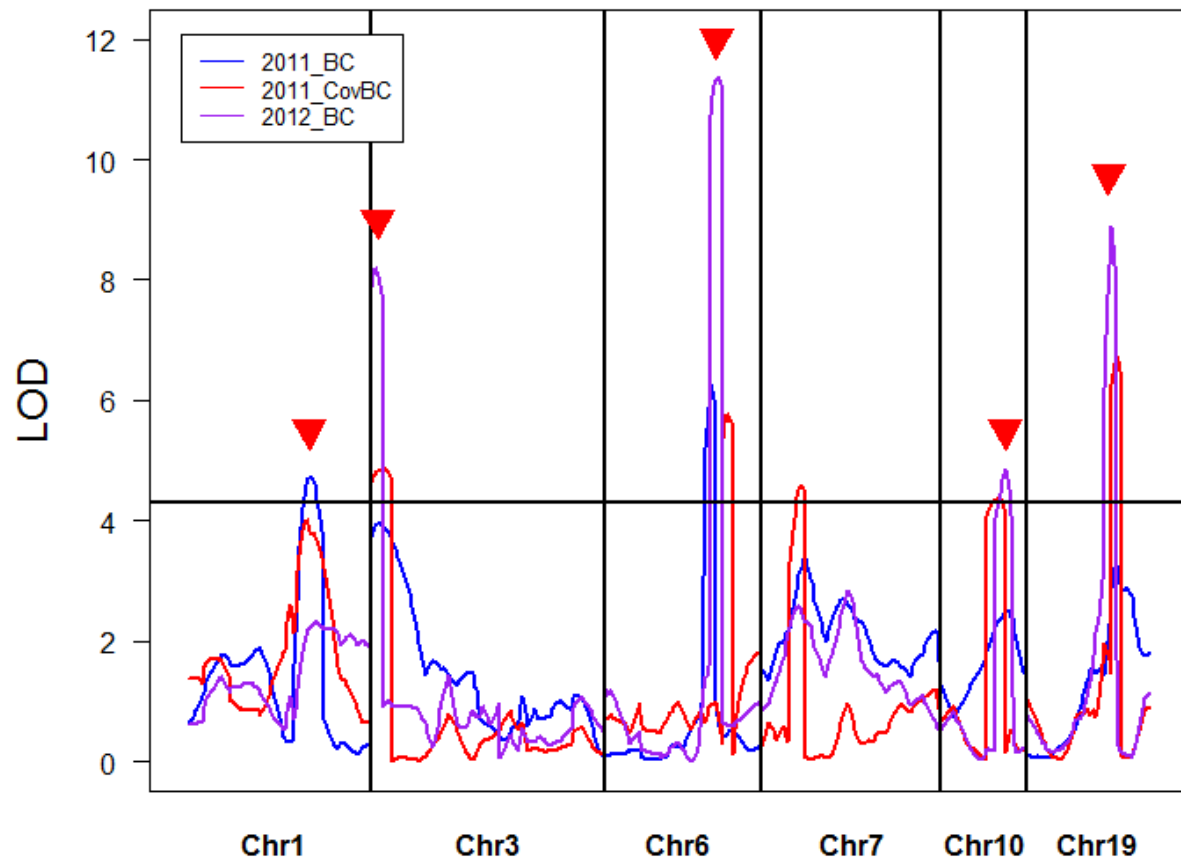
Figure 2.6 Localization of QTLs that were found in multiple years. The genetic distance is shown on the left in centimorgans, the linkage group numbers are shown at the top, horizontal lines on the linkage groups represent the placement of the genetic markers, stars represent the peak LOD of the QTL, and bars to the right indicate the combined 2-LOD support interval.

Location of QTL Stable over Multiple Years on 19 Miscanthus Linkage Groups
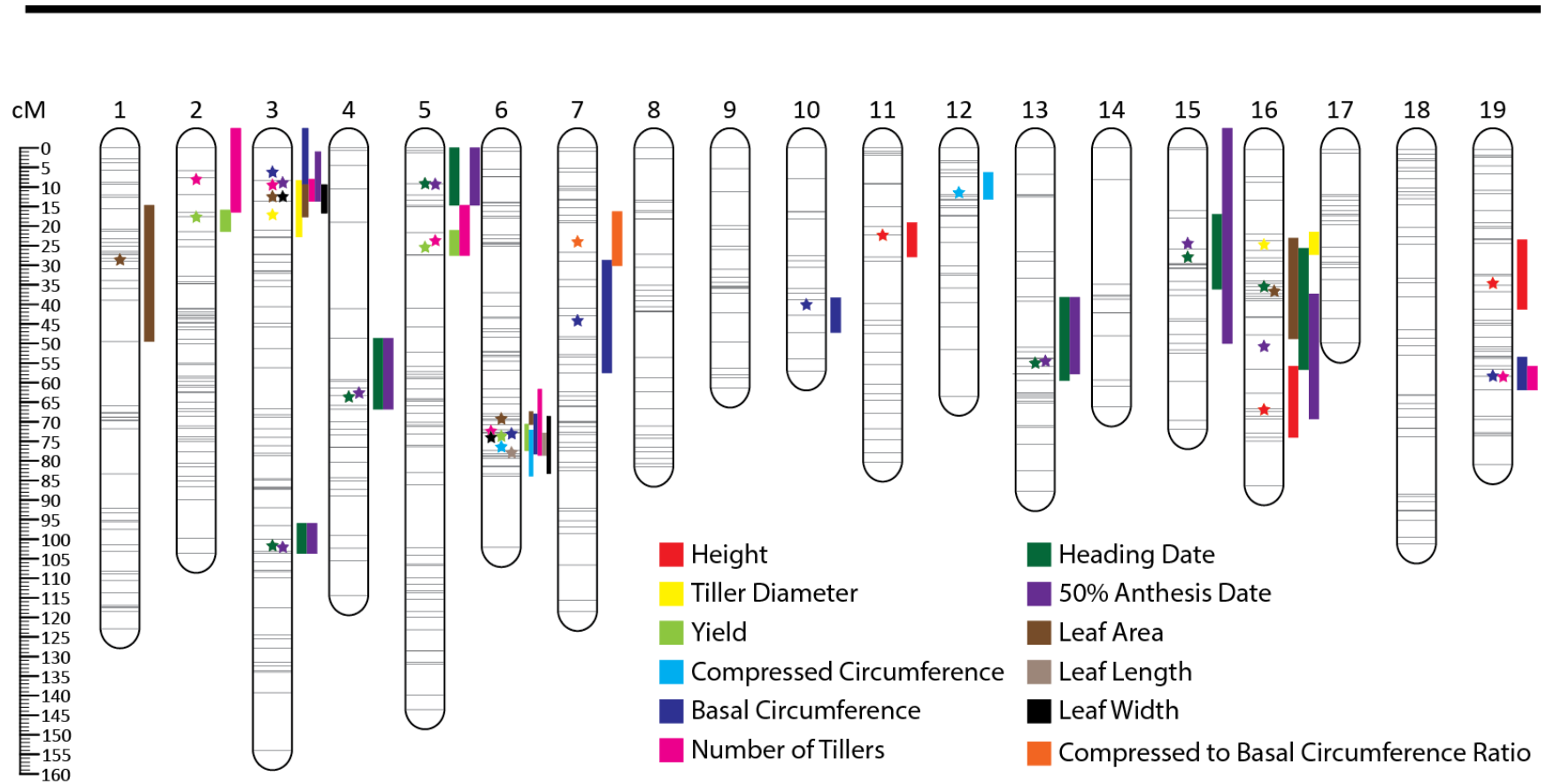
Figure 2.7 Localization of QTLs for 2011 (a) and 2012 (b). The genetic distance is shown on the left in centimorgans, the linkage group numbers are shown at the top, horizontal lines on the linkage groups represent the placement of the genetic markers, stars represent the peak LOD of the QTL, and bars to the right indicate the 2-LOD support interval

Location of 2011 (a), 2012 (b), and 2013 (c) QTL on 19 Miscanthus Linkage Groups
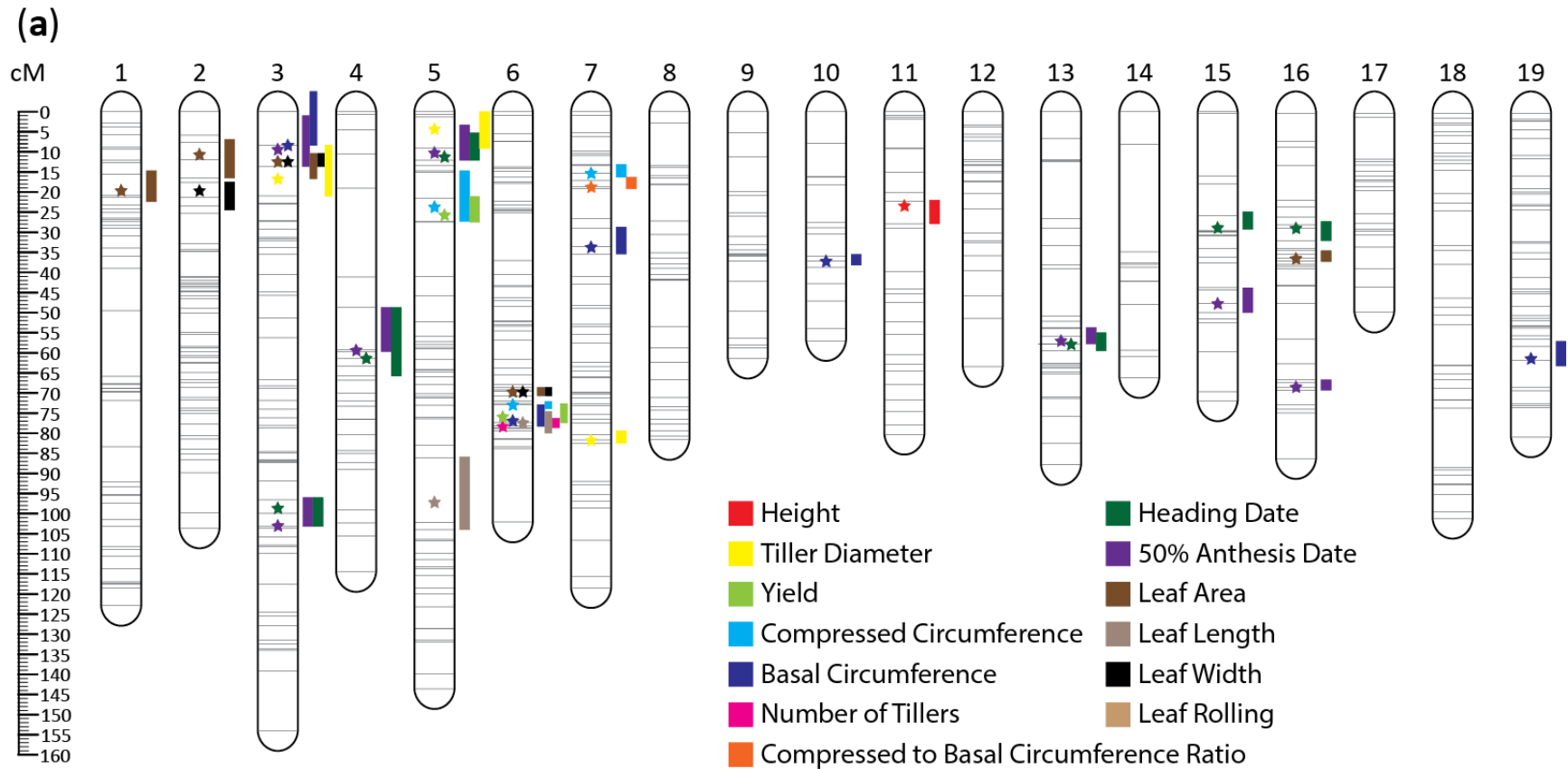
Figure 2.7 (Continued)

Location of 2011 (a), 2012 (b), and 2013 (c) QTL on 19 Miscanthus Linkage Groups

Figure 2.7 (Continued)

Location of 2011 (a), 2012 (b), and 2013 (c) QTL on 19 Miscanthus Linkage Groups

(c)

Legend:
- Height
- Tiller Diameter
- Yield
- Compressed Circumference
- Basal Circumference
- Number of Tillers
- Compressed to Basal Circumference Ratio
- Heading Date
- 50% Anthesis Date
- Leaf Area
- Leaf Length
- Leaf Width
- Leaf Rolling
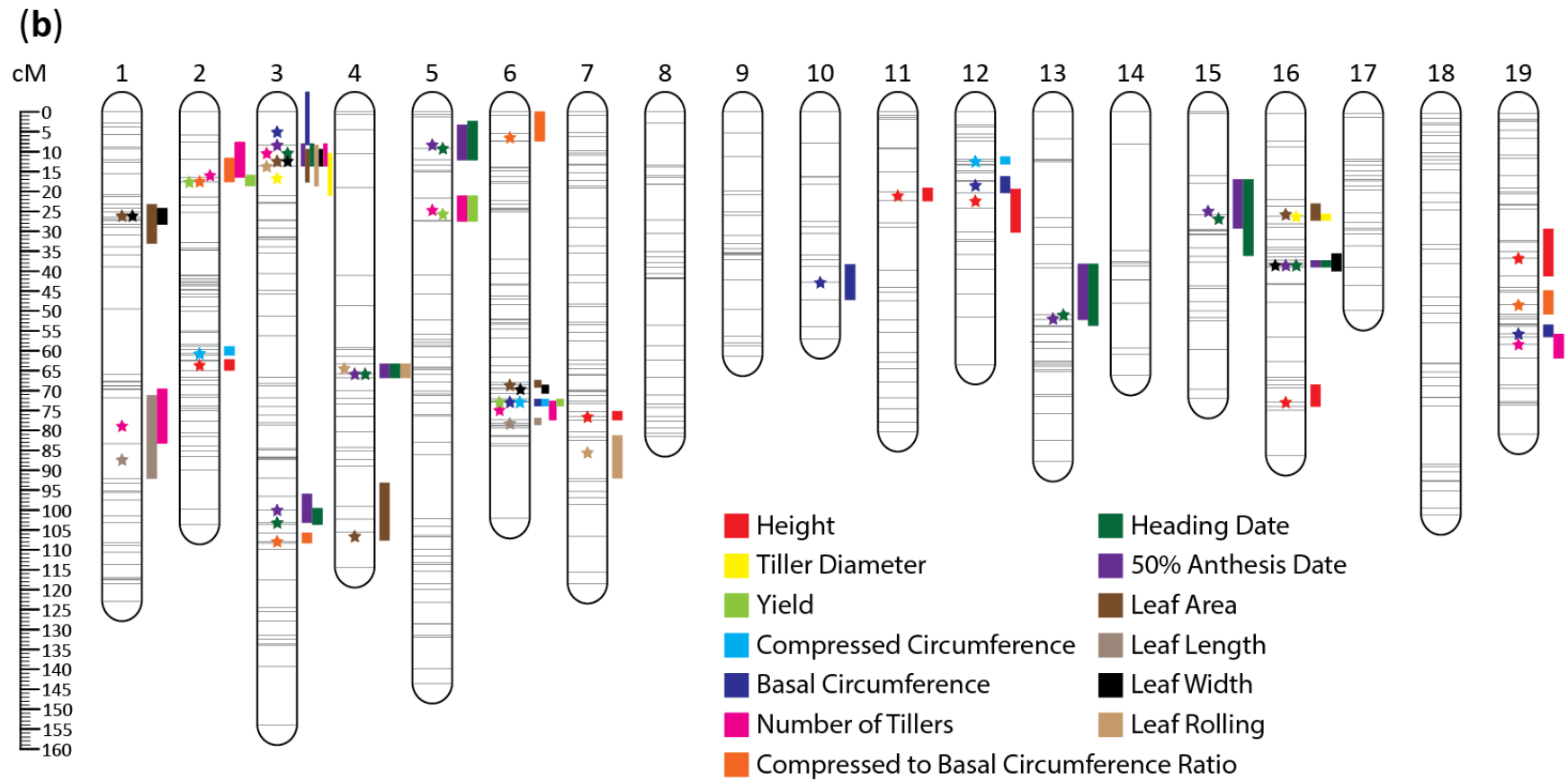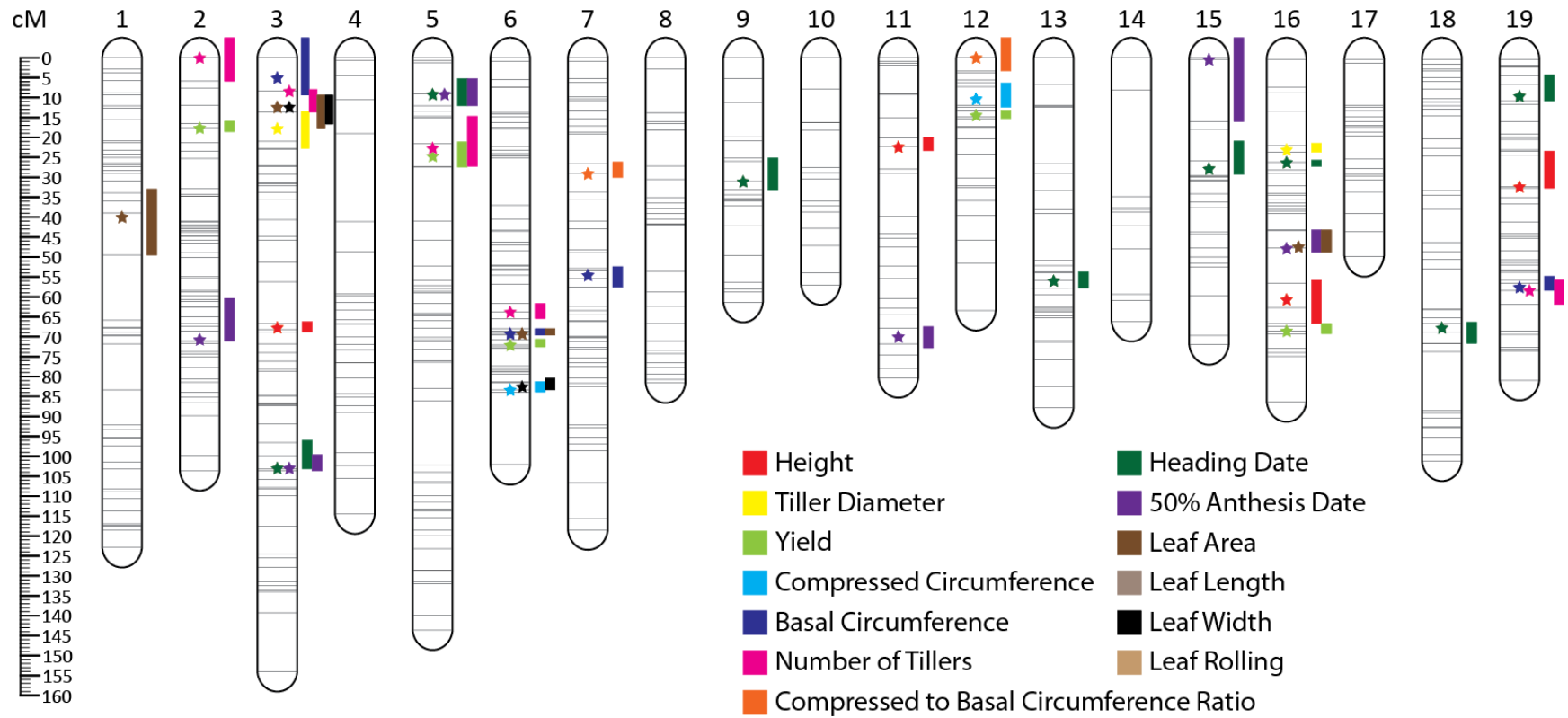
Table 2.1 Plant height (H), tiller diameter (TD), yield (Y), compressed circumference (CC), number of tillers (NT), basal circumference (BC), heading date (HD), 50% anthesis (A), leaf length (LL), average leaf width (ALW), ratio of CC to BC (CCBC), leaf area (LA), and leaf rolling (LR) mean, range, and standard deviation (SD) of the population; means of each parent ('Grosse Fountaine' and 'Undine'); broad sense heritability (H^2); maternal effects (M); and the sum of the percent variation explained by the QTL (ΣPVE) in 2011, 2012, and 2013.

Trait (units)	Year	Mean	Range	SD	'GF'[a]	'UN'[b]	H^2	M[c]	ΣPVE
H (m)	2011	1.5	1.0 - 2.0	0.16	1.9***	1.5	0.52	<0.001	13.1
	2012	1.8	1.3 - 2.3	0.20	2.1***	1.6	0.59	<0.001	47.3
	2013	2.1	1.5 - 2.6	0.19	2.6***	2.0	0.60	0.004	35.6
TD (mm)	2011	6.2	4.5 - 8.9	0.75	7.3***	5.7	0.35	NS[d]	31.6
	2012	4.4	3.1 - 5.8	0.51	5.0***	4.1	0.35	NS	26.2
	2013	4.6	3.4 - 6.2	0.46	5.5***	4.3	0.41	NS	21.5
Y (kg)	2011	1.2	0.1 - 2.1	0.36	1.7***	0.8	0.34	<0.05	37.3
	2012	1.2	0.1 - 2.3	0.40	1.7***	0.7	0.40	<0.01	45.1
	2013	2.1	0.2 - 3.6	0.62	3.0***	1.4	0.53	NS	52.6
CC (cm)	2011	35.8	7.3 - 53.0	7.62	39.0***	28.5	0.32	NS	29.6
	2012	37.6	11.6 - 51.2	6.49	40.9**	31.4	0.40	NS	35.9
	2013	44.9	12.1 - 64.4	8.16	47.7	43.3	0.38	NS	26.9
NT	2011	78.1	6.4 - 165.8	41.80	49.1***	80.7	0.28	NS	14.2
	2012	166.1	22.0 - 262.6	44.51	209.9***	120.0	0.51	<0.05	63.7
	2013	236.3	91.6 - 394.6	54.05	291.9***	181.6	0.49	NS	49.1
BC (cm)	2011	80.2	33.0 - 116.0	13.20	86.8***	64.2	0.51	NS*	48.6
	2012	100.5	39.3 - 138.7	15.24	118.8***	78.7	0.55	NS*	48.5
	2013	121.8	56.7 - 158.9	15.35	142.2***	101.8	0.61	NS*	46.3
HD (days)	2011	8/24	7/22 - 10/7	16.03	8/20***	8/4*	0.80	NS	60.4
	2012	8/25	7/13 - 9/18	11.38	8/24**	8/16*	0.67	NS	53.7
	2013	8/15	7/23 - 9/6	8.82	8/19***	8/4*	0.92	NS	66.8
A (days)	2011	9/8	7/31 - 10/14	14.88	8/31*	8/23*	0.77	NS	62.7
	2012	9/10	8/17 - 10/3	8.44	9/7**	9/2	0.70	<0.05	56.8
	2013	8/29	8/10 - 9/13	6.37	9/2***	8/23	0.90	NS	85.1
LL (cm)	2011	61.5	41.0 - 82.0	6.55	74.2***	55.8	0.43	<0.05	19.7
	2012	59.4	43.9 - 76.7	6.62	69.3***	47.2	0.38	<0.001	21.3
	2013	57.8	40.5 - 85.6	6.91	65.6***	50.1*	0.28	<0.001*	0
ALW (cm)	2011	0.8	0.2 - 1.6	0.34	0.9***	0.5*	0.90	NS	87.2
	2012	0.8	0.3 - 1.7	0.29	0.9***	0.6	0.91	NS	88.5
	2013	0.8	0.3 - 1.6	0.28	0.9***	0.7	0.88	NS	69.4
CCBC	2011	0.5	0.2 - 0.6	0.06	0.4	0.4	0.18	NS	9.5
	2012	0.4	0.3 - 0.5	0.04	0.3*	0.4*	0.22	NS	36.7
	2013	0.4	0.2 - 0.5	0.05	0.3***	0.4	0.22	NS	21.0
LA (cm²)	2011	50.8	13.5 - 118.4	22.15	67.7***	31.7*	0.82	NS	91.7
	2012	49.5	16.4 - 122.2	19.84	65.6***	29.3	0.80	NS	84.4
	2013	51.2	19.4 - 116.7	19.16	66.1***	36.8*	0.73	NS	67.9
LR	2012	7.0	1.3 - 9.0	1.46	7.0	7.6	0.44	<0.05	35.4

[a] Indicate significant difference between means of the parents
[b] Indicate significant difference between variance of parents
[c] Indicate significant difference in variance of the populations separated by maternal lineage
[d] Non-significant
[e] No data
*P-value <0.05, **P-value<0.01, ***P-value<0.001

Table 2.2 Pearson correlation coefficients between selected traits are below the diagonal while genetic correlations are above the diagonal. All Pearson correlations are significant with a P-value <0.05. The 11, 12, and 13 refer to 2011, 2012, and 2013 and denote the year in which the data was collected.

	BC11	BC12	BC13	CC11	CC12	CC13	E11	H11	H13	HD11	HD13	NT13	TD13	Y11	Y12	Y13
Basal Circ. (BC11)	1	0.91	0.91	0.81	0.73	0.71	-0.78	0.25	0.19	0.11	0.12	0.73	0.17	0.71	0.65	0.69
Basal Circ. (BC12)	0.87	1	0.93	0.73	0.82	0.80	-0.77	0.25	0.25	0.11	0.11	0.83	0.20	0.77	0.76	0.78
Basal Circ. (BC13)	0.83	0.84	1	0.78	0.72	0.80	-0.68	0.35	0.29	0.08	0.12	0.68	0.24	0.67	0.70	0.76
Comp. Circ. (CC11)	0.68	0.65	0.61	1	0.87	0.80	-0.62	0.23	0.20	0.11	0.17	0.49	0.20	0.85	0.79	0.76
Comp. Circ. (CC12)	0.60	0.67	0.63	0.68	1	0.86	-0.63	0.15	0.20	0.07	0.10	0.65	0.26	0.87	0.88	0.79
Comp. Circ. (CC13)	0.50	0.58	0.61	0.48	0.51	1	-0.71	0.34	0.32	0.03	0.09	0.63	0.35	0.74	0.83	0.92
Emergence (E11)	-0.52	-0.44	-0.40	-0.45	-0.35	-0.29	1	-0.40	-0.35	-0.07	-0.08	-0.64	-0.22	-0.59	-0.54	-0.53
Height (H11)	0.38	0.39	0.40	0.38	0.39	0.24	-0.35	1	0.84	-0.07	-0.01	0.11	0.61	0.31	0.36	0.42
Height (H13)	0.17	0.26	0.26	0.26	0.31	0.28	-0.15	0.66	1	0.10	0.24	0.04	0.77	0.33	0.42	0.46
Heading Date (HD11)	NS	NS	NS	-0.08	-0.08	NS	0.17	-0.21	NS	1	0.88	0.25	0.08	0.19	0.21	0.25
Heading Date (HD13)	0.10	0.08	0.08	0.09	NS	NS	NS	NS	0.16	0.81	1	0.26	0.15	0.22	0.27	0.29
Number of Tillers (NT13)	0.65	0.70	0.75	0.50	0.54	0.68	-0.30	0.27	0.21	NS	0.17	1	-0.13	0.66	0.68	0.73
Tiller Diameter (TD13)	0.08	0.13	0.15	0.14	0.17	0.26	NS	0.29	0.48	NS	NS	-0.12	1	0.29	0.41	0.42
Yield (Y11)	0.66	0.66	0.63	0.81	0.76	0.44	-0.45	0.55	0.39	NS	0.13	0.53	0.13	1	0.94	0.79
Yield (Y12)	0.60	0.68	0.62	0.68	0.86	0.47	-0.31	0.55	0.44	NS	0.17	0.59	0.20	0.83	1	0.93
Yield (Y13)	0.61	0.68	0.69	0.56	0.59	0.81	-0.32	0.38	0.44	NS	0.19	0.85	0.30	0.59	0.66	1

Table 2.3 QTL results for 2010-2013 height (H), tiller diameter (TD), yield (Y), compressed circumference (CC), basal circumference (BC), compressed circumference to basal circumference ratio (CCBC), number of tillers (NT), heading date (HD), 50% anthesis (A), leaf area (LA), leaf length (LL), leaf width (LW), and leaf rolling (LR).

QTL	Year	LG[a]	cM[b]	LOD[c]	PVE[d]	2 LOD CI[e]	Par[f]	GF[g]	UN[h]	I[i]
H1	2011	11	23.3	5.9	13.1	22 ↔ 27.9	Yes	-0.05	0.21	-0.08
H1	2012	11	21.0	7.3	9.1	19.1 ↔ 22.4	Yes	0.00	0.25	-0.04
H1	2013	11	22.3	7.4	11	20 ↔ 23.3	Yes	-0.01	0.27	-0.02
H2	2012	2	63.6	5.6	6.9	62.2 ↔ 64.9	No	-0.02	-0.17	-0.11
H3	2012	7	76.6	6.3	7.5	75.2 ↔ 77.4	No	0.09	0.19	0.07
H4	2012	12	22.4	5.6	7.3	19.4 ↔ 30.3	Yes	-0.10	-0.19	-0.02
H5	2012	16	72.9	8.1	9.9	68.5 ↔ 74	No	0.05	0.25	0.04
H5	2013	16	60.7	5.9	8.4	55.8 ↔ 66.7	No	0.06	0.22	0.00
H6	2012	19	36.8	4.8	6.6	29.4 ↔ 41.3	No	-0.10	0.06	-0.18
H6	2013	19	32.3	6.2	8.5	23.5 ↔ 32.7	No	-0.06	0.19	-0.12
H7	2013	3	67.7	5.3	7.7	66.2 ↔ 68.9	No	-0.18	0.09	-0.08
TD1	2011	3	16.7	7.3	11.9	8.3 ↔ 21	No	-0.61	0.82	-0.15
TD1	2012	3	16.7	8.5	16.4	10.4 ↔ 21	No	-0.39	0.65	-0.22
TD1	2013	3	17.7	6.2	11.9	13.4 ↔ 22.8	No	-0.43	0.47	0.11
TD2	2011	5	4.3	6.8	11.2	0 ↔ 9.2	No	-0.56	-0.82	0.35
TD3	2011	7	81.7	5.6	8.5	79.4 ↔ 82.5	No	0.58	-0.62	0.39
TD4	2012	16	26.2	5.8	9.8	25.7 ↔ 27.3	No	-0.18	0.60	0.07
TD4	2013	16	23.0	5.4	9.6	21.5 ↔ 23.7	No	-0.09	0.56	0.01
Y1	2011	5	25.6	8.8	14.9	21.1 ↔ 27.5	Yes	-0.23	-0.40	0.11
Y1	2012	5	25.6	11.9	17.6	21.1 ↔ 27.5	Yes	-0.28	-0.54	0.23
Y1	2013	5	24.6	12.6	16	21.1 ↔ 27.4	Yes	-0.44	-0.86	0.22
Y2	2011	6	75.9	13.2	22.4	72.6 ↔ 77.4	Yes	0.50	0.34	-0.18
Y2	2012	6	72.9	14.9	21.7	72.1 ↔ 73.9	Yes	0.66	0.31	-0.32
Y2	2013	6	72.1	14.6	17.9	70.6 ↔ 72.6	Yes	0.77	0.72	-0.40
Y3	2012	2	17.6	4.4	5.8	15.9 ↔ 18.6	No	0.32	-0.19	-0.09
Y3	2013	2	17.6	5.2	5.8	15.9 ↔ 18.6	No	0.41	-0.41	-0.17
Y4	2013	12	14.4	5.8	6.6	13.2 ↔ 15.2	No	-0.51	-0.42	-0.12
Y5	2013	16	68.5	5.6	6.3	66.7 ↔ 69.3	No	0.20	0.61	-0.09
CC1	2011	5	23.6	4.9	8	14.7 ↔ 27.3	Yes	-3.62	-5.50	2.72
CC2	2011	6	72.9	9.5	15	72.1 ↔ 73.9	Yes	7.91	6.49	-3.12
CC2	2012	6	72.9	13.1	20.1	72.1 ↔ 73.9	Yes	9.29	6.55	-4.53
CC2	2013	6	83.3	9.6	15.7	81.3 ↔ 83.9	Yes	8.90	8.73	-5.13
CC3	2011	7	15.3	4.2	6.6	13.1 ↔ 16.3	No	-1.88	-6.03	-2.41
CC4	2012	2	60.7	6.7	9.6	58.9 ↔ 61.2	No	7.39	-0.50	-4.46
CC5	2012	12	12.4	4.4	6.2	11.3 ↔ 13.2	No	-5.50	-1.91	-2.26
CC5	2013	12	10.3	6.7	11.2	6.3 ↔ 12.4	No	-9.61	-6.33	-0.21
BC1	2011	3	8.3	9.3	11.8	END[j] ↔ 8.4	No	-15.06	-2.96	-0.78

Table 2.3 (Continued)

BC1	2012	3	5.0	10.4	13.3	END ↔ 8.3	No	-19.83	-10.95	-3.32
BC1	2013	3	5.0	8.0	12	END ↔ 9.4	No	-18.87	-8.33	-6.93
BC2	2011	6	76.9	9.4	12.1	72.9 ↔ 78.3	No	11.83	10.90	-1.14
BC2	2012	6	72.9	12.0	13.6	72.1 ↔ 73.9	No	15.68	15.05	-11.85
BC2	2013	6	69.2	12.8	17.6	68 ↔ 69.6	No	18.54	16.33	-11.33
BC3	2011	7	33.7	5.1	6.3	28.7 ↔ 35.4	No	-10.21	-3.98	-0.77
BC4	2011	10	37.1	6.4	7.9	35.5 ↔ 38.2	No	-0.01	-12.34	-3.60
BC4	2012	10	42.8	5.0	5.1	38.3 ↔ 47.2	No	4.32	-13.78	-5.39
BC5	2011	19	61.4	8.1	10.5	55.8 ↔ 61.9	No	-12.33	-6.83	-3.17
BC5	2012	19	55.8	8.8	10.1	53.5 ↔ 56.5	No	-16.00	-11.63	-1.32
BC5	2013	19	57.5	7.4	9.6	54.8 ↔ 58.4	No	-16.30	-11.06	-0.29
BC6	2012	12	18.4	5.8	6.4	16.2 ↔ 20.4	No	-16.04	-5.15	2.93
BC7	2013	7	54.5	5.5	7.1	52.4 ↔ 57.5	No	-12.59	-2.65	9.15
CCBC1	2011	7	18.7	4.5	9.5	16.3 ↔ 19.2	No	0.02	-0.07	-0.03
CCBC1	2013	7	29.1	5.6	10.6	26.2 ↔ 30.1	No	0.05	-0.03	0.01
CCBC2	2012	2	17.4	4.4	6.1	11.6 ↔ 17.6	No	0.03	0.01	0.03
CCBC3	2012	3	107.9	4.6	6.7	105.7 ↔ 108.3	No	-0.01	-0.04	0.00
CCBC4	2012	6	6.4	6.8	10.1	0 ↔ 7.4	No	0.04	0.04	0.01
CCBC5	2012	19	48.5	9.0	13.8	44.9 ↔ 50.8	No	0.06	0.02	0.02
CCBC6	2013	12	0.0	5.3	10.4	END ↔ 3.3	No	-0.05	-0.04	-0.01
NT1	2010	12	0.0	5.7	12.6	END ↔ 5.3	No	-1.55	-3.10	-6.68
NT2	2011	6	78.3	7.4	14.2	76.3 ↔ 78.6	Yes	140.56	195.80	-49.51
NT2	2012	6	74.9	14.1	15.8	72.6 ↔ 77.4	Yes	52.16	46.00	-25.01
NT2	2013	6	63.8	10.8	15	61.7 ↔ 65.4	Yes	44.00	66.08	-40.35
NT3	2012	1	78.9	6.8	11.6	69.6 ↔ 83.3	Yes	26.40	52.19	23.29
NT4	2012	2	15.9	8.1	8.5	7.6 ↔ 16.5	Yes	46.32	-23.74	7.20
NT4	2013	2	0.0	4.4	6.3	END ↔ 5.9	Yes	36.16	-30.52	-22.12
NT5	2012	3	10.4	9.8	11.9	8 ↔ 13.7	No	9.39	-49.68	22.28
NT5	2013	3	8.4	6.5	8.4	8 ↔ 13.7	No	-3.44	-58.32	14.58
NT6	2012	5	24.6	6.8	7.2	21.1 ↔ 27.5	Yes	-9.74	-38.90	22.08
NT6	2013	5	22.6	8.7	11.9	14.7 ↔ 27.3	Yes	-13.58	-73.27	2.98
NT7	2012	19	58.4	7.7	8.7	55.8 ↔ 61.9	No	-33.81	-37.90	-5.65
NT7	2013	19	58.4	6.0	7.5	55.8 ↔ 61.9	No	-52.01	-28.07	-19.36
HD1	2010	5	6.6	7.1	12	0 ↔ 14.7	No	-0.81	-0.73	0.29
HD1	2011	5	11.2	21.6	21.9	5.26 ↔ 12.1	No	-7.26	-5.53	1.04
HD1	2012	5	9.2	13.0	15.4	2.3 ↔ 12.1	No	-4.72	-2.57	0.59
HD1	2013	5	9.2	27.8	24.1	5.3 ↔ 12.1	No	-3.71	-3.51	0.78
HD2	2010	16	47.8	9.5	14.5	43.2 ↔ 56.7	No	0.46	1.17	-0.15
HD2	2011	16	28.9	11.4	9.7	27.3 ↔ 32.1	No	0.98	5.73	-1.51
HD2	2012	16	38.5	7.8	8.2	37.3 ↔ 39	No	0.14	3.59	-1.00

Table 2.3 (Continued)

HD2	2013	16	26.2	16.6	10.9	25.7 ↔ 27.2	No	-0.06	3.64	0.07
HD3	2010	19	55.8	4.4	6.8	49.3 ↔ 58.4	Yes	-0.60	-0.37	0.44
HD4	2011	3	98.6	12.4	11.3	96 ↔ 103.2	Yes	6.23	0.99	-0.03
HD4	2012	3	103.2	7.7	8	99.6 ↔ 103.7	Yes	3.99	-0.63	0.27
HD4	2013	3	103.0	13.7	9.8	96 ↔ 103.2	Yes	3.27	-0.35	0.41
HD5	2011	4	61.3	6.0	5.6	48.7 ↔ 65.8	Yes	-1.97	-4.33	1.23
HD5	2012	4	65.8	5.7	5.4	63.3 ↔ 66.8	Yes	-2.57	-2.10	1.65
HD6	2011	13	57.8	7.4	6.1	55 ↔ 59.5	No	-0.63	-4.73	0.18
HD6	2012	13	50.9	4.6	4.5	38.2 ↔ 53.7	No	-0.20	-2.85	-0.21
HD6	2013	13	56.0	5.1	3.2	53.7 ↔ 57.8	No	-0.75	-1.74	-0.70
HD7	2011	15	28.8	6.5	5.8	24.9 ↔ 29.3	No	-0.81	-4.59	0.02
HD7	2012	15	26.8	5.1	5.8	17 ↔ 36.1	No	-1.08	-3.04	-0.39
HD7	2013	15	27.8	10.6	7.9	20.9 ↔ 29.3	No	0.31	-2.85	1.23
HD8	2012	3	10.4	6.1	6.4	8 ↔ 13.7	Yes	-1.82	3.09	1.10
HD9	2013	9	31.0	4.4	2.6	25.2 ↔ 33.1	No	1.11	1.15	0.83
HD10	2013	18	67.7	4.8	3.6	66.3 ↔ 71.7	Yes	-0.61	1.78	0.30
HD11	2013	19	9.6	6.6	4.7	4.4 ↔ 10.9	Yes	0.40	1.40	-1.74
A1	2010	5	9.2	6.6	11.4	0 ↔ 14.7	No	-0.77	-0.93	0.35
A1	2011	5	10.2	20.8	22.9	3.3 ↔ 12.1	No	-6.77	-4.71	1.56
A1	2012	5	8.3	16.5	19.6	3.3 ↔ 12.1	No	-3.61	-2.57	0.86
A1	2013	5	9.2	30.8	34.2	5.3 ↔ 12.1	No	-2.73	-3.26	0.13
A2	2010	16	47.8	12.5	19.6	45.2 ↔ 56.7	No	0.40	1.70	-0.02
A2	2011	16	68.5	13.3	10.8	66.7 ↔ 69.3	No	1.93	5.95	-0.46
A2	2012	16	38.5	9.1	8.8	37.3 ↔ 39	No	0.19	2.93	-0.20
A2	2013	16	47.8	7.9	7	43.2 ↔ 48.8	No	0.28	1.70	0.73
A3	2011	3	9.3	4.7	4.1	1 ↔ 13.7	Yes	-0.71	3.63	2.44
A3	2012	3	8.3	4.9	4.6	8 ↔ 13.7	Yes	-0.81	2.16	0.49
A4	2011	3	103.0	8.9	6.8	96 ↔ 103.2	Yes	5.26	-0.66	-1.30
A4	2012	3	100.0	9.1	9.3	96 ↔ 103.2	Yes	3.12	0.20	-0.04
A4	2013	3	103.0	13.8	15	99.6 ↔ 103.7	Yes	2.52	0.08	0.62
A5	2011	4	59.3	6.2	6.1	48.7 ↔ 59.7	Yes	-1.80	-4.20	0.78
A5	2012	4	65.8	4.6	4	63.3 ↔ 66.8	Yes	-1.82	-0.97	1.36
A6	2011	13	57.0	8.5	8.5	53.7 ↔ 57.8	No	-0.89	-4.87	-0.14
A6	2012	13	51.9	5.3	4.9	38.2 ↔ 52.2	No	-0.27	-2.20	-0.10
A7	2011	15	47.7	4.6	3.5	43.8 ↔ 50	No	-1.72	-3.16	0.77
A7	2012	15	24.9	5.0	5.6	17 ↔ 29.3	No	-0.23	-2.35	0.14
A7	2013	15	0.4	9.6	17.3	END ↔ 16	No	-0.43	-3.01	0.70
A8	2013	2	70.7	6.1	5.1	60.4 ↔ 71.1	No	0.86	1.47	-0.85
A9	2013	11	70.0	7.8	6.5	67.5 ↔ 72.8	No	1.66	0.81	0.11
LA1	2011	1	19.6	10.7	9.5	14.7 ↔ 22.4	Yes	-91.51	-32.26	-8.65

Table 2.3 (Continued)

LA1	2012	1	26.1	7.2	6.7	23.2 ↔ 33	Yes	-97.58	36.57	-12.10
LA1	2013	1	39.9	5.4	5.8	33 ↔ 49.5	Yes	-1.60	-17.37	-6.28
LA2	2011	2	10.6	5.0	3.1	6.9 ↔ 16.5	Yes	-12.92	-9.54	-1.13
LA3	2011	3	12.4	51.2	66.4	10.4 ↔ 16.7	Yes	-55.56	38.62	-20.05
LA3	2012	3	12.4	41.5	55.1	9.4 ↔ 17.7	Yes	-46.42	33.39	-11.67
LA3	2013	3	12.4	34.7	50	9.4 ↔ 17.7	Yes	-42.72	30.40	-11.62
LA4	2011	6	69.6	11.2	9.1	68.6 ↔ 70.7	No	21.62	7.38	-11.20
LA4	2012	6	68.6	9.3	9	67.4 ↔ 69.2	No	17.02	7.37	-13.45
LA4	2013	6	69.2	6.2	6.5	68 ↔ 69.6	No	15.93	4.82	-10.06
LA5	2011	16	36.5	5.7	3.6	34.6 ↔ 37.3	No	0.41	16.42	-1.15
LA5	2012	16	25.7	6.5	5.8	23.1 ↔ 27.3	No	-4.53	16.38	2.96
LA5	2013	16	47.4	5.1	5.6	43.2 ↔ 48.8	No	-0.81	16.47	-0.35
LA6	2012	4	106.6	5.4	7.8	93.2 ↔ 107.6	Yes	-75.69	-8.65	5.86
LL1	2011	5	97.1	4.9	9.3	85.9 ↔ 104	Yes	-0.48	-7.19	2.75
LL2	2011	6	77.4	5.9	10.4	72.9 ↔ 78.3	Yes	4.51	4.37	-6.73
LL2	2012	6	78.3	5.6	10.7	76.9 ↔ 78.6	Yes	3.44	6.32	-5.52
LL3	2012	1	87.3	4.2	10.6	71.2 ↔ 92.1	No	-7.92	-1.62	3.77
LW1	2011	2	19.6	9.1	4.7	17.5 ↔ 24.5	Yes	-0.15	1.26	-0.21
LW2	2011	3	12.4	65.5	79	10.4 ↔ 13.7	No	-0.89	0.66	-0.35
LW2	2012	3	12.4	63.6	75.2	9.4 ↔ 13.7	No	-0.75	0.59	-0.26
LW2	2013	3	12.4	45.9	66	9.4 ↔ 16.7	No	-0.71	0.52	-0.19
LW3	2011	6	69.6	7.0	3.5	68.6 ↔ 70.7	No	0.24	0.05	-0.08
LW3	2012	6	69.6	7.7	3.8	68.6 ↔ 70.7	No	0.19	0.08	-0.13
LW3	2013	6	82.5	4.7	3.4	80.4 ↔ 83.3	No	0.17	0.05	-0.12
LW4	2012	1	26.1	14.5	6.5	24.2 ↔ 28.3	Yes	-1.47	0.49	-0.17
LW5	2012	16	38.5	6.6	3	35.6 ↔ 40	No	-0.03	0.21	0.03
LR1	2012	3	13.7	7.6	13	8.3 ↔ 18.7	Yes	1.25	-0.67	1.49
LR2	2012	4	64.4	7.5	11.5	63.3 ↔ 66.8	Yes	-1.84	-1.10	0.53
LR3	2012	7	85.5	6.6	10.9	81.3 ↔ 92	No	-0.69	-1.80	-0.24

[a] Linkage group
[b] Peak position in centiMorgans
[c] Likelihood of odds
[d] Percentage of phenotypic variance explained by QTL
[e] 2-LOD support confidence interval
[f] QTL for same trait present on paralogous chromosome in either year
[g] Difference in the effect of the alleles inherited from 'Grosse Fountaine'
[h] Difference in the effect of the alleles inherited from 'Undine'
[i] Interaction of alleles
[j] CI goes beyond the last marker on the LG

Table 2.4 QTL distribution across the 19 linkage groups for 2011 (A), 2012 (B), and 2013 (C).

LG	Observed Number of QTL per 10 cM																Exp	X^2
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120	121-130	131-140	141-150	151-160		
A)																		
1	0	1	0	0	0	0	0	0	0	0	0	0	0	-	-	-	0.08	0.01
2	0	2	0	0	0	0	0	0	0	0	0	-	-	-	-	-	0.18	6.49*
3	2	3	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0.44	6.76*
4	0	0	0	0	0	1	1	0	0	0	0	0	-	-	-	-	0.17	0.06
5	1	2	2	0	0	0	0	0	0	1	0	0	0	0	0	-	0.40	2.88
6	0	0	0	0	0	0	2	5	0	0	0	-	-	-	-	-	0.64	201.29***
7	0	2	0	1	0	0	0	0	1	0	0	0	-	-	-	-	0.17	5.42*
8	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-
9	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
10	0	0	0	1	0	0	-	-	-	-	-	-	-	-	-	-	0.17	0.03
11	0	0	1	0	0	0	0	0	-	-	-	-	-	-	-	-	0.13	0.02
12	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
13	0	0	0	0	0	2	0	0	0	-	-	-	-	-	-	-	0.22	5.49*
14	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
15	0	0	1	0	1	0	0	0	-	-	-	-	-	-	-	-	0.25	0.13
16	0	0	1	1	0	0	1	0	0	-	-	-	-	-	-	-	0.25	0.13
17	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-
18	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-
19	0	0	0	0	0	0	1	0	-	-	-	-	-	-	-	-	0.13	0.02

Table 2.4 (Continued)

LG	Observed Number of QTL per 10 cM																Exp	X^2
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120	121-130	131-140	141-150	151-160		
B)																		
1	0	0	2	0	0	0	0	1	1	0	0	0	0	-	-	-	0.31	0.98
2	0	3	0	0	0	0	2	0	0	0	0	-	-	-	-	-	0.45	11.12*
3	2	6	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0.69	851.71***
4	0	0	0	0	0	0	3	0	0	0	1	0	-	-	-	-	0.33	18.82***
5	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	-	0.27	9.46**
6	1	0	0	0	0	0	2	5	0	0	0	-	-	-	-	-	0.73	112.97***
7	0	0	0	0	0	0	0	1	1	0	0	0	-	-	-	-	0.17	0.06
8	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-
9	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
10	0	0	0	0	1	0	-	-	-	-	-	-	-	-	-	-	0.17	0.03
11	0	0	1	0	0	0	0	0	-	-	-	-	-	-	-	-	0.13	0.02
12	0	2	1	0	0	0	0	-	-	-	-	-	-	-	-	-	0.43	1.32
13	0	0	0	0	0	2	0	0	0	-	-	-	-	-	-	-	0.22	5.49*
14	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
15	0	0	2	0	0	0	0	0	-	-	-	-	-	-	-	-	0.25	4.98*
16	0	0	2	3	0	0	0	1	0	-	-	-	-	-	-	-	0.67	4.43
17	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-
18	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-
19	0	0	0	1	1	2	0	0	-	-	-	-	-	-	-	-	0.50	0.33

Table 2.4 (Continued)

LG	Observed Number of QTL per 10 cM																Exp	X^2
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120	121-130	131-140	141-150	151-160		
C)																		
1	0	0	0	1	0	0	0	0	0	0	0	0	0	-	-	-	0.08	0.01
2	1	1	0	0	0	0	0	1	0	0	0	-	-	-	-	-	0.27	0.24
3	2	3	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0.50	7.26*
4	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-
5	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0		0.27	9.46**
6	0	0	0	0	0	0	3	1	2	0	0	-	-	-	-	-	0.55	6.15*
7	0	0	1	0	0	1	0	0	0	0	0	0	-	-	-	-	0.17	0.06
8	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-
9	0	0	0	1	0	0	0	-	-	-	-	-	-	-	-	-	0.14	0.02
10	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-
11	0	0	1	0	0	0	1	0	-	-	-	-	-	-	-	-	0.25	0.13
12	1	2	0	0	0	0	0	-	-	-	-	-	-	-	-	-	0.43	1.32
13	0	0	0	0	0	1	0	0	0	-	-	-	-	-	-	-	0.11	0.01
14	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-
15	1	0	1	0	0	0	0	0	-	-	-	-	-	-	-	-	0.25	0.13
16	0	0	2	0	2	0	2	0	0	-	-	-	-	-	-	-	0.67	7.28**
17	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-
18	0	0	0	0	0	0	1	0	0	0	0	-	-	-	-	-	0.09	0.01
19	1	0	0	1	0	2	0	0	0	-	-	-	-	-	-	-	0.44	0.46

* P-value <0.05, **P-value<0.01, ***P-value<0.001

CHAPTER 3: EXPLORATION OF A *BARLEY STRIPE MOSAIC VIRUS* MEDIATED VIGS SYSTEM IN *MISCANTHUS*

3.1 Abstract

Despite the rapid expansion of genetic information pertaining to the *Miscanthus* genus over the last decade, a rapid, highly efficient functional genomics tool remains elusive. Limited success has been achieved with traditional bombardment and *Agrobacterium*-mediated transformation, but the low efficiencies of these protocols in combination with the genotype specificity and lengthy time requirements make these techniques imperfect. Virus induced gene silencing (VIGS) may present an ideal alternative. This technique is rapid and robust to polyploidy, but it is untested in *Miscanthus*. Here we attempted to adapt an existing VIGS vector with a large monocotyledonous host range for use in *Miscanthus*. *Barley stripe mosaic virus* (BSMV) has been reported to infect over 240 species within the Poaceae family. This vector was previously modified to include a ligation independent cloning (LIC) site to aid in cloning small plant gene fragments into the virus. In theory, during viral replication, double-stranded RNA will trigger the post-transcriptional gene silencing response, resulting in down regulation of viral transcripts and the cloned plant gene. In order for this process to work, the virus must be able to infect *Miscanthus*. Five *Miscanthus*, two maize, and two sorghum genotypes were screened for susceptibility to BSMV to no prevail. Simultaneously, a number of gene fragments were attempted to clone into the LIC site. After extensive trouble shooting, it was determined that a deletion in the original LIC BSMV clone was preventing successful integration of gene fragments. Two strategies are weighed moving forward: test additional genotypes of *Miscanthus* or acquire one or more additional VIGS vectors to screen *Miscanthus* for susceptibility. Neither method is guaranteed to work. The decision comes down to the amount of resources and effort available to dedicate towards developing a VIGS system for *Miscanthus*. Screening additional

genotypes with the current BSMV VIGS vector will require the fewest resources but may have a smaller chance of success.

3.2 Introduction

Miscanthus breeding and genetics have come a long way in recent years. Four genetic maps have been published since 2012 (Kim *et al*., 2012; Ma *et al*., 2012; Swaminathan *et al*., 2012; Liu *et al*., 2015). In addition to providing a valuable tool for QTL studies, these works revealed that *Miscanthus* recently underwent a genome duplication event. In terms of breeding progress, new triploid *M.* ×*giganteus* have begun to be developed and tested (Chae *et al*., 2013; Clark *et al*., 2015), providing much needed genetic diversity within triploid *M.* ×*giganteus*. Deep sequencing of the mapping population described by Swaminathan *et al*. (2012) has been completed and the data is in the process of being compiled into a high density genetic map and a draft of the *Miscanthus* gene space (Kankshita Swaminathan, personal communication). Despite these efforts, no reliable functional genomics tools are available to *Miscanthus* breeders or geneticists. Limited success has been achieved in transforming *M*. *sinensis* through particle bombardment of callus tissue (Wang *et al*., 2011), but no further reports have been published. Efforts to reproduce the published transformation protocol failed to yield any transgenic plants (John Juvik, unpublished data), possibly as result of genotype specific responses to transformation. *Agrobacterium*-mediated transformation of *M*. *sinensis* has also been reported (Hwang *et al*., 2014), but this method also suffers from low transformation efficiency and strong genotype specificity. As more genetic information continues to be developed, it is imperative that researchers have a quick, reliable, and affordable functional genomics tool. Virus induced gene silencing (VIGS) is a promising candidate to provide such a tool.

Analogous to RNA interference in animals, post-transcriptional gene silencing (PTGS) is a process in which RNA molecules inhibit gene expression in plants. PTGS plays a key role in a plants innate defense against viral infection, and it is this mechanism that VIGS relies on to silence genes in a sequence specific manner (Baulcombe, 1999). Prior to explaining VIGS, the mechanisms behind PTGS must be understood. While PTGS is not limited to defense against viral infection, the ensuing discussion will focus on the viral defense aspect of PTGS. Our understanding of this process has progressed extensively since it was first observed in transgenic petunia flowers in the late 1980s (Napoli *et al*., 1990). PTGS is triggered when double-stranded RNA (dsRNA) is recognized and cleaved by Dicer-like (DCL) ribonucleases (Figure 3.1; Xie *et al*., 2004). Four DCL genes have been identified in *Arabidopsis* (Blevins *et al*., 2006). While DCL4 appears to be the predominant ribonuclease involved in the recognition of virus-derived dsRNA, DCL2-4 all cleave virus-derived dsRNAs. Each DCL ribonuclease produces a characteristic short-interfering RNA (siRNA). DCL2-4 produce 22 nucleotide (nt), 24 nt, and 21 nt primary siRNAs, respectively (Bouche *et al*., 2006; Brodersen & Voinnet, 2006; Vaucheret, 2006). DCL1 plays only a minor role, if any, in processing virus-derived dsRNA (Bouché *et al*., 2006). DCL ribonucleases initiate PTGS by recognizing and cleaving dsRNA into siRNA.

Endogenous plant RNA-dependent RNA polymerases (RDRs) are essential in amplifying the virus-induced PTGS signal. RDRs synthesize long, dsRNA from single stranded viral RNA exhibiting perfect homology to primary siRNAs. DCL ribonucleases then convert the long, dsRNAs to secondary siRNAs, resulting in siRNAs that cover the entire viral genome. In fact, dsRNAs produced from RDRs are the preferred substrate for DCL ribonucleases (Kasschau *et al*., 2007). Although siRNAs are abundant and cover the entire genome of the virus, the distribution of siRNAs across the viral genome is uneven. Often the siRNA cover both positive and negative

sense, but occasionally the copy number of siRNAs are biased towards positive sense copies (Donaire *et al.*, 2009; Qi *et al.*, 2009). *Arabidopsis* contains six RDR genes (Wassenegger & Krczal, 2006), but only three have been shown to be functional (Llave, 2010). There are differences in the specificity of the RDR genes to certain viral genomes, but little is known about the underlying mechanisms. It has been postulated that the different specificities are related to the susceptibility of certain RDRs to virus-encoded silencing suppressors (Díaz-Pendón & Ding, 2008). RDRs work in conjunction with DCL ribonucleases to amplify the number of siRNAs homologous to virus encoded genes.

Both primary and secondary siRNAs are loaded into RNA-induced silencing complexes (RISC) to regulate gene expression (Hutvagner & Simard, 2008). At minimum, active RISC contain a cleavage-competent argonaute (AGO) protein and a siRNA (Miyoshi *et al.*, 2005; Rivas *et al.*, 2005). *Arabidopsis* contains ten paralogous AGO proteins. Some of these appear to act redundantly while others are involved in different PTGS pathways (Brodersen & Voinnet, 2006). Homology based degradation of mRNA occurs through endonucleolytic cleavage by AGO, resulting in greatly reduced translation (Brodersen *et al.*, 2008). To summarize, PTGS is initiated when DCL ribonucleases recognize and cleave dsRNA into siRNA. Additional copies of siRNA are then made by DCL ribonucleases through the synthesis of dsRNA from ssRNA by RDRs. These siRNA are then incorporated into RISC, and mRNA homologous to the loaded siRNA is cleaved by AGO, rendering the mRNA nonfunctional.

Researchers can take advantage of this defense mechanism by introducing a fragment of a plant gene into the genome of a virus. The components of PTGS cannot differentiate between the viral genome and the transgenically introduced fragment; therefore, the plant gene fragment is cleaved, incorporated into RISC, and RNA showing homology is degraded, resulting in a

reduction of both viral RNA and the corresponding plant RNA (Figure 3.2). It is this process that

is coined VIGS. It is unknown how RDRs differentiate single-stranded viral RNA from

endogenous single-stranded RNA (ssRNA), but it is known that RDRs cannot distinguish RNAs

with or without a 5' cap or poly-(A) tail (Curaba & Chen, 2008). Because endogenous genes are

protected from RDRs, secondary amplification of siRNAs are limited to the introduced fragment

and do not cover the entire mRNA of the endogenous gene targeted for silencing (Vaistij *et al*.,

2002; Kościańska *et al*., 2005; Miki *et al*., 2005). The fact that endogenous genes are protected

from RDRs is important in maintaining the specificity of the gene silencing.

VIGS was first ascertained experimentally using *Tobacco mosaic virus* to introduce a

fragment of the phytoene desaturase (PDS) gene into *Nicotiana benthamiana*. Plants infected

with the modified viral vector showed extensive photobleaching of the leaves and contained only

a small fraction of the PDS mRNA present in control plants (Kumagai *et al*., 1995). Many

additional viral vectors have been developed for VIGS since this inaugural experiment, and

improvements have been made to VIGS systems to make them more broadly applicable, efficient,

and affordable (Senthil-Kumar & Mysore, 2011; Lange *et al*., 2013).

The first step in developing a VIGS system is genetically engineering a virus to include a

cloning site that allows easy integration of plant gene fragments. While the majority of plant

virus genomes are positive-sense, single-stranded RNA, both RNA and DNA viruses have been

successfully used to induce gene specific silencing (Purkayastha & Dasgupta, 2009). In order for

a virus to be appropriate for use in a VIGS system, it cannot have any potent anti-silencing genes

or those genes must be knocked out (Anandalakshmi *et al*., 1998; Kasschau & Carrington, 1998).

Ideally, viral vectors would cause slight to mild symptoms on the host plant but build up to high

levels. This would limit confusion in regards to if the source of the phenotypic change was due to

the virus or the gene that was knocked out and cause more complete silencing. One way to potentially reduce viral symptoms while increasing infectivity would be through mutagenesis and shuffling of the viral genome. Toth *et al*. (2002) were able to use mutagenesis of the movement protein of the *Tobacco mosaic virus* to improve the spread of the virus within the plant. Further, viruses that are not transmitted by insects or seed would be easier to control and lead to fewer regulations (Ramanna *et al*., 2013). Lastly, viruses with a large host range would be the most useful. In total, more than 30 viruses have shown potential utility as vectors in a VIGS system (Yuan *et al*., 2011). Considering the time it takes to develop a VIGS system compared to the time it takes to apply an extant system, researchers wishing to incorporate this technique into their research or breeding programs should first seek collaboration with scientists whom have published VIGS vectors.

Once an appropriate vector is acquired or created, cloning the gene fragment of interest into the vector is often quite simple. Although the exact process depends on the type of cloning site, it usually includes PCR amplification, restriction enzymes, and ligases. Silencing can be achieved by inserts as short as 23 nucleotides (Thomas *et al*., 2001), but sequences between 200-350 nucleotides provide the most stable and complete silencing (Burch-Smith *et al*., 2004; Liu & Page, 2008; Zhang *et al*., 2010). When choosing a fragment to clone into the vector, it is important to avoid introns (Ruiz *et al*., 1998). While some studies recommend inserting the antisense fragment (Lee *et al*., 2012), others have been unable to detect a difference in the effectiveness of either sense or antisense inserts (Scofield *et al*., 2005). Also, the uniqueness of the fragment should be considered. If the goal is to silence one particular gene, then the sequence of the gene fragment should be unique. On the other hand, entire gene families can be silenced if the sequence of the gene fragment is unique to the family but not unique to any particular family

member (Lu *et al*., 2003). Down regulation has been observed with fragment similarity as low as 80% (Burton *et al*., 2000), which can be advantageous in crops lacking a reference genome. Software has been developed that predicts the suitability of fragments and specificity of fragments prior to performing any experimentation (Xu *et al*., 2006).

Subsequent to introducing a plant gene fragment into the viral genome, the recombinant virus needs to be inoculated into the plant of interest. The first VIGS vectors required costly in vitro transcription (Janda *et al*., 1987; Kumagai *et al*., 1995). Many of the modern VIGS vectors utilize *Agrobacterium*-mediated inoculation. The full length of the viral cDNA is cloned into the T-DNA region of the Ti plasmid between a plant active promoter and a transcription terminator (Turpen *et al*., 1993). An intermediate host can be utilized for plants that are recalcitrant to *Agrobacterium* transformation. Generally, *N. benthamiana* is utilized as it is highly susceptible to *Agrobacterium tumefaciens* and viral infections (Senthil-Kumar & Mysore, 2011). Sap from infected *N. benthamiana* can be rub-inoculated onto the leaves of the desired host. At this point the infected sap contains replication-competent, recombinant virions. Other methods of inoculation do exist but are omitted due to their infrequent use.

When it comes to silencing genes with viral vectors, two factors help ensure relatively uniform silencing. First, viruses have means of spreading within the plant. This movement can be broken into three stages: intracellular movement, intercellular movement through plasmodesmata, and systemic movement through the vascular system (Figure 3.3; Harries & Ding, 2011). There is no clear consensus on how viruses move within a cell. There is evidence that microtubules (Boyko *et al*., 2007) and microfilaments (Liu *et al*., 2005) are involved with movement in association with viral movement proteins (MPs), but there is also evidence that contradicts this (Gillespie *et al*., 2002; Hofmann *et al*., 2009). The MPs are also involved with

intercellular movement as they have been shown to modify the size exclusion limit of the plasmodesmata (Wolf *et al*., 1989). Systemic movement mainly occurs through the phloem, but some viruses utilize the xylem for long distance transport (Harries & Ding, 2011). Comparatively little is known about this long distance transport, but there is some evidence that plants have defense proteins located in the phloem to limit viral movement (Chisholm *et al*., 2000). As expected some viruses have evolved means to overcome these plant defense mechanisms (Decroocq *et al*., 2009). It is important to note that most viruses are prevented from invading meristematic tissue (Matthews, 1991). As the virus spreads throughout a plant, RNAi mechanisms are triggered in a greater number of cells, resulting in more uniform gene silencing.

The second factor that helps ensure relatively uniform silencing is the transport of a silencing signal. This too can be broken into three stages: short-range silencing, extensive local silencing, and systemic silencing (Kalantidis *et al*., 2008). Without the amplification of the silencing signal by RDRs, gene silencing is limited to a thin layer of cells (10-15 cells) surrounding the exogenous dsRNA (Himber *et al*., 2003). It is believed that siRNA travel through plasmodesmata (Voinnet *et al*., 1998) similar to soluble proteins between 27 and 54 kDa (Kobayashi & Zambryski, 2007). Extensive local silencing is characteristic of sink organs and relies on the amplification of 21 nt siRNA (Himber *et al*., 2003). It has been observed that silencing in the source tissue remains limited to the 10-15 cell range even though the amplified signal is phloem-transported throughout the plant. It is unclear on how the silencing signal moves through the sink tissue without causing silencing (Tournier *et al*., 2006). The systemic spread of silencing was recognized shortly after VIGS was introduced (Palauqui *et al*., 1997; Voinnet & Baulcombe, 1997). Systemic silencing of viral RNA appears to be caused by siRNA traveling by phloem. Some forms of silencing, such as the silencing of transgenes, result in

permanent silencing through DNA methylation, but virus-mediated silencing does not appear to result from changes to the host DNA (Brosnan & Voinnet, 2011).

VIGS has quickly become one of the most widely used functional genomics tools in plant research (Scofield & Nelson, 2009; Di Stilio, 2011). It holds many advantages over alternative functional genomics tools. First and foremost, it is easy and rapid. Gene fragments can be cloned into vectors in a day or two, and the phenotype generally arises within one or two weeks after inoculation, allowing entire experiments to be completed within a month (Lu *et al*., 2003). This technique can be used in lieu of producing stable transformants, particularly in species where transformation protocols have yet to be established, or as a way to prescreen genes prior to undertaking lengthy transformation experiments. As alluded to previously, VIGS can be used to silence individual genes or gene families and is particularly useful in polyploid plants. Finally, genes that are essential for plant growth and development can be studied using VIGS. In traditional transformation, mutations in essential genes would cause cells to die prior to the regeneration of transformed plants. Peele and colleagues (2001) were able to silence two essential genes, one encoding a subunit of magnesium chelatase and the other encoding a proliferating cell nuclear antigen.

To the author's knowledge, VIGS has never been attempted on any *Miscanthus* species although they appear to be perfect candidates as they are recalcitrant to transformation, recently underwent a genome duplication event, and have limited genomic sequence data. Since there are no published reports of any of the available VIGS vectors being tested on *Miscanthus*, the first step is identifying candidate vectors that have the highest probability of infecting *Miscanthus*. The best candidates are vectors that have a large host range, particularly within

monocotyledonous plants. The two most widely used monocot viruses are *Barley stripe mosaic virus* (BSMV) and *Brome mosaic virus* (BMV; Scofield & Nelson, 2009).

BSMV was first developed and tested as a VIGS vector in *Hordeum vulgare* (barley), representing the first VIGS in any monocot species (Holzberg *et al*., 2002). Modifications of this vector have since been used in *Triticum aestivum* (wheat; Scofield *et al*., 2005), *Haynaldia villosa* (Wang *et al*., 2010), *Brachypodium distachyon* (Yuan *et al*., 2011), and the dicot *N. benthamiana* (Yuan *et al*., 2011), among others. Furthermore, BSMV is known to infect over 240 species within Poaceae, including maize (*Zea mays*), rice (*Oryza sativa*), and oat (*Avena sativa* & *A. strigosa*; Jackson & Lane, 1981). Considering the exceptionally large host range, including close relatives of *Miscanthus*, and its long history as a vector for VIGS, BSMV is the ideal vector to apply to *Miscanthus*.

Barley stripe mosaic was first speculated to be caused by a virus in 1924 (Jackson & Lane, 1981). BSMV is the type member of the Hordeivirus genus. This genus is characterized by individually encapsidated, tripartite, positive-sense RNA genomes with a hordei-like triple movement gene block (Jackson *et al*., 2009). The three short rigid rods that encapsulate the α, β, and γ RNAs vary in length according to the size of the nucleic acid within. Symptoms can vary depending on the host and strain, but generally BSMV elicits mild mosaic symptoms. Transmission of the virus is either through seed or mechanical plant to plant spread (Jackson *et al*., 1989). In field situations, this spread can occur rapidly, leading to heavy yield losses. The spread of the virus through airborne pollen has been suggested, but it appears to not play a role in the dispersal of the disease (Slack *et al*., 1975). Fortunately, this disease can be effectively controlled by eliminating infected seed prior to planting through simple diagnostic screens as it cannot survive in the absence of living plants for long periods of time (Carroll, 1986).

Although genomic sequencing of BSMV revealed more open reading frames, the three genomic RNAs encode seven major proteins (Figure 3.4; Gustafson & Armour, 1986; Gustafson *et al*., 1987). All three RNAs are necessary for plant infection, but the coat protein, located on the β-RNA, is unnecessary for the virus to spread systemically throughout an infected plant; therefore, the RNAs can travel cell to cell without the protective capsid (Petty & Jackson, 1990). All three movement proteins are necessary for local and systemic spread. These proteins are also located on the β-RNA (Lawrence & Jackson, 2001). The γ-RNA contains two proteins γa and γb. The former interacts with αa to form the RNA-dependent-RNA-polymerase necessary for replication (Jackson *et al*., 2009). There is evidence suggesting that replication occurs in membrane-associated viroplasms attached to chloroplasts, which would make sense considering the mosaic symptoms (McMullen *et al*., 1978; Lin & Langenberg, 1985). The γb protein is involved in suppression of the RNAi response, among other things (Yelina *et al*., 2002; Bragg & Jackson, 2004). Despite this function, it has been observed that the effectiveness of VIGS is decreased when this gene is knocked out (Yuan *et al*., 2011).

The BSMV VIGS tool has begun to be applied to study genes other than the traditional resistance genes investigated by plant pathologists. Oikawa *et al*. (2007) utilized BSMV VIGS to investigate cell wall synthesis in barley. In another barley study it was utilized to investigate inorganic phosphorus uptake (Pacak *et al*., 2010). Beyond silencing genes in leaf tissue, Bennypaul *et al*. (2012) successfully silenced a root specific gene. Gene silencing in floral tissue has also been reported (Lee *et al*., 2012). Due to the biology of BSMV, there are even reports of silencing continuing beyond the initial generation (Bruun-Rasmussen *et al*., 2007). Small proteins have also been expressed using BSMV in what is called virus-mediated over expression or VOX (Haupt *et al*., 2001; Lawrence & Jackson, 2001). Unfortunately, the stability of the

construct decreases with increasing protein size, so expression is not always uniform or long lasting. Lee *et al*. (2012) reported that a group was working on separating the γa and γb genes to separate RNA molecules. In theory, the reduced size of the RNA molecule should allow for larger proteins to be expressed without any penalty to the fitness of the virus. To the author's knowledge, this has yet to be confirmed in practice.

Thanks to Professor Dawei Li of China Agricultural University, Beijing and Professor Andrew Jackson of the University of California, Berkeley the BSMV VIGS vector published by Yuan *et al*. (2011) was acquired. This vector is designed for *Agrobacterium* infiltration and contains a ligation independent cloning (LIC) site to improve the efficiency of cloning host gene fragments into the vector. To broaden the applicability of this vector beyond species susceptible to *Agrobacterium* infection, *N. benthamiana* is used for initial inoculation. Sap extracted from infected *N. benthamiana* can then be used to introduce the recombinant viral vector into the plant of interest.

Yuan *et al*. (2011) can be referenced for a detailed explanation of the BSMV vector construction. Briefly, full-length α, β, and γ cDNAs were amplified from the ND18 strain of BSMV (Petty *et al*., 1989) and inserted into the pCass4-Rz (Annamalai & Rao, 2005) *Agrobacterium* plasmid. A 21 nt LIC site containing the *Apa*I restriction enzyme sequence (GGGCCC) was added to the 3' end of the γb gene. LIC relies on T4 DNA polymerase, which has both polymerization and exonuclease activity, to create complimentary overhangs between the vector and the insert. *E. coli* then repairs the remaining nicks between the joined vector and gene fragment during transformation, creating scarless recombinant plasmids (LIC, 2014).

The objectives of this chapter are to identify a VIGS vector that is suitable for gene silencing in *Miscanthus* and then optimize the protocol to ensure maximum gene silencing. First,

Miscanthus will be challenged by BSMV to determine if and to what extent the virus can infect *Miscanthus* cultivars. If successful, PDS, a commonly used marker gene to test the efficiency of VIGS, will be used to evaluate the ability of BSMV to induce gene silencing in *Miscanthus*. PDS is an important gene in the carotenoid biosynthesis pathway. Knockdown of PDS expression results in photobleaching of leaves due to the lack of photoprotective carotenoid pigments (Misawa *et al*., 1993). If successful, this will provide the first reliable functional genomics tool in *Miscanthus*.

3.3 Material and methods

Plant material

N. benthamiana accession PI 555478 and PI 555684 were acquired from the United States Department of Agriculture's (USDA) National Genetic Resources Program. Seeds were germinated and grown in LC1 Sunshine Mix (Sun Gro Horticulture, Agawam, MA) in a growth chamber with 14 hrs light and $25/20^oC$ day/night temperature. Both 'Black Hulless' and 'Golden Promise' varieties of barley were used as positive controls. *Miscanthus* seedlings and clonal divisions were tested for susceptibility to BSMV. Seed collected from a controlled cross between *M. sinensis* 'Grosse Fountaine' and *M. sinensis* 'Undine' as well as seed harvested from open pollinated *M. sinensis* 'Silberturm' were planted at a depth of 1-2 mm in Metro-Mix 510 (Sun Gro Horticulture) in six inch pots and sprinkled with ferrous sulfate to reduce the pH and provide sufficient iron for normal plant growth and development. The pots were grown in the Turner Hall Greenhouse, which was maintained between 22.2-29.4°C with supplemental light (threshold of 600 W/m2). Divisions of *M. sacchariflorus* 'Golf Course,' *M. sinensis* 'Grosse Fountaine,' and *M. ×giganteus* 'Illinois' were planted in 36-cell plug trays in Metro-Mix 510 and sprinkled with one tablespoon ferrous sulfate per tray. These plugs were grown in the aforementioned greenhouse.

One week prior to inoculation, *Miscanthus* plugs were transplanted to four inch round pots containing the same potting mix and moved to a growth chamber with the previously described conditions. In addition to *Miscanthus*, maize and sorghum were also tested. B73 and MO17 were tested for maize, and K037 Camjin (PI 533839) and BTx623 (PI 564163) varieties were tested for sorghum. Both maize and sorghum were grown in six inch pots with MetroMix 510 in the same greenhouse as previously mentioned.

Agroinfiltration of *N. benthamiana*

In order to produce infective virions, all three viral RNAs must be present in one plant cell. This means that three separate cultures of *A. tumefaciens* must be prepared in order to inoculate *N. benthamiana*, each culture containing a different plasmid. The α and β containing plasmids do not change throughout the experiment. Only the γ containing plasmid changes as there are two versions of this plasmid. The original version is unmodified while the γLIC version contains the LIC site. Initial screening of plants was conducted with an unmodified γ genome segment in order to investigate the ability of the virus to infect *Miscanthus*, sorghum, and maize under controlled conditions.

The process of infecting *N. benthamiana* is the same whether the LIC site is absent, empty, or contains a plant gene fragment. Single colonies of recombinant *A. tumefaciens* strain EHA105 containing α, β, and γ plasmids were picked from streak plates containing rifampicin (25 μg/ml) and kanamycin (100 μg/ml). These colonies were cultured overnight at 28°C in three ml of liquid LB media (Table A.2) containing the same antibiotics. 0.5 ml starter culture was then used to inoculate 50 ml of liquid LB media containing rifampicin and kanamycin, continuing to grow the *A. tumefaciens* clones separately based on the plasmid that they contained. The bacteria were allowed to grow for 10-12 hours at 28°C. After the required amount of time,

the three cultures were centrifuged at 2,200 g for 10 minutes and resuspended in infiltration

buffer (10 mM MgCl$_2$, 10 mM MES, 0.1 mM acetosyringone, pH 5.2) to an OD$_{600}$ of either

~0.07 or ~0.7. Once the proper optical density was achieved, the mixtures of bacteria and

infiltration buffer were allowed to sit at room temperature in the dark for three hours. Equal parts

of the α, β, and γ agroinfiltration suspensions were combined to agroinfiltrate *N. benthamiana*

plants at the four to twelve leaf stage. A 1-ml needleless syringe was used to agroinfiltrate the

leaves (Figure 3.5). The leaves directly above the cotyledons as well as more mature leaves were

infiltrated. Five to twelve days post inoculation the infiltrated leaves and the leaves immediately

above were harvested and ground in extraction buffer (20 mM sodium phosphate buffer, pH 7.2,

containing 1% 600 mesh carborundum). The extracted sap was either frozen at -80°C for later

inoculation of *Miscanthus*, sorghum, or maize or used immediately. Immediate use was

preferable.

Miscanthus, sorghum, and maize inoculation

Miscanthus, sorghum, and maize inoculations were attempted by rubbing the extracted

sap onto leaves by various means, at various developmental states, and after various treatments.

Generally, the leaf immediately above the cotyledon of young seedlings at the 3-4 leaf stage was

inoculated (Figure 3.6) although older plants (6-10 leaf stage seedlings and clonal propagates of

Miscanthus) and younger leaves were also attempted. Further modifications that were attempted

included shaking additional carborundum onto the plant prior to rub inoculation and exposing

plants to a 24 hour dark period prior to rub inoculation. The sap was either rubbed on by hand,

with a cotton ball, or a glass rod. Shortly after inoculation the leaves were rinsed with water to

remove any residual sap. Ten to fourteen days post-inoculation, plants were investigated visually

for symptoms and the presence of the virus was tested for using RT-PCR with primers specific to

BSMV. Barley plants were used as positive controls when seedlings were available and sap extracted from control *N. benthamiana* plants was used as a negative control.

RT-PCR

RNA extractions were performed using the RNeasy Plant Mini Kit (QIAGEN Inc., Valencia, CA) according to the manufacturer's protocol. A Nanodrop 1000 (Thermo Fisher Scientific Inc., Waltham, MA) was used to ensure only high quality RNA was used. First strand cDNA synthesis was performed using SuperScript III First-Strand Synthesis SuperMix (Life Technologies, Carlsbad, CA) according to the manufacturer's protocol. PCR was performed using Taq 5X Master Mix (New England BioLabs Inc., Ipswich, MA) in a MJ Research PTC-100 thermal cycler (MJ Research, Inc., St. Bruno, Quebec, Canada). The resulting PCR product was ran on a 1.2% agarose gel alongside Bioline's HyperLadder™ 1kb DNA ladder in a Mupid-exU submarine electrophoresis system (Mupid CO., LTD., Tokyo, Japan) containing 1X TAE buffer. Gels were imaged using a Molecular Imager® Gel Doc™ XR+ (Bio-Rad Laboratories, Inc., Hercules, CA).

Primer design

Strategies for primer development varied depending on the sequence information that was available. In the case of BSMV, sorghum, and maize where sequence information was available, NCBI's Primer-BLAST tool was utilized to design primers. The primer pair specificity checking parameter was employed to ensure the primers were specific to the sequence of interest. The primers that were designed from this program were then run through three additional screening steps. Individual primers were screened for PCR suitability using Sequence Manipulation Suite: PCR Primer Stats (Stothard, 2000), and Life Technologies' Multiple Primer Analyzer was utilized to screen primer pairs. Primer pairs that passed both screening steps were

input into New England BioLabs Tm Calculator to ensure that the primer pairs had annealing temperatures within 5°C of each other. Primers needed to amplify *Miscanthus* sequences were designed from conserved sequences between maize, wheat, and sorghum using GEMI software (Sobhy & Colson, 2012). Primers suggested by the GEMI software underwent the same three step screening process as previously described. In some instances, primers that would amplify an appropriately sized fragment and pass all screening steps could not be designed from conserved sequences. In these cases, degenerate primers with the fewest ambiguous nucleotides were used. The PCR product was gel purified with the QIAquick PCR Purification Kit (QIAGEN Inc.), and the DNA was sent to the Roy J. Carver Biotechnology Center for sequencing. The *Miscanthus* sequence was then input into NCBI's Primer-BLAST tool to design *Miscanthus* specific primers. The primer pair most suitable for PCR based on the parameters output by the three screening programs was ordered from Life Technologies.

Ligation independent cloning

In order for the LIC cloning to work, additional fusions have to be added to both the forward and reverse primers. This particular LIC system requires 5'-AAGGAAGTTTAA-3' fused to the 5' end of the forward primer and 5'-AACCACCACCACCGT-3' fused to the 5' end of the reverse primer. The PCR template can either be cDNA or genomic DNA as long as the amplification product does not span an intron. PCR product resulting from amplification with gene specific primers with the additional fusions was purified with the QIAquick PCR Purification Kit (QIAGEN Inc.) after the product was run on a 0.6% agarose gel. In order to generate sticky ends, the purified PCR product was treated with T4 DNA polymerase (New England Biolabs Inc.) at room temperature in 1X reaction buffer containing 5 mM dATP for 30 minutes. The DNA polymerase was inactivated by a ten minute heat treatment at 75°C.

The γLIC vector plasmid was purified from *Agrobacterium* using the QIAprep Spin Miniprep Kit (QIAGEN Inc.) and linearized with *Apa*I (New England Biolabs Inc.) overnight. Sticky ends were created in a similar fashion as previously described except 5 mM dTTP as opposed to dATP was added to the reaction buffer. 200 ng of PCR product and 20 ng of the linearized γLIC vector were mixed at 66°C for two minutes. Subsequent cooling to room temperature allows the complementary ends to anneal. 10 µl aliquots were used to clone into ultra-competent *Escherichia coli* DH10B.

Ultra-competent *E. coli* DH10B cells were made competent via the Inoue method (Inoue *et al*., 1990). Frozen bacterial cells were streaked onto lysogeny broth (LB) agar plates. Single colonies were picked and grown under dark conditions in 25 ml liquid LB. Flasks were shaken at 200 rpm, and the temperature was maintained at 37°C for 6-8 hours. Three 500 ml flasks containing 100 ml SOB media (Table A.2) were inoculated with 0.8, 1.6, and 4 ml of the starter culture to ensure one flask would be at the appropriate stage the following morning. These flasks were grown at 18°C overnight with shaking at 150 rpm. The OD_{600} was checked on a Nanodrop 1000 the following morning. Once the OD_{600} of one of the cultures reached 0.055, the culture was put on ice for ten minutes to stop growth. The bacteria were centrifuged for ten minutes at 2,500 g and 4°C. The supernatant was discarded and the bacteria were gently resuspended in 32 ml of ice cold Inoue buffer (55 mM MnCl2*4H$_2$O, 15 mM CaCl$_2$*2H$_2$O, 250 mM KCl, 10 mM PIPES) and placed on ice for ten minutes. The bacteria were centrifuged using the same parameters as before and the supernatant was again discarded. The bacteria were resuspended in eight ml of ice cold Inoue buffer, 600 µl DMSO was added to the suspension, and it was placed on ice for another ten minutes. Aliquots of 100 µl were put into Eppendorf tubes and were flash frozen with liquid nitrogen and stored at -80°C.

Plasmids were introduced into *E. coli* via the heat shock method. 50 µl thawed, ultra-competent *E. coli* were mixed with 25 ng of plasmid. Tubes were gently inverted to mix and incubated on ice for 30 minutes. A 42°C water bath was used to heat shock the bacteria cells for 45 seconds. The bacteria were then moved back on ice for two minutes to cool prior to adding 600 µl liquid SOC media (Table A.2) although both LB and SOB media also worked. Cells were grown for one hour at 37°C with gentle shaking to prevent bacteria from settling. 50 µl was plated onto LB agar plates containing 100 µg/ml kanamycin. Transformed *E. coli* were tested by colony PCR using primers that flanked the cloning site (Table 3.1). Positive transformants were grown in three ml LB media containing 100 µg/ml kanamycin, and the transformed plasmid was isolated with the QIAprep Spin Miniprep Kit (QIAGEN Inc.). Additionally, the selected colonies were prepared for long-term storage at -80°C in a 25% glycerol/media mixture.

Electroporation was then used to transform the γLIC plasmid containing the gene fragment of interest into electrocompetent *A. tumefaciens*. One µl of plasmid (100-500 ng plasmid) was added to 10 µl thawed, competent *A. tumefaciens* strain EHA105 on ice. The mixture was incubated for one to two minutes on ice and then transferred to a chilled 0.2 cm cuvette. The Gene Pulser II Electroporation System was set to 2.5 Kv/ 400 Ω/ 25 µF to deliver the required shock. After electroporation, the cuvette was washed down with one ml chilled liquid LB and then the contents was transferred to a three ml microcentrifuge tube for incubation (28°C for 3 hour at 150 rpm). The *A. tumefaciens* cells were spun down at 2,700 g for five minutes, resuspended in 100 µl LB, plated on LB agar plates containing 100 µg/ml kanamycin and 25 µg/ml rifampicin, and incubated for one to two days at 28°C. Transformants were selected and prepared for -80°C storage in a 25% glycerol/media mixture.

Electrocompetent *A*. *tumefaciens* EHA105 was created as follows. A single colony of *A*. *tumefaciens* EHA105 was used to inoculate 20 ml of LB media containing 25 µg/ml of rifampicin. The starter culture was allowed to grow for 24 hours at 28°C. Ten ml of starter culture was then used to inoculate 200 ml of LB media with the same concentration of rifampicin. This culture was grown under the same conditions to an OD_{600} of 0.06. Once the bacteria reached the appropriate optical density, the cells were chilled on ice for 30 minutes and centrifuged for ten minutes at 2,200 g and 4°C. The supernatant was decanted and the bacteria were resuspended in 50 ml ice cold 1 mM HEPES buffer (1 mM HEPES, pH 7.5) by gently pipetting. The mixture was again centrifuged for ten minutes at 2,200 g and 4°C, the supernatant was decanted, and the bacteria were resuspended in 25 ml ice cold 1 mM HEPES buffer by gently pipetting the mixture. The centrifugation and resupension steps were repeated again but the bacteria were resuspended in two ml ice cold 1 mM HEPES buffer/10% glycerol. A final repetition of the centrifugation and resupension steps were repeated but the bacteria were resuspended in 550 µl ice cold 1 mM HEPES buffer/10% glycerol. Aliquots of 55 µl of the bacterial mixture were flash frozen in chilled Eppendorf tubes and stored at -80°C until needed.

3.4 Results

Gene fragment cloning

High quality RNA was routinely extracted from young leaf tissue of all species. RNA quality was evaluated based on 260/280 and 260/230 ratios provided by Nanodrop 1000 readings. Nearly all 260/280 ratios were ~2.0 and the 260/230 ratios were ~2.2. The RNA quality was not visualized on a gel since the end use of the RNA was PCR amplification. Prior to PCR amplification, cDNA synthesis was performed. Due to the small volume of cDNA synthesized per reaction, the quality of the cDNA was based purely off of its ability to act as a substrate for

PCR. The primers designed to amplify PDS based on the aligned sequences of other grasses successfully amplified PDS in *Miscanthus*. Initial attempts to isolate the amplified fragment from the gel were unsuccessful due to the low recovery rate of the gel purification. To combat this, the number of cycles in the PCR was increased and cDNA from four lanes was combined into a single extraction. This provided an ample quantity of DNA for sequencing purposes, but often the 260/230 ratio was low. In this case an ethanol precipitation was performed to further purify the DNA although the sequencing quality did not seem to be overly affected by the 260/230 ratio. Primer design for LIC based off of the *Miscanthus* sequence data nearly always gave a strong single band when using the previously purified amplicon. After testing a portion of the PCR product for quality, the remaining portion of the PCR product was utilized for column-based purification. At this point, the gene fragment was ready for LIC cloning. In addition to PDS, fragments of CRTISCO, IspF, and IspH genes were purified, sequenced, and prepared for LIC. Those sequences can be found in table A.3.

Ligation independent cloning

The α, β, γ, γLIC4, and γLIC5 containing plasmids were received from Professor Andy Jackson in *A. tumefaciens* EHA105 as stab cultures. The γLIC4 and γLIC5 containing plasmids represent independent ligation events associated with the insertion of the LIC site into the γ genome segment. The bacteria containing the α plasmid failed to grow from the initial stab culture as the result of a freezer malfunction in Professor Jackson's lab. Bacteria from a backup freezer were still viable and were subsequently shipped. Primers specific to the viral segment contained within each plasmid were used to ensure that each bacterium contained the correct plasmid (Figure 3.7). One of the two γLIC containing bacterium, γLIC4, turned out to be either mislabeled or contaminated by bacteria containing the β plasmid.

The basic steps associated with LIC are as follows: the γLIC5 plasmid had to be isolated from *A. tumefaciens*, the gene fragment had to be inserted into the cut plasmid, and the plasmid/gene fragment hybrid had to be transformed into *E. coli* for in vivo ligation. The ligated plasmid then has to be isolated from *E. coli* and transformed back into *A. tumefaciens*. In order to accomplish all of these steps, competent *E. coli* and *A. tumefaciens* had to be prepared. To test their competency, the α, β, γ containing plasmids were transformed into each bacterium. In the case of *E. coli*, hundreds of transformants were produced (Figure 3.8). Likewise, so many *A. tumefaciens* transformants were produced that individuals colonies could not be identified (Figure 3.8).

Prior to inserting any gene fragments into the γLIC5 plasmid, the plasmid had to be cut by *Apa*I. The manufacturer's protocol suggested an incubation time of 10-15 minutes but this did not appear to be a long enough time for the restriction enzyme to cut all of the plasmid; therefore, *Apa*I was allowed to incubate overnight (Figure 3.9). Once linearized plasmid was obtained, it was treated with T4 DNA polymerase and combined with the similarly treated gene fragment of interest. After which it was transformed into *E. coli* for ligation. Putative transformants were tested via colony PCR. Initial colony PCR using the same primers as used to amplify the gene fragment showed 100% of the putative transformants contained the gene fragment. The plasmids were then isolated and electroporated into *A. tumefaciens*. Colony PCR was used to test putative transformants. None of the *A. tumefaciens* contained the gene fragment of interest. Subsequent testing revealed that the plasmid isolated from *E. coli* did not contain the gene fragment either. The small amount of unincorporated gene fragments that were being plated out alongside the heat-shocked *E. coli* were causing false positive on the colony PCR (Figure 3.10). To avoid future false positives, primers were designed to straddle the LIC site. While troubleshooting the

LIC step, a difference in size of the control plasmid from the expected fragment size was noticed (Figure 3.10). Subsequent sequencing of the fragment revealed that a portion of the viral sequence that included the LIC site had been deleted (Figure 3.11).

Agroinfiltration and rub inoculation

Leaves of *N. benthamiana* were inoculated with a mixture of *A. tumefaciens* EHA105 containing the α, β, and γ plasmids at two densities: OD_{600} 0.07 and 0.7. The higher density appeared to have symptoms more similar to bacterial yellowing as opposed to BSMV infection; therefore, an OD_{600} of 0.07 was routinely used. Leaves that were inoculated as well as the leaf directly above them were harvested for sap extraction. A portion of the leaves were also saved for RNA extraction to test for the presence of the virus. RT-PCR of the leaves confirmed BSMV infection (Figure 3.12). Despite all of the variations of rub inoculation, plant stage, and growth conditions tested, none of the grasses tested showed susceptibility to BSMV (Figure 3.13) either visually or with RT-PCR.

3.5 Discussion

Ligation independent cloning

Despite its recent genome duplication event and a lack of a reference genome, there was little difficulty in developing primers suitable for amplification of *Miscanthus* gene fragments by using consensus sequences of related plants. Making bacteria competent in-house to move plasmids between *E. coli* and *A. tumefaciens* also posed little difficulty. In terms of transforming plasmids into either bacterial species, the more concentrated the plasmid the higher the transformation efficiency (data not shown). Regardless, the transformation efficiency was more than suitable for the applications herein.

Initial attempts to ligate *Miscanthus* gene fragments into the LIC site appeared to be successful when using colony PCR to test the *E. coli* that survived the selective media. Interestingly, when moving the plasmid from *E. coli* to *A. tumefaciens*, the fragment appeared to be lost. While trouble-shooting, it became obvious that it was never there in the first place, and the primer design for the colony PCR needed to be altered. By using the same primers that were used to amplify the original gene fragment to test colonies, false positives were occurring because some of the original gene fragments that were mixed with the cut plasmid and introduced to *E. coli* were still in the media when the bacteria were plated onto the selection plates. When the colonies were picked for colony PCR, enough of the gene fragments were also being picked to act as a template in the PCR reactions. Further, the intensity of the band could not be reliably used as an indicator of success.

 In typical ligation reactions, three primers are used to test for successful insertion: a primer that anneals to the inserted gene fragment and one on each side of the restriction site. PCR amplification of a product only occurs if the fragment has actually been ligated into the restriction enzyme site. Further, the direction of the insert can be determined based on which primer pair produces the amplicon. In the case of LIC, the direction is determined by the primer design, i.e. the five prime overhangs on the primers used to amplify the gene of interest; therefore, primers flanking the LIC site can be used to test for the successful insertion of any gene into the LIC site. Insertion of the gene fragment will lead to a larger amplicon than that of the control γLIC plasmid. This strategy reduces the number of primers needed and avoids false positives.

Despite designing and testing multiple primer sets, no PCR amplification could be detected in either colony PCR or from the extracted plasmid. As a result primers were designed

that amplified the α, β, and γ genome segments to ensure that the plasmids received were indeed

the correct plasmids and that contamination had not occurred during their cultivation and

maintenance. Indeed, the γLIC4 plasmid was found to be the β containing plasmid, and it

appeared to have been the entire time as stocks stored in the -80°C freezer were also the β

containing plasmid. Unfortunately, the original stab cultures were no longer viable, so it cannot

be determined beyond a doubt that the contamination occurred prior to or after receiving the

plasmids, but considering that the plasmids were prepared for long term storage from the onset, it

seems likely that the contamination occurred prior to their shipment. Fortunately, the γLIC5

plasmid did test positive for the presence of the γ containing plasmid. Likewise, all of the other

plasmids contained the expected BSMV sequences.

LIC with the γLIC5 plasmid was attempted, yet primers flanking the LIC site revealed

that none of the *E. coli* colonies growing on the selection plate contained the inserted gene

fragment. Considering the fact that LIC generally has a very low false positive rate (Aslanidis &

DeJong, 1990), the logical conclusion to draw was that the restriction enzyme digest of the

γLIC5 plasmid was not functioning properly. Certain enough, gel electrophoresis of digested

γLIC5 plasmid revealed that less than half of the plasmid was being cut when following the

manufactures protocol. It was these uncut plasmids that were being transformed into the *E. coli*,

thereby causing the false positives. Extending the incubation time alleviated this concern

although the minimum time required to cut all of the plasmid could not be determined due to the

few time intervals tested.

Surprisingly, LIC with the completely digested plasmid still led to 100% false positives.

At this point, the region containing the LIC site was sequenced to be sure that it was accurate.

Unfortunately, a deletion had occurred that removed 479 base pairs immediately following the γb

gene that included the LIC site. Unfortunately, the 48 base pair fragment that was inserted in its place contained an *Apa*I restriction site. Due to the site of the deletion, it likely occurred during the creation of the γLIC5 plasmid. Deletions associated with the transformation of *E. coli* with linear plasmids have been noted in the literature for many years (Conley *et al*., 1986). This deletion may have gone unnoticed if the plasmids were identified by restriction enzyme digestion.

Having the incorrect LIC cloning vector prevented the insertion of *Miscanthus* gene fragments into the viral sequence. In many cases, mistakes such as these can be identified through PCR and restriction enzyme digests. Odd circumstances prevented these in this particular situation. The amount of time and effort put forth troubleshooting this procedure compared to the price of sequencing plasmids begs the question of whether all plasmids received from collaborators should be sequenced or not moving forward.

BSMV susceptibility

The fact that *N. benthamiana* could be infected using a mixture of the α, β, and γ plasmids showed that the plasmids were functioning properly despite all of the issues with the two LIC plasmids. This allowed us to test the susceptibility of *Miscanthus*, sorghum, and maize to the ND18 strain of BSMV. Unfortunately, all genotypes tested proved recalcitrant to infection, which is surprising given the large monocot host range of BSMV (Jackson & Lane, 1981). When a fragment including the LIC site of the γ containing plasmid was sequenced, a couple of point mutations did appear. These point mutations could have arisen during PCR amplification of the virus sequence during the construction of the γ plasmid. If these point mutations were not an artifact of sequencing error, they very well could lead to a decrease in fitness of the virus. The concentration of the virus could also have been insufficient to cause infection on the grasses tested (McKinney & Greeley, 1965).

Two alternative scenarios could explain these results. One or all three of the species may be susceptible to BSMV, but the genotypes tested were not, or the plant genotypes tested may be susceptible to BSMV, but the ND18 strain was not virulent to these genotypes. Evidence of differences in virulence between BSMV strains is common in the literature (McKinney & Greeley, 1965; Chiko, 1984; Petty *et al*., 1994). Likewise, differences in susceptibility of plant genotypes to the same viral strain has been reported numerous times (Kang *et al*., 2005). The question of which one is the culprit in this particular case cannot be determined as of now. Considering the effort required for both, it would be easier to test additional plant genotypes. Initial efforts to develop a detached leaf assay based on a modification of the protocol described by Lindhout *et al*. (1988) in *Miscanthus* may be possible as mature, detached *Miscanthus* leaves stay fresh ~14 days when kept moist and in the dark. This would make screening additional varieties considerably easier. The more challenging aspect is testing whether the detached leaf assay works as barley leaves have a significantly shorter shelf life after removal (~3 days). One of the reasons that the ND18 strain was chosen by Yuan *et al*. (2011) to develop into a VIGS vector was its lack of strong interfering symptoms in BSMVs natural host range. While this is conducive to evaluating the results of down regulating genes in natural hosts, it may have a negative effect on host range.

Future research

The development of a reliable functional genomics tool for *Miscanthus* should remain a high priority. It would open many additional research avenues and allow the ever increasing genetic information to be leveraged more effectively. Despite the lack of susceptibility to BSMV reported herein, *Miscanthus* remains an excellent candidate for VIGS. When considering future research plans regarding the VIGS work began here, two levels of dedication need to be

considered. The first is a minimalist approach that will take very little additional effort but has a much lower chance of succeeding. The second approach would be more comprehensive and would have a better chance of success.

The minimalist approach would be to continue attempting to infect *Miscanthus* with the BSMV clones used herein by screening additional *Miscanthus* genotypes, continue to modify the inoculation procedure, or attempt various pretreatments of the plants. One potential option would be to concentrate the virus prior to inoculation (Lawrence & Jackson, 1998). If a particular genotype or treatment is found to be affective, then the correct LIC plasmids could be acquired. Permits have already been attained to receive the functional γLIC plasmid. While this plan would require little extra work, the likelihood of success is proportionally low.

A more comprehensive approach would have a much better chance of success, but it would require a dedicated effort and substantial funding. Ideally, a cross-disciplinary team comprised of *Miscanthus* breeders, virologists, and geneticists would be assembled to screen a multitude of current VIGS vectors and germplasm. In addition, viruses that naturally infect *Miscanthus* could be evaluated for their potential as a VIGS vectors. Some of the current raw genetic data may even be utilized to accomplish this. Raw sequences that do not align to sorghum could be blasted against known viruses. In this manner, certain viruses may be identified to naturally infect *Miscanthus* with minimal effort and cost.

One alternative VIGS vector is *Brome mosaic virus* (BMV). Similar to BSMV, BMV has a large monocot host range (Lane, 1974) and has a long history of use as a VIGS vector (Ding *et al*., 2006; Benavente & Scofield, 2011). Permits have been completed to acquire the BMV VIGS system utilized by Ding *et al*. (2006). Beyond BSMV and BMV there are only a few viruses that have been utilized in VIGS for monocotyledonous hosts (Ramanna *et al*., 2013). The number of

viruses that can actually be tested is limited by time, money, and willing collaborators; therefore, potential alternatives as well as collaborators will have to be evaluated and ranked based on their chance of success.

A compromise between a large collaboration and continuing attempts with the same VIGS system may be available. The Norwich clone of BSMV has been shown to have a wider host range than ND18, and it has already been developed into a VIGS vector (Professor Andy Jackson, personal communication). In fact, the main difference between the ND18 clone and the Norwich clone is the helicase motif on the β genome segment (Mi-Yeon *et al*., 2012). This means that the same protocols used herein can be attempted with the only difference being a swap of the ND18 β containing plasmid with that of the Norwich β containing plasmid. Again, the appropriate permits to obtain the Norwich vector have been obtained and should be shipping soon. Regardless of the scale in which the development of a VIGS system suitable for *Miscanthus* is pursued, it remains an attractive tool for functional genomics in *Miscanthus* and other grasses.

3.6 Figures and tables

Figure 3.1 A diagram depicting post-transcriptional gene silencing (PTGS) as a defense against plant viruses. When double stranded RNA is created by RNA dependent RNA polymerases (RDR) during viral replication or transcription, it is recognized and cut by Dicer-like (DCL) proteins into 21-24 base pair siRNA. These short viral sequence fragments can then be loaded into the RNA-induced silencing complex (RISC), resulting in homology dependent degradation of viral RNA. Additionally, endogenous RNA-dependent RNA polymerases (eRDR) can amplify the silencing signal by creating additional dsRNA. The silencing signal can then travel cell to cell via plasmodesmata, eventually reaching the phloem.
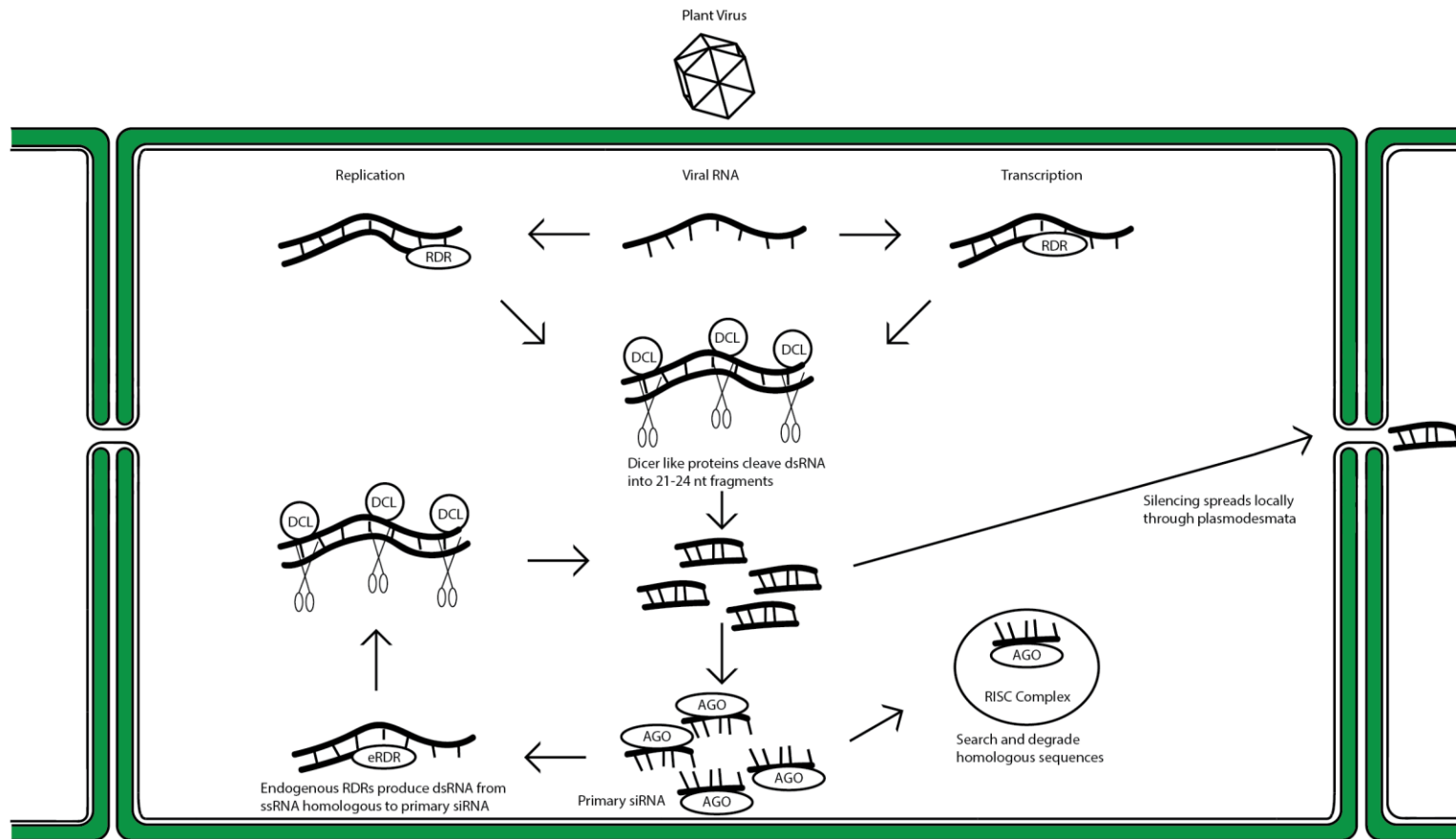
Figure 3.2 A diagram depicting virus induced gene silencing (VIGS) in plants. A fragment of a plant gene, depicted in red, is inserted into the viral genome. The post-transcriptional gene silencing machinery cannot differentiate between the plant gene fragment and the viral genome; therefore, when double stranded viral RNA is recognized and cut by Dicer-like (DCL) proteins into 21-24 base pair siRNA, the plant gene fragment is included. These short plant gene fragments can then be loaded into the RNA-induced silencing complex (RISC), resulting in homology dependent degradation of the virus and the plant gene, or they can move cell to cell through plasmodesmata.

110
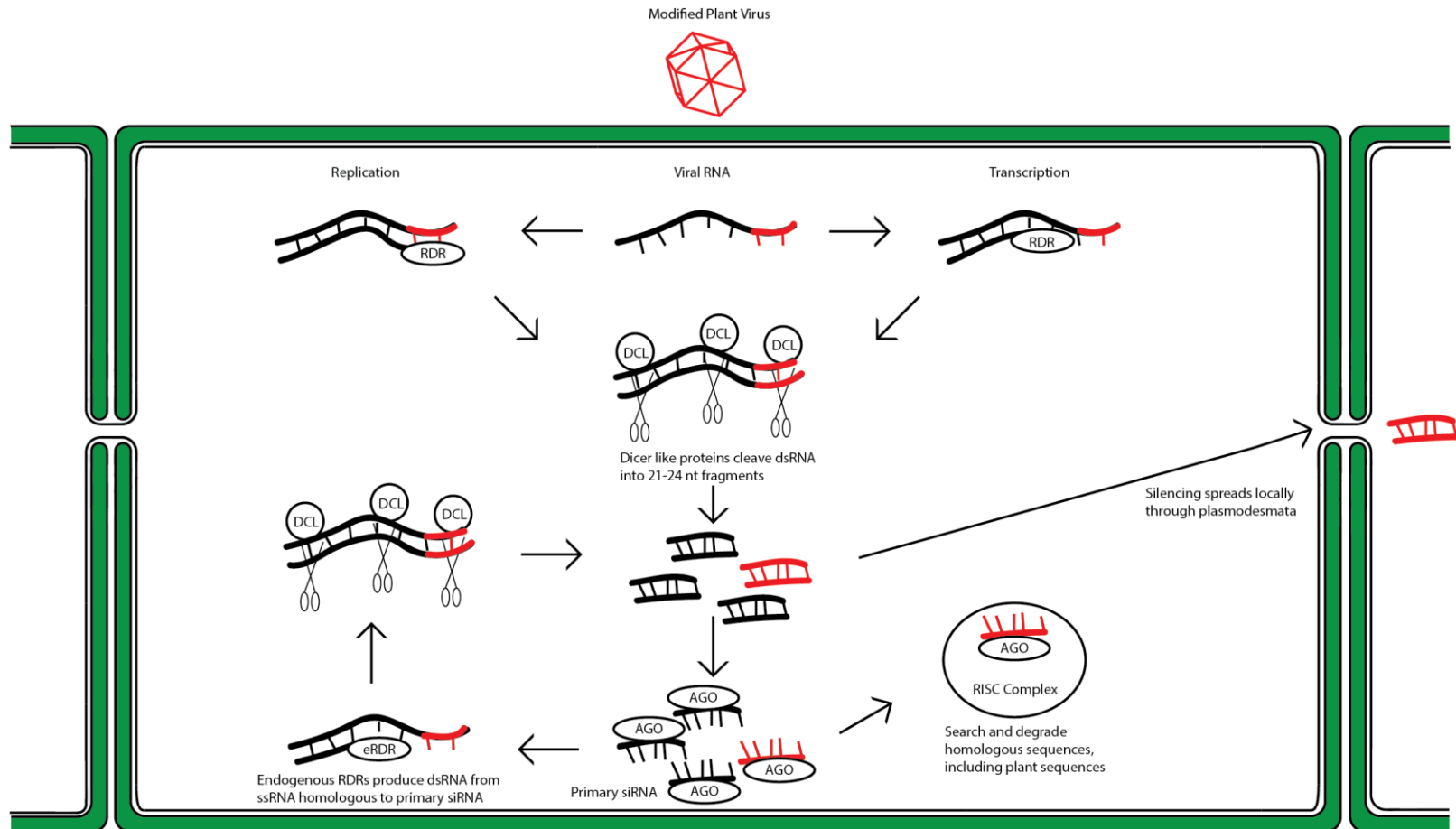
Figure 3.3 Viruses move cell to cell via plasmodesmata, ultimately reaching the phloem and traveling systemically [As seen in Harries & Ding, 2011].
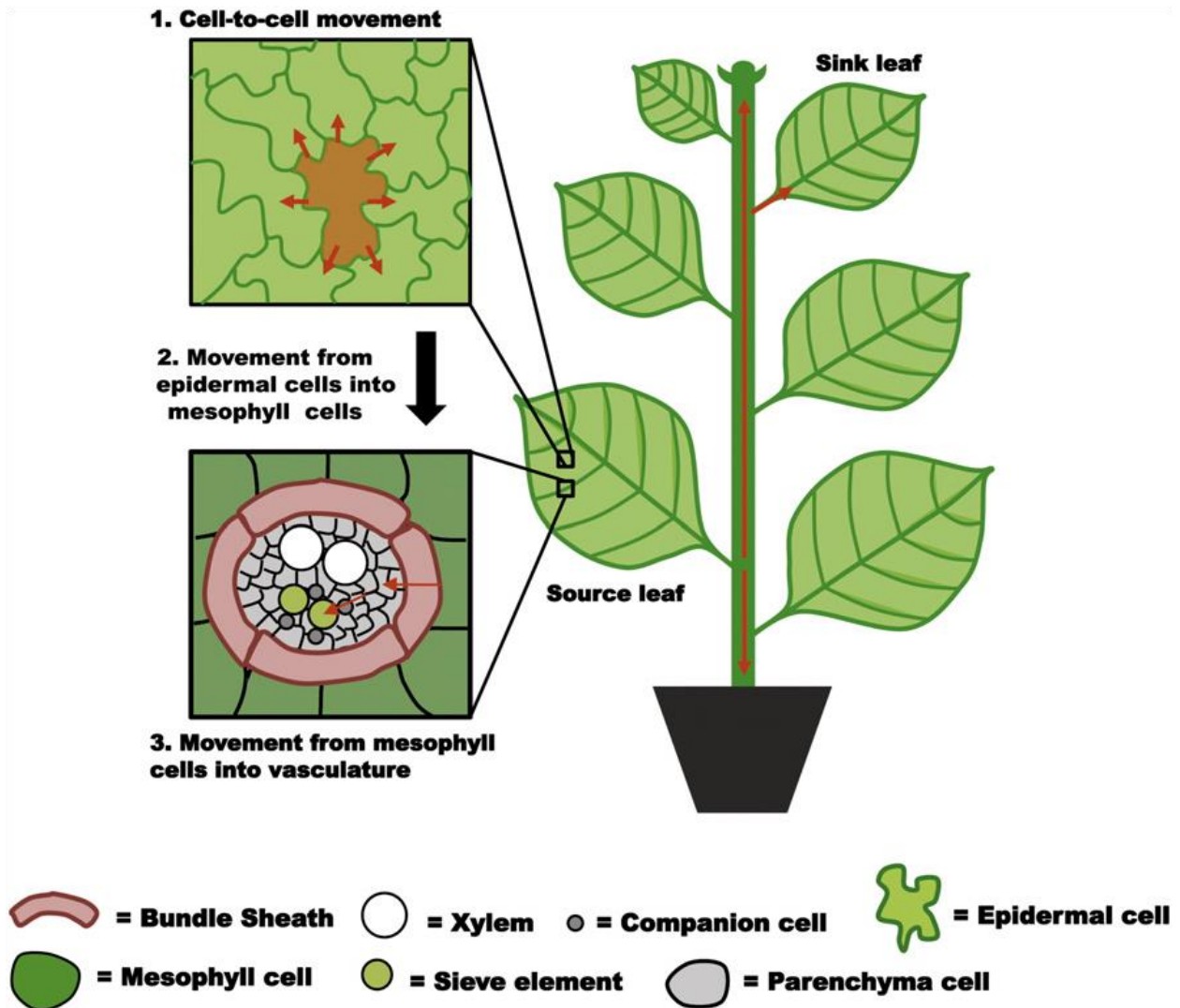
Figure 3.4 (A) The genomic structure of the α, β, and γ genome segments of BSMV along with the (B) γLIC genome segment with the ligation independent cloning site added immediately after the γb gene. Each genome segment was cloned into a plasmid containing a kanamycin resistance gene between a double 35S promoter and a self-cleaving ribozyme. [Slightly modified from Yuan *et al.*, 2011]
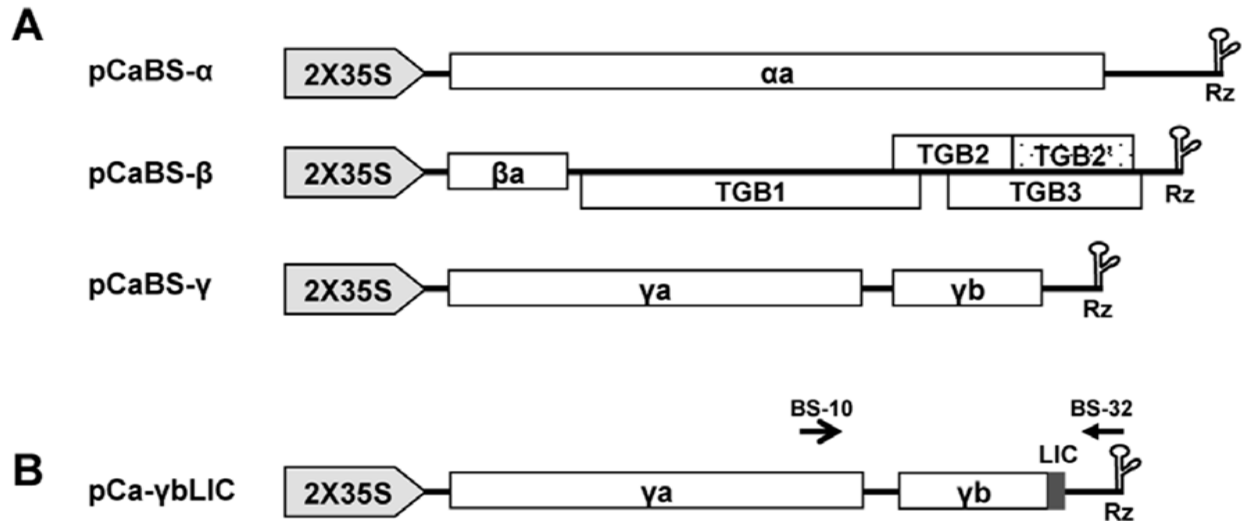
Figure 3.5 An overview of the key steps in the agroinfiltration process: (A) *Agrobacterium tumefaciens* containing the α, β, and γ containing plasmids are suspended in the infiltration buffer and mixed in equal parts, and (B) the mixture of bacteria is injected into the leaves of *Nicotiana benthamiana* in the four to twelve leaf stage with a needleless syringe. (C) After leaves are infiltrated, they appear water soaked.
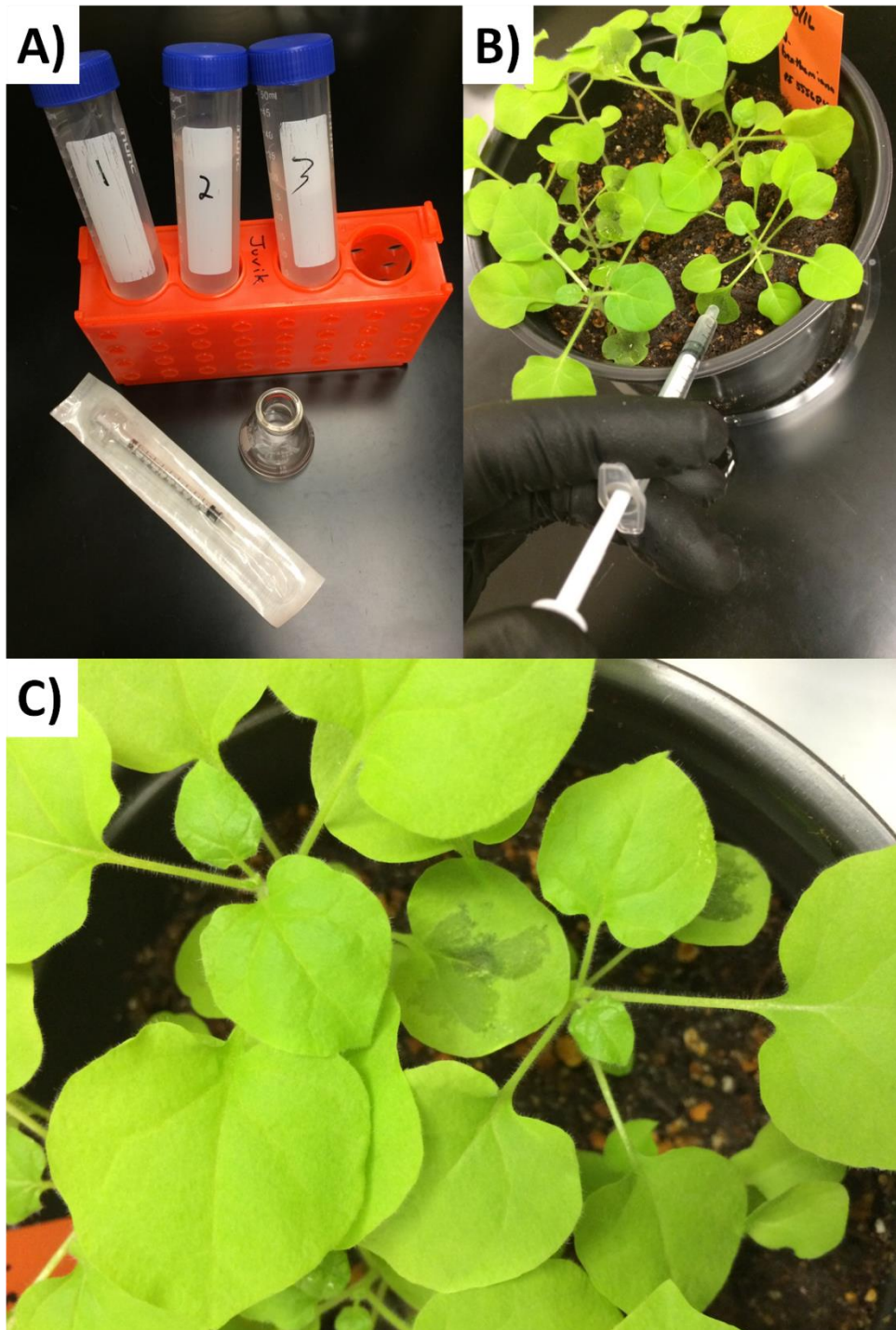
Figure 3.6 BSMV infected *Nicotiana benthamiana* leaves were ground in extraction buffer and the resulting sap (A-inset) was rub inoculated onto young seedlings (A). *Miscanthus* (B), sorghum (C), and maize (D) seedlings prior to inoculation.
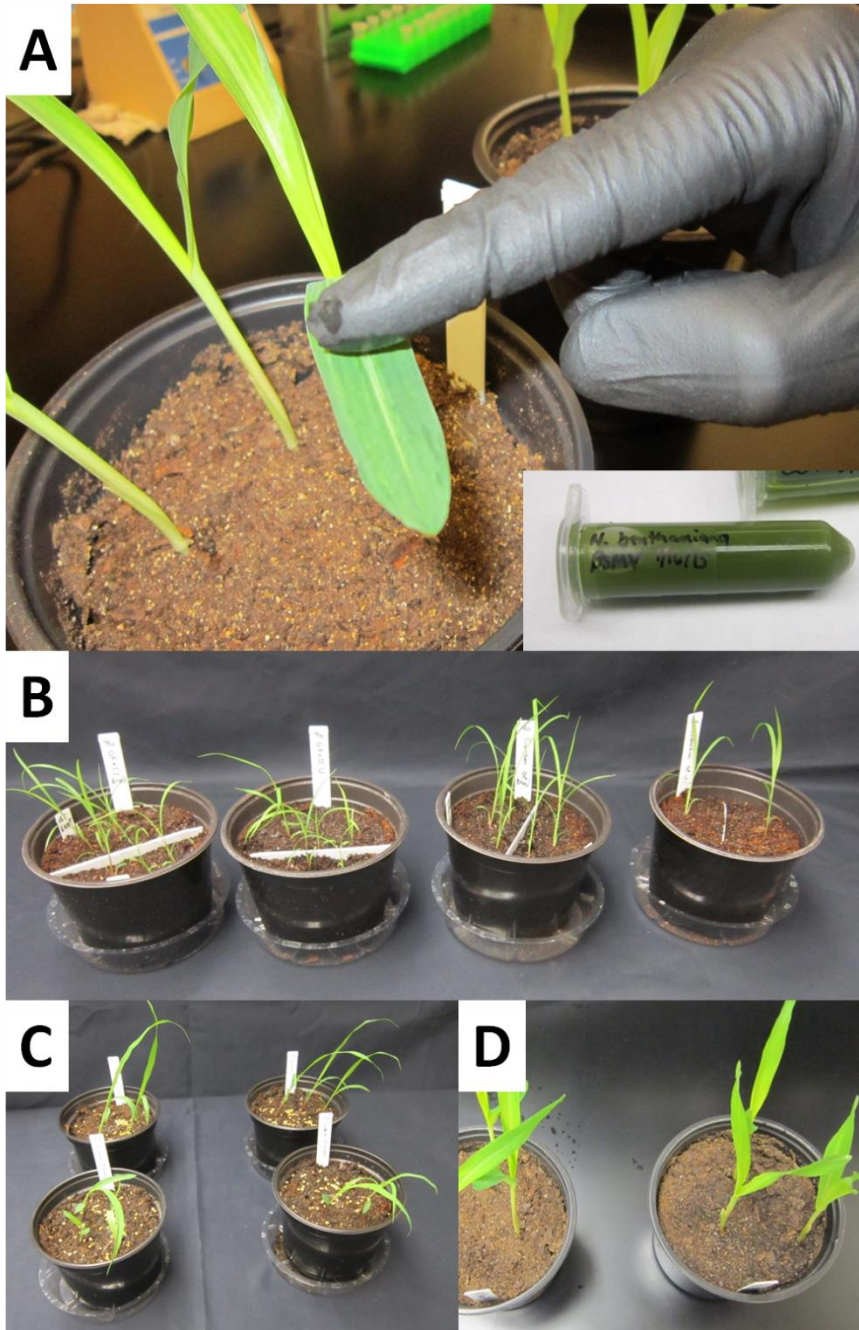
Figure 3.7 A 1.2% agarose gel stained with SYBER Safe and imaged under ultraviolet light showing that BSMV γLIC4 was actually the plasmid containing the β BSMV segment. Further, the γLIC5 segment was ~400 base pairs smaller than the γ BSMV segment.

| Lane	1	2	3	4	5	6	7	8	9	10	11	12
Primer	NA	BSMV_α	BSMV_α	BSMV_α	BSMV_α	BSMV_α	BSMV_β	BSMV_β	BSMV_β	BSMV_β	BSMV_β	InTest4
Template	L[1]	E. coli[2]	Agro_α[3]	Plasmid_α[4]	E. coli_α[5]	Plasmid_4[6]	E. coli	Agro_β	Plasmid_β	E. coli_β	Plasmid_4	E. coli
Lane	13	14	15	16	17	18	19	20	21	22	23	
Primer	InTest4	InTest4	InTest4	InTest4	InTest4	InTest4	InTest4	InTest4	InTest4	InTest4	NA	
Template	Agro_γ	Plasmid_γ	E. coli_γ	Agro_4	Plasmid_4	E. coli_4	Agro_5[7]	Plasmid_5	E. coli_5	Fr Agro_4[8]	L	

1 DNA Ladder
2 Control E. coli
3 A. tumefaciens containing BSMV plasmid α, β, or γ
4 Isolated α, β, or γ plasmid
5 E. coli containing BSMV plasmid α, β, or γ
6 Isolated γLIC4 plasmid
7 A. tumefaciens containing BSMV plasmid γLIC5
8 A. tumefaciens containing BSMV plasmid γLIC4 that was frozen as soon as it was received

Figure 3.8 (A) *E. coli* DH10B and (B) *A. tumefaciens* EHA105 transformed with the BSMV gamma containing plasmid growing on selective media showing the effectiveness of the protocols employed to make the bacteria competent for transformation. The BSMV gamma containing plasmid contains a gene inferring resistance to kanamycin. *A. tumefaciens* EHA105 contains a gene inferring resistance to rifampicin. The LB plates include (A) 100 µg/ml kanamycin and (B) 100 µg/ml kanamycin and 25 µg/ml rifampicin, respectively. The negative control in (A) is not pictured while the center petri dish is the negative control in (B).
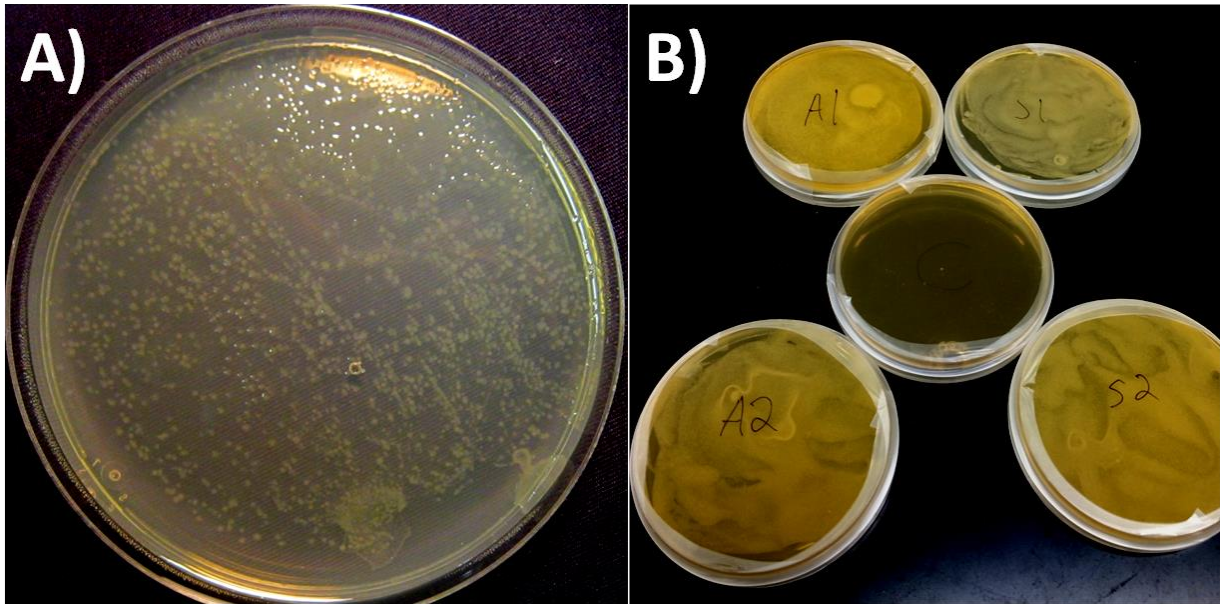
Figure 3.9 A 1.2% agarose gel stained with SYBER Safe and imaged under ultraviolet light showing the effect of incubation time on *Apa*I's ability to cut the γLIC5 plasmid: Lane 2—1 µg uncut γLIC, Lane 3—1µg cut γLIC incubated for ten minutes, Lane 4—1 µg cut γLIC incubated overnight, and Lane 5—0.5µg cut γLIC incubated overnight.
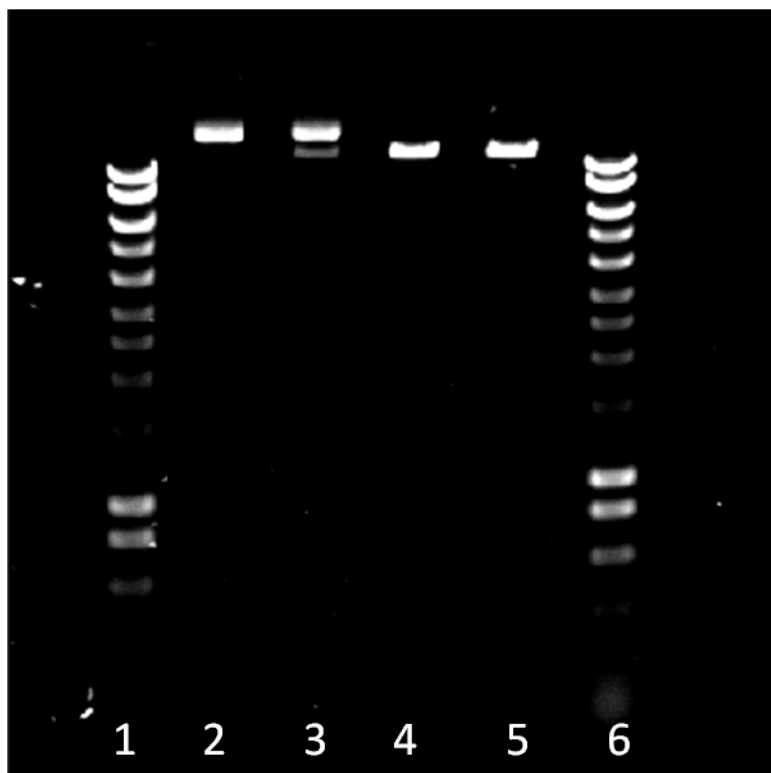
Figure 3.10 Two 1.2% agarose gels stained with SYBER Safe DNA gel and imaged under ultraviolet light showing the results of colony PCR testing the successful insertion of the *Miscanthus* CRTISO gene fragment into the ligation independent cloning (LIC) sight of the γLIC5 plasmid. Gel (A) shows the misleading results that can occur if the primers used to amplify the gene fragment are the same used to test its insertion into the LIC sight. Lane two is the positive control showing PCR product of the MiCRTISO gene fragment, lane three is the negative control (*E. coli* DH10B), lane four is the result of touching the selective media with a pipette tip and introducing that tip to the PCR mixture, and lanes five thru seven are independent *E. coli* colonies. Gel (B) shows a similar test but the primers used to test the colonies for the insertion of the gene fragment flank the LIC sight. Lane two is a negative control (empty γLIC plasmid), lane three is also a negative control (*E. coli* DH10B), and 4-17 are independent *E. coli* colonies. While LIC appears to have been successful based on gel (A), this is a result of residual gene fragment on the selective media. If the colonies truly had the gene fragment in the LIC sight, then the bands in gel (B) would be ~230 bp larger.
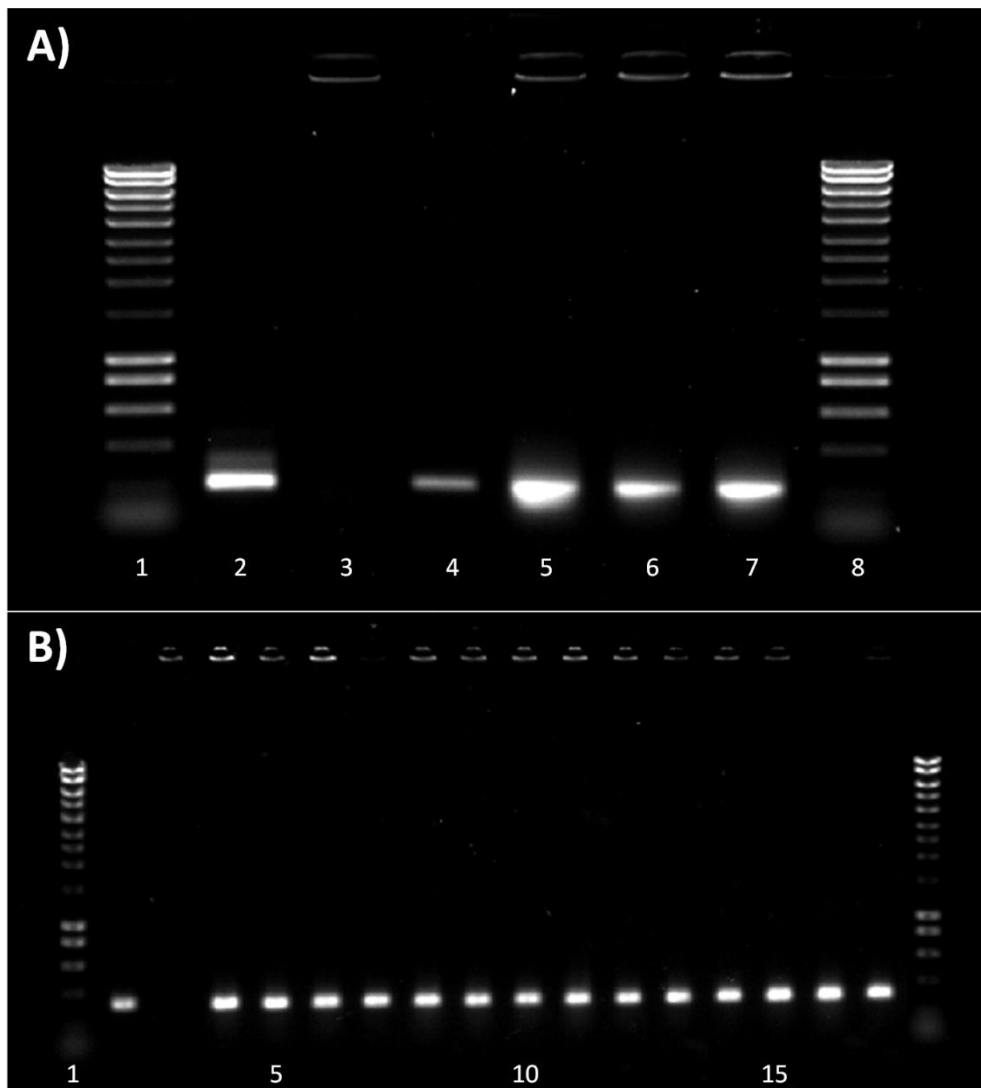
Figure 3.11 Aligned sequences of the γ BSMV segment with the LIC site as published by Yuan *et al*., 2011 (black) and the γLIC5 BSMV segment showing a 479 base pair deletion in the γLIC5 BSMV segment. The addition of a 48 base pair fragment results in a difference of 431 base pairs. Despite this deletion, the γLIC5 does have an *Apa*I restriction site (underlined).

```
AGTGTGACGCAGCTACCGGAAGTTGTAGCTTACATACCAATTCCTAAATTTTCTCTCCAGAATCCGTTAAGATTCAT
------------------------------------CAATTCCTAAATTTTCTCTCCAGAGTCCGTTAAGATTCAT

GGTTTCCAATTCAGGCATCGTTTTCAAGTTCGATTATAGTGGACTTGCAAACACTCCCATCATATGGTTGATGGGCA
GGTTTCCAATTCAGGCATCGTTTTCAAGTTCGATTATAGTGGACTTGCAAACACTCCCATCATATGGTTGATGGGCA

CCATCAGATTTGAATGATCTGATCAAACATTTTTTTTTTTTTAaccaccaccaccGGGCCCttcCAACTTAGAAAC
CCATCAGATTTGAATGATCTGATCAAACATTTTTTTTTTTTTA-------------------------------

GGAAGAAGAATCATCACATCCAACAGAATCTTCAAAAGAAGAAGCTACGGACTTACGTATTGCGTTAACCTCACTTT
-----------------------------------------------------------------------------

CAAGCTTAGCCATTTTTACGATATGAGAAAGTTTCAGCTCCTGCATCTTCTTCTGGAGAAATTCAAGAAGTAACTCC
-----------------------------------------------------------------------------

TGTTCAGAACGTTTCAGAAGTGAATCATACAGATCCGCATGCTTTTGGCCACAGAACCTACTCACGATTGGCAGTTG
-----------------------------------------------------------------------------

ATCACAAGCCTCTTCCGCAATGGAGCATGGCATTCCACAACTGTGTCCAACGATTCCATTCAGGGCGCATTTTTGCG
-----------------------------------------------------------------------------

GTTCCAATAGATACTTCTTGTAAAGTTCCAATCTCTTATTTCTTGTTTCAGAATATACATGCTTTCGCTCACATCTC
-----------------------------------------------------------------------------

TTACCACAGTAAGTACTTGTAGTTGAGGTACCACAACACACACAAGAGAAAGTAGCCATCA---------------
----------------------------------------------------------------GCTAGCTGAGCAAGAG

------------------------------TGCGAAGGTAAATACAGTAGATTTAAACATCAGGACCCAGAGTTC
CAACAGGGGCCCTCTTGTCGTCGTCAGGATCCTGCGAAGGTAAATACAGTAGATTTAAACATCAGGACCTAGAGTTC

ACCACTCGAAGTCCTTTCTCA--
ACCACTCGAAGTCCTTTCTCAAA
```

Figure 3.12 *Nicotiana benthamiana* plants ten days after (A) agroinfiltration in comparison to (B) control plants infiltrated with the buffer alone. (C) A close-up of the inoculated leaf (left) and a control leaf for comparison (right). (D) RT-PCR results confirming the presence of the virus: Lane 2—positive control; Lane 3—control *N. benthamiana*; Lane 4 and 5—agroinfiltrated *N. benthamiana*.
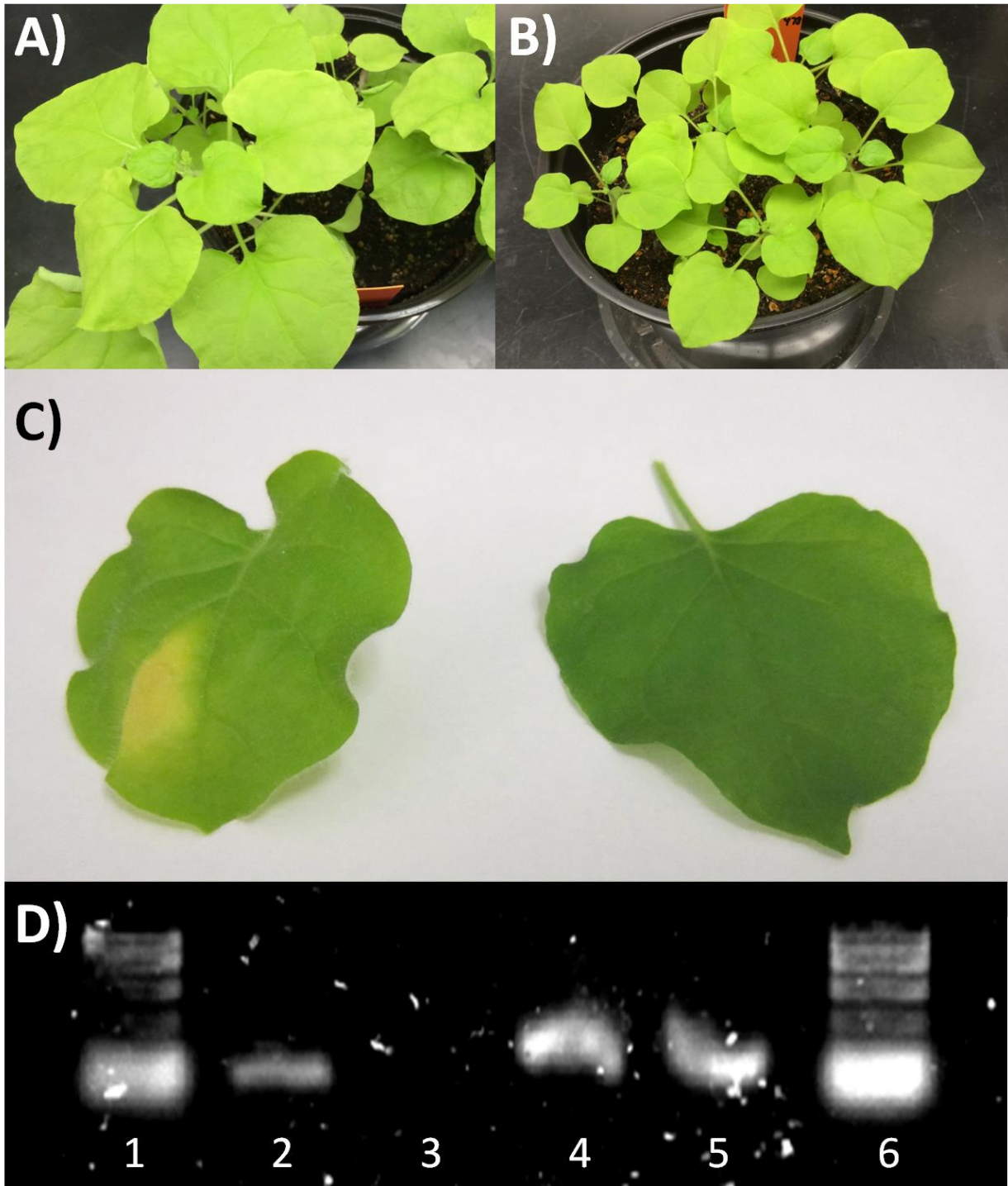
Figure 3.13 *Miscanthus* (A-E), sorghum (F), and maize (G-H) 14 days after inoculation with BSMV infected or control sap. The two leftmost pots in (A) and (B) and (C) are control plants while the two rightmost pots in (A) and (D) and (E) were inoculated with BSMV. In contrast, the left leaf in (F), (G), and (H) were inoculated with BSMV while the leaves on the right were inoculated with control sap. No differences were apparent between the control and inoculated plants. This was backed up with RT-PCR results.

Table 3.1 The primers, along with their uses, that were utilized to test the VIGS system on *Miscanthus*, sorghum, and maize.

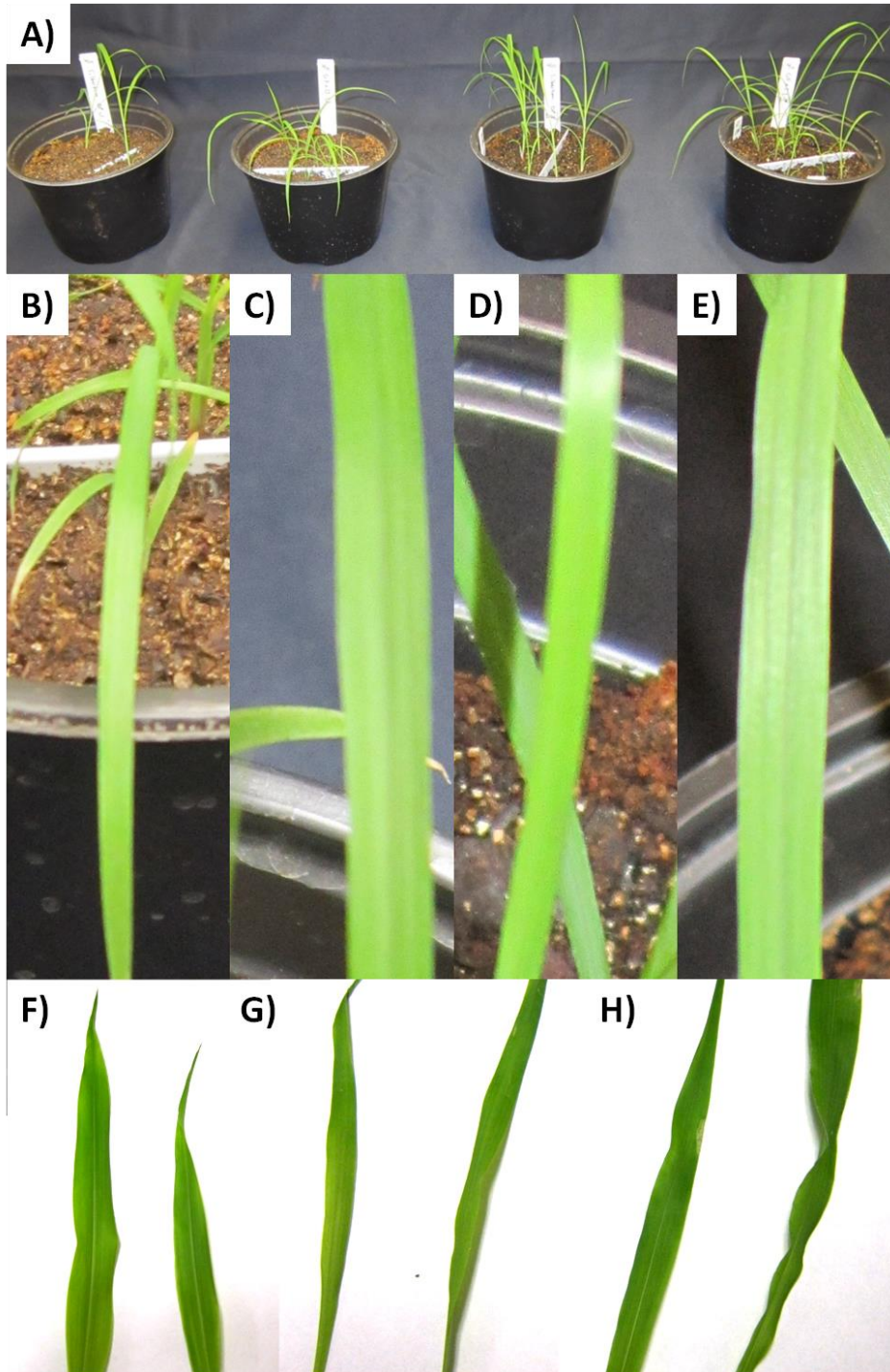| Primer | Primer Sequence (5'-3') | Amplicon | | | Usage |
		Species	Gene/Segment	Length	
EF1a_F	GTGGTTTTGAGGCTGGTAT	Miscanthus	Elongation factor-1	253	Positive control
EF1a_R	GACCTCTCAATCATGTTGTC	Miscanthus	Elongation factor-2	253	Positive control
InTest4_F	AGTGTGACGCAGCTACCG	BSMV	Gamma	728	Testing insertion into LIC site
InTest4_R	TGAGAAAGGACTTCGAGTGGTG	BSMV	Gamma	728	Testing insertion into LIC site
BSMV_α3_F	GCCTTTATGGCAGGTGAGGT	BSMV	Alpha	329	Testing for the alpha BSMV segment
BSMV_α3_R	TGAGAGTAACCGGCATTCGG	BSMV	Alpha	329	Testing for the alpha BSMV segment
BSMV_β2_F	GAAGACGCCCAAGTCTGTGA	BSMV	Beta	400	Testing for the beta BSMV segment
BSMV_β2_R	CAAAGCACAAACAGCCGGAA	BSMV	Beta	400	Testing for the beta BSMV segment
BSMV_γ_F	CTACGGAGGACATTAACATG	BSMV	Gamma	594	Testing for the gamma BSMV segment
BSMV_γ_R	CATTCAGTTCATCAGCAGTC	BSMV	Gamma	594	Testing for the gamma BSMV segment
PDS_F	CAGCAATGSTTGGWGGBCAA	Miscanthus	Phytoene desaturase	492	Amplification of Miscanthus PDS for sequencing
PDS_R	CAGGAACTCCCACYARCTT	Miscanthus	Phytoene desaturase	492	Amplification of Miscanthus PDS for sequencing
MiPDS_F	AAGGAAGTTTAATTTCTTCAGGAGAAGCATGG	Miscanthus	Phytoene desaturase	242	LIC of PDS fragment into gamma-LIC plasmid
MiPDS_R	AACCACCACCACCGTGCACAAACATAAGCATCTCC	Miscanthus	Phytoene desaturase	242	LIC of PDS fragment into gamma-LIC plasmid
CRTISO2_F	GCAAGCATGGTCTTGTGC	Miscanthus	Carotenoid isomerase	650	Amplification of Miscanthus CRTISO for sequencing
CRTISO2_R	TTGGTGAGCCTACCTCYTT	Miscanthus	Carotenoid isomerase	650	Amplification of Miscanthus CRTISO for sequencing
MiCRTISO2_F	AAGGAAGTTTAAAGGCTGTTGGGGTGAGGTTA	Miscanthus	Carotenoid isomerase	228	LIC of CRTISO fragment into gamma-LIC plasmid

Table 3.1 (Continued)

MiCRTISO2_R	AACCACCACCACCGTGCAGGCAGAACT GACGCTTT	Miscanthus	Carotenoid isomerase	228	LIC of CRTISO fragment into gamma-LIC plasmid
IspF_F	GACCTCCACCGCCTSGAG	Miscanthus	2-C-methyl-D-erythritol 2,4-cyclodiphosphate	312	Amplification of Miscanthus IspF for sequencing
IspF_R	TGTCTCCTTGAATGGGCTA	Miscanthus	2-C-methyl-D-erythritol 2,4-cyclodiphosphate	312	Amplification of Miscanthus IspF for sequencing
MiIspF_F	AAGGAAGTTTAAGCGAGGCTCACTCTG ACGG	Miscanthus	2-C-methyl-D-erythritol 2,4-cyclodiphosphate	237	LIC of IspF fragment into gamma-LIC plasmid
MiIspF_R	AACCACCACCACCGTGCAAGATCAGCG TAGCATCAAG	Miscanthus	2-C-methyl-D-erythritol 2,4-cyclodiphosphate	237	LIC of IspF fragment into gamma-LIC plasmid
IspH_F	GTCGAGCGCGCCGTGCA	Miscanthus	1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase	596	Amplification of Miscanthus IspH for sequencing
IspH_R	CCAATTTCTTCAGTTTCTCCTTT	Miscanthus	1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase	596	Amplification of Miscanthus IspH for sequencing
MiIspH_F	AAGGAAGTTTAAGTTGTGTTGCCTGCAT TTGGA	Miscanthus	1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase	260	LIC of IspH fragment into gamma-LIC plasmid
MiIspH_R	AACCACCACCACCGTGCCTCTGCCATAT TCTTCACAA	Miscanthus	1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase	260	LIC of IspH fragment into gamma-LIC plasmid
ZmPDS_F	AAGGAAGTTTAAAAGCAGGGTGTTCCT GATCG	Maize & Sorghum	Phytoene desaturase	249	LIC of PDS fragment into gamma-LIC plasmid
ZmPDS_R	AACCACCACCACCGTCTCTCCACCCCTA GACCGAA	Maize & Sorghum	Phytoene desaturase	249	LIC of PDS fragment into gamma-LIC plasmid
SbMidrib_F	AAGGAAGTTTAACGGAAAGACAGCGAG CAAGA	Sorghum	Sobic.006G147400	304	LIC of Midrib fragment into gamma-LIC plasmid
SbMidrib_R	AACCACCACCACCGTTGATCGAAAACC GACGACGA	Sorghum	Sobic.006G147400	304	LIC of Midrib fragment into gamma-LIC plasmid

CHAPTER 4: INVESTIGATING SELF-INCOMPATIBILITY IN *MISCANTHUS*

4.1 Abstract

The adoption of *Miscanthus* as a dedicated bioenergy crop is impeded by, among other things, the necessity to establish fields via clonal propagation. Plants grown from seed segregate widely, resulting in a decrease in plot yield. Further, seed established plots increase the invasive potential of this non-native genus of grasses. Both of these concerns could be alleviated by developing hybrid seed from inbred lines. In order to accomplish this goal, the strong self-incompatibility (SI) mechanism acting in *Miscanthus* would need to be manipulated. To better understand this mechanism, SI relationships between full siblings of two biparental populations were utilized to map the locus/loci responsible for self-recognition. A single locus was found to be segregating in both mapping populations, and conserved synteny among the grasses does not suggest that this locus corresponds to the S- or Z-loci found to operate in self-recognition of the Pooideae subfamily of Poaceae. Using rice and sorghum as bridge species, this region appears to be homologous to the T-locus, previously believed to only have two alleles: a functional allele and a non-functional allele. Increasing evidence suggests that sorghum has a deletion in the region homologous to the mapped SI locus in *Miscanthus*, explaining its lack of SI. Despite only detecting a single locus in each mapping population, an investigation of pollen tube growth in a more diverse set of crosses revealed two separate crosses in which 75% of the pollen was compatible. This can only occur in a two locus SI system, suggesting that at least one more locus is responsible for SI in *Miscanthus*. In light of these new revelations, the model in which the common ancestor of all grasses had a four locus SI system is supported. After the split of the Pooideae and Panicoideae subfamilies, these four SI loci underwent distinct evolutionary pathways, resulting in two related but distinct SI systems.

4.2 Introduction

Self-incompatibility (SI) is the general name given to the genetic mechanisms that prevent self-fertilization in flowering plants, thereby avoiding inbreeding depression. These mechanisms can act anywhere from pollen adhesion on the stigma to ovule fertilization (Klaas *et al*., 2011). It was first described in print by Darwin (1876) and is hypothesized to have played a major role in the angiosperm radiation of the Cretaceous period as it engenders diversity (Whitehouse, 1950). Roughly half of the flowering species exhibit SI (Hiscock & Allen, 2008). Despite its importance in creating and maintaining genetic diversity, few of the major food crops exhibit SI (Yang *et al*., 2008). In fact, plant breeders take advantage of self-compatibility (SC) to exploit hybrid vigor or simply to fix favorable alleles in elite germplasm.

The underlying genetic mechanisms of SI have been extensively explored. Classic genetic studies investigating segregation ratios revealed that two main types of SI exist: sporophytic (SSI) and gametophytic (GSI) SI (de Nettancourt, 2001). Further, a single locus is often responsible for the self-recognition that triggers the incompatible reaction of self-pollen or pollen expressing the self SI alleles. This locus is often referred to as the S-locus in all SI systems regardless of the loci's evolutionary history. Frequently, there is no relationship between the S-loci of plants belonging to different families. Despite the differences in the various SI mechanisms, the S-locus consists of two tightly linked genes that control SI. One gene is expressed in the pollen and is called the male determinant while the other gene is expressed in the pistil and is called the female determinant (Takayama & Isogai, 2005). The phenotype of the pollen grain is determined by the genotype of the diploid progenitor in SSI systems while the genotype of the haploid pollen grain determines the phenotype of the pollen grain in GSI systems. In spite of extensive research, the genes underlying this process have only been identified in a

few plant families. Three separate mechanisms have been described at the molecular level and named according to the plant family in which it was first discovered.

Brassicaceae-type SI is the only SSI system identified at the molecular level (Kachroo *et al*., 2002; Hiscock & McInnis, 2003; Takayama & Isogai, 2003). SI species within the Brassicaceae family preferentially inhibit the hydration and pollen tube growth of self-pollen and pollen from related individuals exhibiting the same SI alleles. The S-locus glycoprotein (SLG), a small glycoprotein with 12 highly conserved cysteine residues among alleles, was the first protein linked to the S-locus (Nasrallah *et al*., 1987; Takayama *et al*., 1987). Sequence similarity with this protein led to the identification of another S-linked protein coined the S-locus receptor kinase (SRK). SRK is an integral membrane protein with three distinct regions: an extracellular SLG-like domain, a transmembrane domain, and an intracellular kinase domain (Stein *et al*., 1991). A transgenic gain of function study in *Brassica rapa* later excluded SLG from being the female determinant responsible for SI but proved that SRK was indeed the female determinant, i.e. the protein expressed in the stigma that allows the plant to recognize self- and cross-pollen (Takasaki *et al*., 2000). Transgenic expression of SLG_{28} in *B. rapa* did not induce an incompatible reaction with pollen of the same haplotype, yet transgenic SRK_{28} expression did inhibit growth of pollen originating from plants with the same S-locus haplotype. Further evidence that SLG was not the female determinant came from the fact that it was not present in all genera in the family that displayed Brassicaceae-type SI (Kusaba *et al*., 2001; Suzuki *et al*., 2000; Suzuki *et al*., 2003).

The first step in deducing the male counterpart in Brassicaceae-type SI came from a bioassay (Stephenson *et al*., 1997). SI could be overcome by interposing soluble pollen coat protein (<10kDa) between self-pollen and the stigma if the extract originated from nonself-pollen.

126

The exact gene remained enigmatic for an additional two years until two groups simultaneously discovered it. One group cloned and sequenced the S-locus (Suzuki *et al.*, 1999) while the other group utilized fluorescent differential display to identify the male determinant (Schopfer *et al.*, 1999). Because of this dual discovery, the male determinant is frequently called by multiple names: S-locus protein 11 (SP11), S-locus cysteine-rich protein (SCR), or SP11/SCR. This small, cysteine-rich protein is highly variable in an S-haplotype specific manner. Alleles share as little as 19.5% amino acid identify (Shiba *et al.*, 2002). Both transgenic gain of function and bioassay studies were performed to unequivocally identify the function of this protein. The gain of function study was similar to that conducted to prove the function of the female determinant (Schopfer *et al.*, 1999; Shiba *et al.*, 2001). The bioassay entailed the production of SP11/SCR in transgenic *E. coli*. Applying the bacterially produced SP11/SCR to stigmas sharing the same S-haplotype inhibited growth of cross-pollen (Takayama *et al.*, 2000; Takayama *et al.*, 2001). The sporophytic nature of the Brassicaceae-type SI comes from the fact that SP11/SCR is expressed and secreted either mostly or exclusively in the diploid anther tapetum tissue (Takayama *et al.*, 2000; Shiba *et al.*, 2002; Iwano *et al.*, 2003).

The interaction between the male and female determinants was demonstrated by testing the strength of the interaction between transgenically produced determinant proteins through both pull-down and enzyme-linked immunosorvent assays (Kachroo *et al.*, 2001). A pull-down assay is similar to immunoprecipitation except in the pull-down assay you use a 'bait' protein instead of an antigen, the SRK protein in this case. Researchers found that the male and female determinants interact in an allele specific manner. Through a yeast two-hybrid screen, a stigma protein, arm repeat-containing protein (ARC1), was found to interact with the intracellular serine/threonine kinase domain of SRK. Upon activation by the self-allele of SP11/SCR, SRK

phosphorylates ARC1 (Gu *et al*., 1998; Muzzurco *et al*., 2001). Antisense downregulation of ARC1 showed a partial loss of SI, suggesting either that an alternative pathway exists or residual ARC1 activity maintains the SI pathway (Stone *et al*., 1999). Four lines of evidence suggest that ARC1 acts to ubiquinate proteins necessary for pollen growth: sequence information suggests ARC1 has ubiquitin ligase activity, ARC1 is concentrated near proteasome/COP9 signalosome when SRK is activated, levels of ubiquinated protein increase only after self-pollination, and proteosome inhibition disrupts the SI response (Stone *et al*., 2003). Other molecules have been implicated in the downstream reactions as well and can be reviewed in Ivanov *et al*. (2010), but the current evidence suggests that pollen tube growth is inhibited upon recognition of self-pollen by ubiquitin mediated degradation of essential pollen growth proteins. A summary figure of the Brassicaceae-type SI system from Takayama and Isogai (2005) has been included (Figure 4.1).

In contrast to the Brassicacaea-type SI, the Solanaceae-type SI behaves gametophytically. Although Solanaceae, Rosaceae, and Scrophulariaceae families all share a common female S-determinant, it is referred to as Solanaceae-type SI because it was first described in *Nicotiana alata*, a member of the Solanaceae family (Takayama & Isogai, 2005). Pollen tube growth of self-pollen or pollen expressing the same SI alleles is inhibited in the style of plants in the three aforementioned families. Like the other molecularly described SI systems, the Solanaceae-type SI system is controlled by a single locus (S-locus) containing two tightly linked genes.

The female determinant was identified as a style glycoprotein (~32 kDa) that segregated with the S-locus using sodium dodecyl sulfate polyacrylamide gel electrophoresis (Anderson *et al*., 1986). Subsequent sequence data analysis revealed homology to a fungal ribonuclease, and testing for ribonuclease activity in vitro confirmed the role of the female determinant as a ribonuclease (McClure *et al*., 1989). Hence, it became known as S-RNase. The S-RNase was

found to be exclusively expressed in the pistil with the highest expression in the distal portion of the style where self-pollen inhibition occurs. Gain and loss of function experiments eliminated any residual doubt that the S-RNase protein was indeed the female determinant underlying Solanaceae -type SI (Lee *et al.*, 1994; Murfett *et al.*, 1994). Again, high allelic diversity was observed for S-RNase; the amino acid sequence similarity between alleles ranges from 38 to 98% in the Solanaceae family (McCubbin & Kao, 2000). One or more hypervariable regions appear to impart haplotype specificity and interact with the male S-determinant (Matton *et al.*, 1997). Additionally, high expression levels of S-RNase are required to convey SI (Huang *et al.*, 1994).

The male S-locus determinant in the Solanaceae-type SI system was difficult to isolate and remained a mystery until genomic analysis in the early 2000s (Ushijima *et al.*, 2003; Yamane *et al.*, 2003). Prior to its discovery it was proposed that the male determinant was an RNase inhibitor because S-RNase was known to enter pollen tubes regardless of their SI genotype (Luu *et al.*, 2000). The genomic sequence surrounding the S-locus was found to have four different F-box genes, but only one of them fit the requirements of the male S-determinant. An F-box gene contains a structural motif, approximately 50 amino acids long, involved in protein-protein interactions with the E3 ubiquitin ligase complex (Bai *et al.*, 1996; Skowyra *et al.*, 1997). Simultaneous discovery of the candidate gene thought to be responsible for the haplotype specific male determinant was made by two groups working on two different species within the *Prunus* genus. This ultimately led to two names: S-locus F-box (SLF) and S-haplotype-specific F-box (SFB; Entani *et al.*, 2003; Ushijima *et al.*, 2003; Yamane *et al.*, 2003). Again, transgenic experimentation led to the final proof of function. In this case, competitive interaction was utilized (Sijacic *et al.*, 2004). Competitive interaction often occurs in tetraploid plants. This is where heteroallelic pollen, pollen expressing two different S-alleles, fails to function in SI (de

Nettancourt, 2001; Entani *et al*., 1999). Hua and colleagues (2007) discovered that SLF/SFB was localized in the cytoplasm of growing pollen tubes.

As mentioned previously, it was believed that SLF/SFB was an RNase inhibitor as both self and non-self S-RNases localized to the cytoplasm of growing pollen tubes (Luu *et al*., 2000). Later studies contradicted this finding, suggesting that compartmentalization might play a role in the selective degradation of non-self S-RNases (Goldraij *et al*., 2006; Meng *et al*., 2009). Meng and colleagues (2009) used GFP tagged S-RNases to visualize the female determinant in the growing pollen tubes. While compartmentalization was evident, it was not allele specific and the transgenic expression of S-RNases did not affect incompatibility, implying that the S-RNases must be in the cytoplasm to properly function, which is sensible considering the location of the male determinant. The main evidence supporting the selective degradation of non-self S-RNases stems from the properties of SLF/SFB. F-box proteins are generally components of Skp1-Cullin-F-box (SCF) complexes, which function to target proteins for degradation by 26S proteosomes (Hershko & Ciechanover, 1998). The F-box domain binds the F-box protein to the SCF complex while a second domain specifies which proteins to ubiquitinate (Tyers & Jorgensen, 2000); therefore, it is reasonable to conclude that SLF/SFB preferential binds non-self S-RNases and targets them for proteosomal degradation. Be as that may, no other protein interacting domain has been identified on SLF/SFB (Zhao *et al*., 2010). The exact interaction is still unclear, and it may not be consistent across all taxa with Solanaceae-type SI (Meng *et al*., 2011).

Additional evidence supporting the selective degradation of nonself S-RNases exists. *In vivo* style S-RNase content decreases after a compatible pollination while the content remains steady in unpollinated and self-pollinated styles (Liu *et al*., 2009). Further, the ubiquitinated protein content of compatible pollinations increases relative to incompatible crosses after

130

pollination. Proteosome inhibitors also restrict the growth of compatible pollen (Qiao *et al*., 2004a, Qiao *et al*., 2004b). Lysine to arginine mutations near the C-terminal end of S-RNase also reduced the amount of ubiquitination (Hua & Kao, 2008). Lastly, protein binding assays revealed allele specific differential protein binding, with self S-RNase showing the weakest affinity to SLF/SFB (Hua & Kao, 2006). In summary, Solanaceae-type SI is accomplished in the style through the degradation of pollen ribosomal RNA by S-RNase. Non-self S-RNases are targeted for destruction through the ubiquitin mediated proteosomal pathway prior to exerting their effect on pollen growth. A summary figure of the Solanaceae-type SI system from Takayama and Isogai (2005) has been included (Figure 4.2).

The final well characterized SI system to discuss was discovered in *Papaver rhoeas* (field poppy); therefore, it is referred to as Papaveraceae-type SI. The SI phenotype is controlled gametophytically with two tightly linked genes at what is referred to as the S-locus. Self-pollen tube growth is inhibited almost immediately in this SI system. The female determent was identified by testing the ability of *E. coli* produced papilla proteins to inhibit pollen growth in an allele specific manner via in vitro bioassays (Foote *et al*., 1994). It was determined that the protein responsible was a small, secreted protein lacking distinct hypervariable regions within the protein. Instead allelic variation is spread throughout the protein. Alleles are highly divergent, generally sharing between 40-50% amino acid similarity (Paape *et al*., 2011). The female determinant was later named *Papaver rhoeas* stigma S determinant (PrsS).

Despite knowing the female determinant and much about the downstream process of pollen inhibition, the male determinant in the Papaveraceae-type SI system remained a mystery for 15 years. Finally, Wheeler and colleagues (2009) were able to identify the male determinant by analyzing a 42-kilobase cosmid clone containing PrsS. An open reading frame was discovered

a mere 457 base pairs from the stigma determinant that was expressed specifically in the pollen, only upon pollen maturation, and contained hydrophobic regions expected of an integral membrane protein. Southern blotting revealed that the gene was present as a single copy, and multiple alleles were cloned in a population segregating for SI. The *Papaver rhoeas* pollen S determinant (PrpS) contained similar allelic diversity as the stigma determinant. Antisense oligonucleotides of PrpS interfered with the SI response, thereby proving the function of PrpS as the male determinant in the Papaveraceae-type SI.

When PrpS expressed in the pollen membrane matches one of the two PrsS alleles secreted from the stigma then a very rapid influx of calcium ions occurs 50 µm from the growing pollen tube tip in the shank region of the pollen tube (Franklin-Tong *et al*., 2002). The oscillating high concentration of calcium at the rapidly growing pollen tip is very important for pollen tube expansion (Takayama & Isogai, 2005). The loss of the calcium gradient results in a rearrangement of the actin cytoskeleton (Geitmann *et al*., 2000). Additionally, p26, a cytosolic pollen protein, is phosphorylated in a calcium dependent manner and is believed to play a role in downstream signaling of the SI response (Franklin-Tong & Franklin, 2003). Thomas and Franklin-Tong (2004) provided evidence that the SI response in Papaveraceae happens on two levels. The initial inhibition of pollen growth is reversible in the first 45 minutes; after the initial time period, the reaction becomes irreversible through the activation of programmed cell death. A summary figure of the Papaveraceae-type SI system from Takayama and Isogai (2005) has been included (Figure 4.3).

The diverse mechanisms responsible for SI across the various plant families discussed herein suggests that this trait evolved multiple times in the various plant lineages (Takayama & Isogai, 2005). The exact number of times is unknown as many plants exhibiting SI have yet to be

132

studied. Comparing the known mechanisms does provide insight that may aid in studying novel SI systems. First, closely related plants tend to share the same SI mechanisms (Hayman & Richter, 1992). In general, plants within the same family share an SI system, yet variation may still exist within a particular type of SI system. Also, related, self-compatible species may retain the downstream signaling genes and non-functional determinant genes responsible for the SI response. All of the loci responsible for SI have two tightly linked genes that appear to have co-evolved: one gene being expressed in the pistil and another expressed in the stamen. The male and female determinants show very high allelic diversity (Takayama & Isogai, 2005). Despite many of them being glycoproteins, the sugars do not appear to be necessary for allele recognition. Finally, the determinant genes generally come from large families of genes (Vanoosthuyse *et al.*, 2001; Gagna *et al.*, 2002). While previously undiscovered SI systems will likely prove to be unrelated to the known systems, these common themes should prove useful in the effort to explain new SI systems, and many of the methods utilized to study the Brassicaceae-, Solanaceae-, and Papaveraceae-type SI systems should be applicable in the search for novel SI systems.

Despite its importance, much less is known about Poaceae-type SI. Watson and Dallwitz (1992) estimated that the Poaceae family is comprised of roughly 10,000 species, making it the fourth largest family of flowering plants and the second largest among monocotyledons (Watson, 1990). Additionally, it contains all of the cereals and many of the economically important forage species (Langridge & Baumann, 2008). SI within this family was first documented well over a century ago in rye grass by Rimpau (Lundqvist, 1954). SI is known to be common within Gramineae (Connor, 1979) but is more common among perennial species (Körnicke, 1890). Grass pollen is characteristically tricellular and short-lived with rapid pollen inhibition upon

incompatible pollen-stigma interactions (Shivanna *et al.*, 1982). At least four of the six

subfamilies contain species that exhibit SI (Langridge & Baumann, 2008), but much of the

research has been conducted on species within Pooideae, a subfamily that is highly divergent

from Panicoideae, Chloridoideae, and Arundinoideae (Figure 4.4; Kellogg, 1998).

A contributing factor to the slow progress in identifying the mechanisms underlying SI in

the grass family is that the most economically important grasses as well as the model grass

systems are self-compatible. *Oryza barthii*, sometimes referred to as *O. perennis* subsp. *barthii*,

is a partially SI, wild relative of *O. sativa*, one of the model grass species (Nayar, 1967; Chu *et

al.*, 1969). As such, it would seem as if it would be a particularly suitable candidate system to

further elucidate the genetic mechanisms of grass SI. In his unpublished thesis, Nayar (1958)

describes a population of *O. barthii* that failed to set seed when selfed. Further testing revealed

only partial SI as selfing led to 15-43% seed set, slightly higher than the natural selfing rate of

12-29%. Differences between pollen tube growth and morphology of self- and cross-pollen

further established *O. barthii* as partially SI (Chu *et* al., 1969). Whether these observations are

the result of the remnants of the grass SI system or another mechanism are yet to be determined.

Much of the research since these initial studies has shifted away from studying SI and towards

the study of these wild relatives' ability to promote outcrossing through physical flower

morphology and temporal barriers to improve hybrid seed development (Oka & Morishima,

1967; Virmani & Athwal, 1973; Kato & Namai, 1987; Uga *et al.*, 2003; Marathi *et al.*, 2015).

Studies on *Secale cereal* and *Phalaris coerulescens* by Lundqvist (1954) and Hayman

(1956), respectively, revealed the gametophytic nature of Poaceae-type SI. GSI is evident when a

fraction of the pollen is incompatible in a single cross, i.e. if only half of the pollen in a cross is

compatible, then SI must be controlled by the haploid genome of the pollen grain. Further, a

cross with 75% pollen compatibility is evidence of a two-locus SI system. While differences in percent compatibility between reciprocal crosses are more likely to occur under a two-locus system, this in itself is not proof of a two-locus system as this scenario can occur in a one-locus system as well. Both Lundqvist (1954) and Hayman (1956) discovered that two unlinked, multi-allelic loci, termed S and Z, were controlling the SI phenotype observed in populations of their respective plants. Both the S and Z alleles expressed in the pollen grain have to match stigma alleles in order to trigger an incompatible reaction. As of 2008, eight species had been confirmed to exhibit the same SI system (Langridge & Baumann, 2008). Another feature of the grass SI system is that polyploid plants retain their SI. Lundqvist (1957) also noticed that homozygosity at one of the SI loci decreased the amount of selfed seed produced. Further, some environmental conditions played a role in the amount of leakage in the SI system. Under high temperatures and humidity, selfed seed production in rye could reach as high as 25% compared to natural rates near 1% (Wricke, 1978; Wilkins & Thorogood, 1992).

A third locus (T) was suggested to be involved in SI by screening self-compatible rye mutants (Lundqvist, 1968). This locus was later confirmed and identified in other species as well (Thorogood & Hayward 1991; Hayman & Richter 1992; Voylokov *et al*. 1993; Thorogood *et al*. 2005), but no allelic variation appears to be present at the T-locus. Therefore, the role of the T-locus is difficult to parse out. It may be essential to the formation or interaction of the S and Z determinants or the downstream signaling. Alternatively, it may be a relic of a larger self vs. non-self recognition system where one allele became fixed. If the latter is true, the mutation would represent a loss of function allele (Hayman, 1992). Interestingly, the pistils of self-fertile mutants maintain their SI, i.e. plants with this mutation can still inhibit pollen with the same S and Z

alleles if the self-fertile allele is not present in the pollen. This suggests that only a single functional copy of the T-locus need be present for SI to properly function.

Biochemical assays have provided some insight into the underlying mechanisms involved in Poaceae-type SI. Wehling *et al*. (1994) observed in rye that protein phosphorylation levels were higher in pollen tubes treated with incompatible stigma extract. Differential protein screening based on phosphorylation levels between compatible and incompatible reactions led to the isolation of four proteins. Wehling and his colleagues (1994) also tested the effect of kinase inhibitors and calcium channel inhibitors on pollen growth. They found that Lavendustin A, a tyrosine kinase inhibitor, and verapamil, a calcium channel blocker, almost completely eliminated the SI response to self-pollen. In contrast, differential phosphorylation was not found in other species or in self-fertile mutants (Baumann, 1995); therefore, signaling cascades based on phosphorylation are unlikely to play a role in the initial SI response although not enough evidence has been accrued to completely rule it out. Calcium signaling may yet prove to be critical to grass SI.

Differential screening of pollen proteins from contrasting genotypes of *P. coerulescens* failed to yield any candidate proteins (Tan & Jackson, 1988). Li *et al*. (1994) attempted to screen pollen RNA, also using *P. coerulescens*. This yielded a candidate gene for the S-locus but failed to identify any Z-locus candidates. Unfortunately, it was later determined that the candidate gene was not the S-determinant (Langridge *et al*. 1999). Yang *et al*. (2009) constructed subtractive hybridization cDNA libraries to identify gene candidates. Since pollen inhibition acts so quickly, it was hypothesized that the molecules involved in the initial steps were already synthesized. Therefore, cDNA from immature stigmas was subtracted from mature stigma cDNA and cDNA extracted almost immediately after an incompatible pollination. Nine genes were found to be

exclusively expressed in stigmas, be differentially expressed between mature and immature stigmas, and map to the S- or Z-locus. Two in particular are promising candidate genes for the female determinant of the S- and Z-loci. Both have calcium dependent protein kinase domains and are maximally expressed two minutes after pollination and decline ten minutes post pollination. To the author's knowledge, no further work regarding these candidates has been published.

Genetic mapping of the SI loci has been accomplished in many grass species. Both the S- and Z-loci were mapped in the 1980s (Cornish *et al*., 1980; Gertz & Wricke, 1989). The S-locus was mapped on chromosome one in *Phalaris*, which is syntenic to rice chromosome five between 2.7 and 4.8 Mb on the physical map (Bian *et al*., 2004). More recently, AFLP markers generated from pollen and pistil transcripts were used to fine map the S-locus in *Hordeum bulbosum* (Kakeda *et al*., 2008). Two transcripts were found to be completely linked to the S-locus but only one transcript represents a potential candidate as the other lacks sufficient allelic variation. Using one of the completely linked transcripts, Kakeda (2009) isolated two F-box genes using rapid amplification of cDNA ends. Similar to studies in *Lolium* (Shinozuka *et al*. 2007) and *Phalaris* (Bian *et al*. 2004), the centromeric location of the S-locus in *Hordeum* results in the suppression of recombination, making map based cloning difficult. Extensive mapping was recently reported in *L. perenne* (Manzanares *et al*., 2012). Using 10,000 genotyped individuals, the researchers were able to reduce the genetic map region to 0.07 cM. The syntenic region in rice only contains six genes. Manzanares (2013) was able to identify four stigma S-locus gene candidates and six pollen S-locus gene candidates.

The Z-locus maps to chromosome two of rye and *Phalaris*, which is syntenic to rice chromosome four between 32.3 and 32.7 Mb on the rice physical map (Hackauf & Wehling,

137

2005). Despite the relatively small syntenic region in rice, the resolution of the genetic maps surrounding the Z-locus is much lower, spanning 1.5 cM due to the differences in recombination rates surrounding the loci. In an attempt to hone in the Z-locus, comparative genomics was utilized to develop locus specific markers. Using publically available sequence information and the conserved synteny present in Poaceae, Shinozuka *et al*. (2010) was able to develop PCR markers specific to the Z-locus. Screening *L. perenne* BAC clones with these locus specific markers allowed researchers to identify three clones that contained Z-locus DNA. The subsequent sequencing of the aforementioned BAC clones led to the identification of two candidate genes, but functional studies are necessary to prove the role of these genes in SI.

To the author's knowledge, no studies have attempted to describe the SI mechanism operating in *Miscanthus*. In fact, few studies have been reported on the mechanism operating within the Panicoideae subfamily of Poaceae, which *Miscanthus* resides. The single study cited in Langridge and Baumann (2008) suggesting that this subfamily contained the S-Z SI system common in other grasses was conducted on *Panicum virgatum*, and the focus of the study was on post-fertilization barriers involved in interploidy crosses as opposed to pre-fertilization barriers (Martinez-Reyna & Vogel, 2002). Differences in reciprocal seed set were taken as evidence of reciprocal differences in pollen compatibility, thereby providing evidence for a two locus SI system. Even when a fraction of the pollen is incompatible, more than enough compatible pollen is present to fertilize ovules; therefore, the reciprocal differences in seed set likely had more to do with the female plants genetic predisposition to set seed or environmental factors than the percentage of compatible pollen present. Little is known about the SI mechanism operating in *Miscanthus* or the subfamily it belongs to.

The objectives of this chapter are to explore the genetic mechanisms underlying SI in *Miscanthus* and map the locus/loci responsible. Seeing as no information regarding this topic is currently available for *Miscanthus* and very little information is available for close relatives, only the broadest assumptions will be made at the outset of this research. The syntenic relationship between *Miscanthus* and sorghum, maize, and rice will be utilized when possible in comparative genetics.

4.3 Materials and methods

Plant material

Two biparental mapping populations were phenotyped for SI. A description of both follows. Two ornamental varieties of *M. sinensis*, 'Grosse Fountaine (GF)' and 'Undine (UN),' originally purchased from local nurseries were reciprocally crossed in isolation in the Plant Science Laboratory (PSL) greenhouses at the University of Illinois at Urbana-Champaign in the fall of 2009. Recently, Clark *et al*. (2014) determined through marker data that both accessions ancestry can be traced to southern Japan. Seeds collected from both parents were planted the ensuing winter in the aforementioned greenhouse. The rhizome of the seedlings was split in order to generate three clonal replicates of each progeny. One week prior to transplanting, the plugs were moved to outdoor ground beds to harden off. Just prior to transplanting the plugs, leaf tissue was cut back to two to four cm to minimize transpiration while the root system developed. In total 221 replicated progenies along with each of the parents were transplanted on 1.52 meter centers in May of 2010 at the Energy Biosciences Institute (EBI) farm in Urbana, IL in a randomized block design. The plot was not watered nor fertilized in 2011 or 2012 although it was watered immediately after transplanting in 2010 and again in the fall of 2013 to promote flowering. 2,4-D was applied to control broadleaf weeds and atrazine and Medal® II AT were

applied as pre-emergent herbicides at the beginning of each growing season. All three were applied at 0.18 kg ha^{-1}. This population will henceforth be referred to as the GFxUN population. A more detailed description of this mapping population can be found in Swaminathan *et al*. (2012), Gifford (2012), and Gifford *et al*. (2014).

A second mapping population was partially phenotyped for SI. This mapping population was also created from two publicly available, ornamental cultivars: *M. sinensis* 'Strictus' and *M. sinensis* 'Kaskade.' These plants were crossed in isolation in the PSL greenhouses at the University of Illinois at Urbana-Champaign in the fall of 2010. 294 individuals grown from seed harvested from *M. sinensis* 'Strictus' were planted on the EBI farm in Urbana, IL in a Latinized row-column design consisting of three replicates of each progeny in the spring of 2011. Marker analysis revealed that 33 of these progeny were the result of self-pollination despite the normally strong SI mechanism operating in *Miscanthus*. A more detailed description of this mapping population (STxKA) can be found in Liu (2015) and Liu *et al*. (2015).

In order to more completely assess the number of loci controlling the SI response in *Miscanthus*, pollen tube growth was investigated in a diallel cross of 12 diverse lines. A cross where 75% of the pollen is compatible could only be possible if multiple, unlinked loci were involved in regulating SI. Due to dry, unseasonably cool conditions, investigation of the full diallel was unable to be completed. A list of the accessions included in the diallel design as well as their origins can be found in table 4.1.

In vitro pollination assay

Although many steps in this assay are time dependent, there is some flexibility in the timing. The times reported are the ones used herein. This protocol was developed by Shui-Zhang Fei and Jianxiong Jang while visiting Shelia McCormick's lab at the University of California-

Berkeley and was only slightly modified to account for the climate in central Illinois. A diagram

of the *Miscanthus* flower anatomy has been included to reduce confusion with terminology

(Figure 4.5).

Two to three racemes per genotype tested provided plenty of mature pistils for testing

pollen incompatibility by imaging pollen tube growth. Raceme collection can be done any time

in the afternoon, but two p.m. is ideal. Only racemes that have open florets at the apex and closed

florets near the proximal end should be selected (Figure 4.6B). This ensures the availability of

mature pistils as *Miscanthus* racemes mature from the distil to proximal end. The developmental

stage of the pistils is critical to accurate phenotyping. Pistils that are directly below the open

florets tend to be mature. The stigmas of mature pistils tend to be purple in color and appear

turgid although the color is genotype and population dependent. Some populations have white or

yellow stigmas at maturity. Immature pistils often lack the ability to discern self-pollen from

cross-pollen. Immature pistils tend to be white or light purple in color. Overly mature pistils can

often result in false results as pollen growth tends to be inhibited regardless of SI alleles. Overly

mature pistils have stigmas that are often very dark, almost black in color, and lack the feathery

look of fresh stigmas. This occurs when florets fail to open at maturity, often due to water or heat

stress. Open florets were avoided as the pollen source would not be controlled. Immediately after

collection, racemes were placed in numbered test tubes with approximately one to two cm of

water to maintain hydration of the florets. The genotype corresponding to each respective

number was carefully recorded as there is no way to reliably distinguish racemes or

inflorescences after they are removed from a plant.

The pistil can be dissected from the spikelet at any time prior to *in vitro* pollination (one

to two a.m.), but contamination with self-pollen increases with time as anthers begin to dehisce

late in the evening. The pistil can be dissected by cutting the base of the spikelet with a sharp razor blade and separating the lemma and palae with fine tweezers. Alternatively, the pistil can be removed from the spikelet by gripping the spikelet between the thumbnail and the index finger and squeezing out of the pistil. In addition to choosing mature stigmas from closed spikelets, care must be taken to choose intact pistils. Dissecting pistils from the spikelet is generally the limiting step in evaluating pollen tube growth as dissection can be rather time consuming. Dissected pistils were immediately placed onto boric acid plates (Figure 4.6C). The ovary should be in contact with the medium, but the stigma can either be upright or laying on the medium as long as it is not pressed into the medium. Pistils dissected from multiple genotypes can be combined into a single petri dish when appropriate, i.e. when multiple genotypes are being tested with a single pollen tester, by labeling the bottom of the petri dish with the corresponding test tube number from which the pistil originated. The boric acid medium consists of 25 ppm boric acid, 1% sucrose, and 1% agarose dissolved in ddH_2O. A microwave was used to dissolve the agarose. The size of petri dish is dependent on the experiment. Four cm petri dishes are suitable for ~50 pistils.

Multiple inflorescences from an individual plant were collected around five p.m. for pollination purposes. There is some flexibility in the timing of the collection. Often, it is best to collect the inflorescences immediately after the racemes have been collected to avoid multiple trips to the field. When temperatures are warmer, it is best to collect the inflorescences later in the day. Also, cool, cloudy days generally result in poor pollen shed. Only inflorescences that are actively undergoing anthesis should be collected. The best inflorescences are those with half of their anthers already dehisced. When collecting inflorescences the stem should be cut at least six inches below the lowest raceme and labeled immediately. The collected inflorescence should

then be vigorously shaken to shed any dead pollen and passed through your lightly closed hand to remove any spent anthers prior to placing it into water. The inflorescences should be placed in the dark after collection.

Pollination of boric acid plates containing dissected pistils should be done between one and two a.m. The timing of pollination is critical as the percentage of viable pollen decreases as the length of time that the anthers have been exuded from the spikelet yet have not dehisced increases. In order to pollinate the pistils, an inflorescence should be shaken a few inches above the boric acid plate (Figure 4.6D). A pollen cloud should be visible (Figure 4.6E). Care should be taken to avoid pollinating in drafty rooms as this can disperse the pollen cloud, potentially leading to fewer pollen grains per stigma than necessary to determine the percentage of compatible pollen. Additionally, the inflorescence should not be allowed to touch the boric acid plate as this results in the removal or scattering of pistils. The desired number of pollen grains per stigma is dependent on the experiment. If genotypes are simply being placed into incompatibility groups, then the number of pollen grains is less important as one only needs to distinguish between 0%, 100%, and any other level of pollen tube growth. On the other hand, if the exact percentage of compatible pollen in a cross is being investigated, the level of pollen should not be so great as to hinder counting of individual pollen grains. The number of pollen grains can be adjusted by either increasing the number of inflorescences used per plant or increasing the distance between the inflorescence(s) and the boric acid plate.

After pollination, the petri dishes holding the pollinated pistils were sealed with parafilm and kept in the dark at room temperature for approximately five hours to allow pollen tube growth. Afterwards, the pistils were transferred to 96 well plates containing 300 µl of 10 M NaOH for softening (Figure 4.6F). This step halts any further pollen growth, makes the pistil

143

more amenable to squashing under microscope cover slips, and facilitates die penetration. Care was taken to transfer the pistils in an orderly fashion as the well number is the only identifier at this point. After four hours of softening, the pistils appear mostly colorless. They should then be transferred to 96 well plates containing 300 µl ddH$_2$O for rinsing, but care should be taken to limit disturbing the pistils as this will cause incompatible pollen to be washed away. As soon as all of the pistils were transferred to the rinsing well, they were moved to clean microscope slides containing 60 µl 0.1% aniline blue dye dissolved in 0.07 M sodium phosphate buffer (pH=9). Cover slips were placed over the pistils, and the slide was placed in a microscope slide box (Figure 4.6G-H). Again, care was taken to maintain the order of the slides as no additional label is present. The microscope slides were incubated for 12-24 hours at 4°C in a non-humidity controlled cooler.

To visualize pollen tube growth, a Zeiss Axiovert 200M fluorescent microscope was used with a DAPI filter. Observations of pollen tube growth were made using either a 5X or 10X objective. To decrease the time spent imaging pollen tube growth, the percentage of compatible pollen was recorded without saving an image in the majority of cases. Images were saved for later analysis when exact counts of compatible vs. incompatible pollen were desired.

Mapping population strategy

The investigation of both biparental populations had two initial goals: to determine how many SI loci were segregating and to determine how many alleles were segregating at each locus. Since *Miscanthus* is an outcrossing genus, there are potentially four alleles segregating at each locus—one from each grandparent. To accomplish these goals, it was necessary to determine how many SI groups were present in each population. Individuals in the populations were placed into the same group if they were completely incompatible with each other, meaning none of the

pollen from one plant could grow down to the ovule of another. The easiest way to determine how many SI groups there were in a population was to pollinate pistils from a large number of individual plants with a single pollen donor. Then it becomes a basic genetics problem. Pistils from 20-30 randomly chosen individuals within the population were initially tested against pollen from a single randomly chosen tester. The number of plants incompatible with the pollen tester was then compared to all possible scenarios, keeping in mind that these scenarios differ based on whether the populations were created from a single or bi-directional cross (Figure 4.7). If there were only two SI genotypes segregating in a population, one would expect 15 of the 30 individuals tested to be completely incompatible with the pollen tester.

Subsequent to determining the number of loci and alleles segregating in the GFxUN population, the remaining individuals that had not been placed into an SI group were screened using pollen from each of the groups. No additional individuals were screened in the STxKA mapping population as the single locus segregating was determined to be the same locus as identified in GFxUN population.

Statistical analysis

Chi-square goodness of fit tests were performed in excel to determine the number of genotypes segregating in the two biparental populations. Student's t-tests were performed using SAS® 9.3 software (SAS Institute Inc., Cary, NC) to differentiate the percentage of compatible pollen in the diallel crosses. QTL analysis was performed using the SI group number as the phenotype with MapQTL 5.0^{TM} (Van Ooijen, 2004) using the cross pollinated (CP) population type and a maximum likelihood mixture model. The permuted, genome-wide significance LOD threshold values based on 1,000 permutations were used to declare a QTL significant (Churchhill & Doerge, 1994). Initially, an interval mapping approach was taken using a step size of 1 cM.

The QTLs found in the initial search were then used in an approximate multiple-QTL model approach to further refine the peak QTL position (Jansen 1993, 1994; Jansen & Stam, 1994). Confidence intervals were calculated based on the 2-LOD dropoff method (Lander & Botstein, 1989). The genotyping of the population and genetic map used in the analysis of the GFxUN population were described previously (Swaminathan *et al*., 2012; Gifford *et al*., 2014). The genotyping of the STxKA population and genetic map used in the analysis of the population were also described previously (Liu, 2015; Liu *et al*., 2015).

4.4 Results

Grosse Fountaine x Undine

Repeated screening of pistils from 20 randomly chosen individuals with one randomly chosen pollen donor from the GFxUN population consistently yielded approximately five individuals that were 100% incompatible with the pollen donor (Figure 4.8A). The pollen donor and the individuals that were 100% incompatible with it were placed into the same SI group. By default, the first pollen tester and the individuals that were incompatible with it were placed into group one. The next pollen tester was then chosen from the group of plants that were not placed into the first group. In this manner, individuals belonging to four separate SI groups were identified. Further, SI reactions were consistent when testing the same 20 individuals with different pollen donors from the same group. Pollen from each group (1-4) was then used to test the remaining unclassified plants in the population. To do this, pistils from plants being tested were dissected and placed onto four separate petri dishes; each petri dish was then pollinated by plants belonging to a different group (1-4).

Of the 221 individuals in the population, 194 were able to be placed into one of the four SI groups during the period in which the population flowered in 2013 (Table A.4). Only one of

the 27 unclassified individuals could not be placed into one of the four groups. The remaining 26 were untested due to time constraints. The single individual, UG031, not fitting into one of the groups was tested for its ability to inhibit self-pollen in 2014. Self-pollen growth of UG031 showed the standard incompatible phenotype. By testing each plant against all four groups, the relationship between the groups became apparent. Pollen from group one was always 100% incompatible with plants in group one but 100% compatible with plants in group three and vice versa (Figure 4.9A). The same relationship held for groups two and four. 50% of pollen from group one and three grew on stigmas from plants in groups two and four and vice versa. In addition to replication, these relationships were exploited to increase the accuracy and confidence of placing individuals into groups.

Using the group number as the phenotype of the 194 individuals, a single QTL was identified on LG 15 that far exceeded the permuted LOD threshold of 4.3 (Figure 4.10A). The peak LOD score is at 58.5 cM between markers EBI 686 and EBI 687; the 2-LOD confidence interval ranges from 57.5 to 61.7 cM. This particular region on LG 15 is not densely covered with markers. The closest markers to the SI QTL have a recombination frequency between 5.1 and 7.7 (Table 4.2). The corresponding position of the SI QTL on sorghum, an SC relative, is on linkage group eight between base pairs 49,763,305 and 53,137,669 on the sorghum physical map.

Interestingly, markers EBI 681-686 could not be placed onto the sorghum physical map. The closest flanking markers—EBI 680 and 687—correspond to 49.8 and 53.1 Mb of chromosome eight on the sorghum physical map, respectively. Of the 846 markers used to create the GFxUN genetic map, only 93 could not be assigned to a unique location on the sorghum genome. Additionally, 62 of the 93 markers that could not be placed occur in isolation, meaning that they are flanked by markers that could be placed onto the sorghum genome. Only eight

unplaced marker pairs, a single run of three consecutive unplaced markers, and two stretches of

six unplaced markers were found to occur consecutively in map order (Figure 4.11). A goodness

of fit test was conducted to determine whether or not the markers that could not be placed onto

the sorghum genome were randomly distributed along the *Miscanthus* genetic map according to a

Poisson distribution function (Isemura *et al*., 2007). The null hypothesis that the unplaced

markers were distributed randomly was rejected with a P-value <0.001. In addition, a simulation

was ran in R to see how often by chance we would expect to see a run of six consecutive

unmapped markers given that 93 of 846 markers were unplaced. Six or more consecutive

unplaced markers occurred in 98 of 100,000 permutations. In none of the 100,000 permutations

did two such runs of six or more consecutive unplaced markers occur in a single permutation.

The conserved synteny within the grass family can be used to compare the location of the

previously mapped S- and Z-loci operating in the Pooideae subfamily. These loci have been

mapped to the fifth and fourth LG, respectively, of rice (Wricke & Wehling, 1985; Gertz &

Wricke, 1989). The fifth rice chromosome corresponds to the third and ninth LG of sorghum

while the fourth chromosome corresponds to the fourth and sixth LG of sorghum. The T-locus

corresponds to rice chromosome 12 between 19.9 and 23.2 Mb (Manzanares, 2013), which is

homologous to sorghum chromosome eight between 44.1 and 48.1 Mb (Figure 4.12). This region

is extremely close to the location of the SI QTL mapped herein. Additionally, a deletion appears

to have occurred in sorghum between 45.63 Mb and 47.19 Mb (Figure 4.12). Evidence for a

deletion in sorghum compared to corn also exists in one homologous region but not the second

homologous region (Figure 4.13). When comparing the genomes of corn and rice, a deletion is

evident in one homologous chromosome but not the other (Figure 4.14).

Strictus x Kaskade

Screening the same 30 randomly chosen individuals with multiple randomly chosen

pollen donors from the STxKA population allowed all 30 individuals to be placed into one of two

SI groups. The pollen donor and the individuals that were 100% incompatible with it were placed

into the same SI group. By default, the first pollen tester and the individuals that were

incompatible with it were placed into group one. Fourteen individuals showed a group one

phenotype while 16 showed a group two phenotype. This data is consistent with the segregation

of a single locus and two genotypes (Table 4.3). Half of the pollen from group one was viable on

pistils belonging to group two individuals and vice versa (Figure 4.9).

Using the group number as the phenotype of the 30 tested individuals, a single QTL was

identified on LG 15 that far exceeded the permuted LOD threshold of 4.3 (Figure 4.10B). The

peak LOD score of 26.0 coincides with the map position of 40.6 when the original map produced

by Liu (2015) is scaled down to approximately the same scale as the map produced with the

GFxUN population. The nearest marker to the peak LOD score is TP95523. Unfortunately, not

all of the markers on LG 15 can be used due to singularity error. Singularity error, the case when

multiple solutions exist to a set of mathematical equations, occurs when all of the markers are

included in analysis. This error commonly occurs when segregation distortion is so extreme that

there is a complete lack of one genotype in the population and when all of the markers segregate

in a single parent (Van Ooijen, 2004). Due to this limitation, the markers were separated and

their coding switched to that of the DH population type. QTL mapping was then carried out on

the two separate maps. The *M. sinensis* 'Strictus' map contained 54 markers on LG 15 while LG

15 contained only eight markers on the *M. sinensis* 'Kaskade' map (Figure 4.10C-D). Using the

'Strictus' map under the DH population type the peak LOD for SI was found to be nearest

marker TP37250 at 50.2 cM. Similarly, the 'Kaskade' DH map resulted in a peak LOD near marker TP32438 at 23.5 cM.

Diallel

The diallel could not be completed due to unfavorable weather conditions for flowering in the fall of 2014. Four of the individuals involved in the diallel cross were tested as male parents. The majority of crosses resulted in the 100% pollen compatibility (Table 4.4). The ten replicates were tested for differences between crosses known to be 50% and 100% compatible using a Student's t-test. The mean compatibility of the crosses known to be 50% compatible was 51.2% while the mean of the ten crosses known to be 100% was 85.3%. Two crosses, *M. sinensis* 'Giraffe' x *M. sinensis* 'Positano' and *M. sacchariflorus* 'PMS-071' x *M. sinensis* 'Positano,' resulted in statistically significant differences between 50% and 100% compatible reactions and had means of 74.5 and 74.1, respectively. Two additional crosses, *M. sinensis* 'NC-2010-004-B(37)-d' x *M. sinensis* 'Positano' and *M. sinensis* 'Sirene' x *M. sinensis* 'Positano,' were found to be 50% compatible, and a single cross, *M. sinensis* 'Sirene' x *M. sinensis* 'Giraffe,' was completely incompatible.

4.5 Discussion

Screening biparental *Miscanthus* populations for segregation at the SI loci using the phenotyping protocol employed herein is useful for quickly determining the number of loci and the number of SI genotypes segregating as a single model was converged upon often within the first day or two of testing. Once the number of SI genotypes and the SI relationships between these genotypes is known, a population can be screened fairly rapidly for purposes of mapping the locus or loci involved if the number of genotypes segregating is manageable, i.e. less than six. When conditions are ideal; not too hot, dry, or overcast; and an experienced individual is

imaging pollen tube growth, then shortcuts could be employed to more rapidly phenotype a population. In the case of the STxKA population, where there were only two SI genotypes in the population, a single pollen donor could be used to screen the entire population as opposed to using pollen from both groups. Likewise, in the case of the GFxUN population, two pollen donors could have been used to delineate individuals into all four SI groups as opposed to using pollen from all four groups.

An important note to make regarding the screening of biparental *Miscanthus* populations is the necessity to ensure that individuals are truly progeny of the intended population. A number of individuals had to be thrown out of the GFxUN population that showed non-parental alleles (Swaminathan *et al.*, 2012). Additionally, a number of individuals resulting from self-pollination had to be discarded in the STxKA population (Liu, 2015). In neither case were these individuals identified prior to genotyping. Including these individuals in the initial determination of the number of loci and the number of SI genotypes segregating could easily lead to incorrect conclusions regarding the underlying segregation of SI in that population. For these reasons, it might be wise to use genetic markers to screen populations prior to phenotyping for SI.

Despite screening the GFxUN population with genetic markers prior to the SI phenotyping, UG031 could not be placed into any of the four SI groups. For this reason, it was hypothesized that this individual had lost its ability to inhibit self-pollen either via a mutation in a gene essential to the SI response or a recombination event occurred between the male and female determinants, thereby breaking self-recognition. Further testing in the following year (2014) revealed that UG031 maintained its ability to inhibit self-pollen; therefore, neither hypothesis was supported. The most likely explanation is that UG031 does fit into one of the

four groups, but the stigmas from the plant used to test were in poor condition, resulting in uncharacteristic pollen tube growth.

An important point to remember when using QTL software on a qualitative trait is that the trait residuals will often break the assumption of normality. This causes the maximum likelihood algorithm to continue iterating to high LOD scores and can be discarded as artifacts (Johan W. Van Ooijen, personal communication). Due to this artificial inflation of LOD scores, little emphasis has been placed on these values beyond declaring the QTL significant. Along the same lines, the accuracy of the 2-LOD confidence interval should also be carefully interpreted. In saying that, there is no doubt that the locus responsible for the SI response in both the GFxUN and STxKA populations resides on the second half of LG 15. Further supporting this is the fact that strong segregation distortion was observed for markers on LG 15 of the STxKA population (Liu, 2015). This segregation distortion is expected due to the fact that half of the *M. sinensis* 'Kaskade' pollen in the initial cross would be expected to be incompatible with the pistils of *M. sinensis* 'Strictus;' therefore, any markers linked to the SI locus on LG 15 would be distorted, and the degree of distortion would be proportional to the extent of genetic linkage. This distortion was not present in the GFxUN population because 100 percent of the pollen in the initial cross was compatible.

It has been well documented throughout the years that increasing the population size increases the power to detect QTL, reduces the confidence interval for the QTL, and decreases the standard error of the QTL effect size (Darvasi *et al*., 1993; Beavis, 1994; Beavis, 1998; Vales *et al*., 2005). Even with as few as 30 individuals, a QTL was able to be detected for SI. In this case, it is not the ability to detect a QTL that is the limiting factor; it is the confidence interval associated with this position that is limited by the number of individuals. The ability to locate the

chromosome in which the locus controlling SI is on with so few individuals should be useful in delineating any other distinct loci associated with SI. This is important because it is impossible to distinguish between loci based on segregation ratios alone. As the power to detect and place QTLs is much larger in the GFxUN population due the larger number of individuals phenotyped, more weight should be put on the location of the SI QTL in this population.

While Davasi and colleagues (1993) found that markers do not increase the power to detect QTLs when at a higher density than one every ten cM, they do decrease the confidence interval up to the resolving power. In other words, increasing the number of markers is only helpful as population size or recombination frequency increases. Since no markers in this analysis were perfectly linked to the SI QTL, the marker density on LG 15 is below the resolving power, and additional markers on LG 15 would be expected to decrease the confidence interval associated with this locus.

Surprisingly, the location of the SI locus identified herein is not homologous to either the S- or Z-loci as would be expected if the SI system was shared across the entire grass family. As previously mentioned, all of the grass species in which SI has been investigated to date have been consistent with the two gene, S and Z SI system (Langridge & Baumann, 2008). The T-locus, on the other hand, is homologous to the region of the genome associated with SI in *Miscanthus*. Previously, this locus has only been shown to have two alleles—a functional and non-functional allele—with the non-functional allele resulting in self-compatible plants (Thorogood *et al*. 2005). Therefore, it was hypothesized that the T-locus was a remnant of an SI system containing more than two loci and this locus had been fixed for a single allele. To the author's knowledge, this locus has always been associated with a breakdown of SI. Its action in *Miscanthus* is, therefore, unique in two ways: allelic diversity exists at this locus and it is not

associated with a breakdown of SI. Due to the shared synteny between the T-locus and the locus identified in *Miscanthus*, this locus will henceforth be referred to as the T-locus

If the locus identified herein and the T-locus are one and the same, then this has some interesting evolutionary consequences. This would suggest that the common ancestor of the Pooideae and Panicoideae subfamilies, if not all grasses, possessed a shared SI mechanism that involved at least three loci. After the grass subfamilies diverged, the fate of the three or more SI loci appear to have followed distinct paths. In the case of the T-locus, it appears that it may have indeed been fixed in the Pooideae subfamily, yet diversity remains at this locus within the Panicoideae subfamily. Since not all SI species within the Pooideae subfamily have been explored, it cannot be unequivocally said that no diversity exists at the T-locus within the entire Pooideae subfamily but only those families studied to date.

If this hypothesis is correct, then additional loci controlling SI in *Miscanthus* will likely be found and could still possibly be homologous to the S- and Z-loci. While only a single locus was segregating in the first two mapping populations investigated, the identification of crosses in which the percent of compatible pollen is 75% suggests that multiple loci are controlling the SI response in *Miscanthus*. Until the second locus is mapped, it cannot be known for sure if it is an artifact of the recent whole genome duplication event that occurred in the *Miscanthus* lineage (Kim *et al*., 2012; Ma *et al*., 2012; Swaminathan *et al*., 2012). The fact that over 85% of crosses investigated in the diallel experiment were 100% compatible suggests that there is high allelic diversity for SI as expected. If this diversity did not exist, then SI would put *Miscanthus* at an evolutionary disadvantage as the effective population size would drastically decrease.

Given that only a single locus was segregating in the first two mapping populations investigated and the locus turned out to be the same in both may indicate that more diversity

exists at the T-locus than the other(s). Since both of these populations were derived from cultivars that trace their origins back to Japan, the lack of diversity at one SI locus may be a result of the initial bottleneck expected to have occurred when *Miscanthus* first migrated to Japan (Clark *et al*., 2014). Until the gene is cloned, it will be difficult to assess the diversity at the loci involved in SI and investigate how this diversity is associated with geography. From the information provided by the diallel crosses, it is unlikely that the second locus is completely fixed in *Miscanthus* native to Japan.

Another aspect with evolutionary consequences is the loss of SI. As stated, the two closest relatives of *Miscanthus* that have been fully sequenced are self-compatible species. SI could be lost in a number of ways, e.g. an insertion or deletion could cause a frame shift mutation, a point mutation could introduce a premature stop codon, recombination could occur between the two genes underlying a single locus, a large deletion could remove the entire region, etc. The fact that six consecutive markers so close to the SI locus cannot be placed on the sorghum genome suggests that a deletion may have occurred in this region, which would account for the self-compatible nature of sorghum. This is further supported by the gap present when comparing the syntenic region of rice. While rice is also a self-compatible species, the T-locus could be fixed with the non-functional allele, thereby negating SI without a large deletion. Since only one of the two paralogous regions in maize has evidence for a deletion in this region, the loss of SI in maize most likely occurred after the maize whole genome duplication event, making the deletion events in maize and sorghum independent. Further, this suggests that maize would likely have been SI until sometime after the whole genome duplication event approximately 5-12 Mya (Swigonova *et al*., 2004b).

An additional region of six consecutive markers on *Miscanthus* LG three could not be placed on the sorghum genome. The region in question is syntenic to sorghum chromosome two between 67.0 and 68.7 Mb. Considering the two locus SI system suggested to be operating in *Miscanthus* by the diallel crosses, it would be logical to suggest that the second locus may map to this position. Although further testing will be required to completely rule this hypothesis out, there is no evidence of a deletion in this region when comparing the rice and sorghum genomes. In addition, this region is not homologous to either the S- or Z-loci. The unplaced markers on *Miscanthus* LG three span an area of 2.8 cM while the unplaced markers on *Miscanthus* LG 15 span an area of 8.6 cM, a region over three times larger. If the local recombination rates are similar in both regions, then it may suggest that the putative deletion on LG three was significantly smaller. The smaller size may interfere with the detection of a deletion when comparing sequences at the whole genome level. Of course, it is imperative to point out that comparisons between physical and genetic distances must be interpreted cautiously as the genetic distance is influenced by local recombination rates, which vary across the genome (Begun & Aquadro, 1992; Tanksley *et al.*, 1992). Until the second locus controlling self-recognition is mapped in *Miscanthus*, a conclusion cannot be reached regarding whether this region is associated with SI.

Limited data is available on the underlying physiological action of the T-locus in grasses in which it has been previously identified in. Thorogood *et al.* (2005) investigated the percentage of compatible self-pollen in a population of *L. perenne*. The population of *L. perenne* was derived from selfing an individual developed from a cross between two inbred lines selected separately for SC. Segregation distortion was noted at the T-locus but not the second locus associated with SC in the population; this second locus was determined to be the S-locus. An

interaction between the T- and S-loci was suggested due to unexpected marker ratios in the population. The author's even suggested an interaction between the T- and Z-loci, but their evidence for this interaction is weak. Besides this potential interaction between the T-locus and the other SI loci, little is known about the action of the T-locus. Considering the different nature of the role that the T-locus plays in other species compared to *Miscanthus*, it is difficult to say how relevant the information from the Pooideae subfamily is.

The focus of the work herein was on the genetics of the SI system in *Miscanthus* as opposed to the underlying physiological action of the SI system. For this reason, it is difficult to speculate how the SI responses differ between *Miscanthus* and other grass SI systems. Many of the incompatible pollen grains fail to penetrate the stigma surface. Pollen tube growth appears erratic when incompatible. When incompatible pollen tubes manage to penetrate the stigma surface, their growth is generally inhibited rather quickly although not in all cases. In agreement with the observation that anthesis is heavily influenced by environmental stresses, pistils that are water or heat stressed have a decreased ability to inhibit incompatible pollen although this generally does not persist to the point of the pollen reaching the ovule.

This initial evaluation of the SI mechanism operating in *Miscanthus* generates many questions and potential avenues for research. One of the more important questions that needs to be addressed is where the second locus is located and if this region shares synteny with any of the other known locations involved in SI or SC in other grasses. While it would theoretically be possible to map the second locus in a population segregating for more than one SI locus, it would be complicated and time consuming placing each individual into an SI group. Ideally a population could be created that segregated exclusively for the unmapped locus. Unfortunately, it is difficult to create such a population with the information we currently have. Populations could

be developed from crossing the individuals that we know are 75% compatible. In this case, we would be confident that the second locus is segregating, but we would still have to deal with the segregation of 12 potential SI genotypes in the population.

A better solution would be to take advantage of the progeny developed from self-pollen in the STxKA cross. It is known that these individuals are homozygous at the second locus. Further, we would expect half of the selfed progeny to be homozygous at the T-locus. The double homozygous lines can easily be identified by investigating their pollen tube growth on *M. sinensis* 'Strictus' pistils. The double homozygous lines would be expected to be 50% compatible (Figure 4.15). Any population developed from one of the double homozygous lines would have a maximum of four SI genotypes segregating. It would be wise to cross very diverse lines when creating populations from the double homozygous plants to ensure that the other parent is not also homozygous with the same alleles for the second locus. Once the populations are developed, a similar strategy as the one used herein could be used to phenotype individuals and map the locus/loci responsible for SI.

In addition to mapping the second locus, it would be beneficial to fine map the T-locus. In order to accomplish this goal, random mating could be carried out within the GFxUN population for a number of generations. After random mating, a biparental population could be developed for fine mapping purposes. In addition to the original markers identified, more markers would have to be created for linkage group 15 to make this strategy effective. In concert, an RNA-seq experiment could be performed to look at differential gene expression between mature and immature pistils and stamens as well as a time course experiment to look at differences in gene expression between compatible and incompatible reactions. In addition to identifying potential gene candidates underlying the T-locus, this data could be mined for additional markers.

Screening BAC libraries for sequences linked to the *Miscanthus* T-locus and sequencing the identified regions would help both with identifying candidate genes and creating new markers to fine map the locus. More importantly, this sequence information would help clarify the initial comparative genomics work and hopefully provide insight into how SI and the loss of SI shaped the evolution of the Poaceae family.

Finally, experiments could be carried out to better understand the physiology of the compatible and incompatible reactions. Lavendustin A, a protein kinase inhibitor, and verapamil, a calcium channel blocker, can be incorporated into the boric acid media at 50 µM and 1 mM, respectively, to test their effect on pollen tube growth. Failure of the pistil to inhibit self-pollen would suggest that these processes play a role in downstream signaling. The results of these as well as other physiology experiments would complement the gene cloning effort as they may hint at functional domains that could be used to further filter gene candidates.

4.6 Conclusion

The SI relationships between progeny of two bi-parental populations were successfully used to map SI in *Miscanthus* for the first time. In fact, to the author's knowledge, this is the first time any SI locus has been mapped in the Panicoideae subfamily of Poaceae. A single locus was segregating in the first population inspected; this locus mapped to LG 15, which is syntenic to LG eight of sorghum. This observation was independently confirmed in a second population via segregation distortion and genetic mapping of a subsample of the population. Utilizing the conserved synteny between the grasses, it was determined that this region was homologous to the previously identified T-locus. Again, to the author's knowledge, this is the first time that allelic diversity beyond functional and non-functional alleles has been identified at this locus. Further, the T-locus has previously only been associated with the breakdown of SI.

Comparative genomics analysis suggests that a deletion in the sorghum genome compared to rice may be responsible for the loss of SI in sorghum. This deletion would have occurred after the divergence of *Miscanthus* and sorghum roughly 5.4 million years ago (Kim *et al*., 2014). If this is the case, then it would be expected that an independent deletion in maize or another interruption in the genes responsible for SI would have caused maize to have lost its SI response. The fact that one paralogous region shows evidence for a deletion compared to rice and the second does not suggests that these events were indeed independent and that they occurred after the whole genome duplication event in maize.

An exploration of a partial diallel has also provided evidence that at least one additional unlinked locus is responsible for controlling SI in *Miscanthus*. Little is currently known about this locus. Due to the differences and similarities between the evidence found in Pooideae and Panicoideae subfamilies of the grass family, it is entirely possible that the common ancestor of these two subfamilies had the four locus SI system that was previously suggested and that after their divergence individual loci suffered different evolutionary fates.

4.7 Figures and tables

Figure 4.1 A working model of Brassicaceae-type self-incompatibility (SI) as published in Takayama and Isogai (2005). The stigma transmembrane protein SRK is the female determinant while SP11, represented by the pink circles and squares, is the male determinant. Recognition of the self-SP11 allele by SRK triggers a phosphorylation signaling cascade, leading to the rejection of self-pollen (left side). No interaction occurs between SRK and SP11 in the case of cross-pollen (right side) and pollen growth continues as normal.
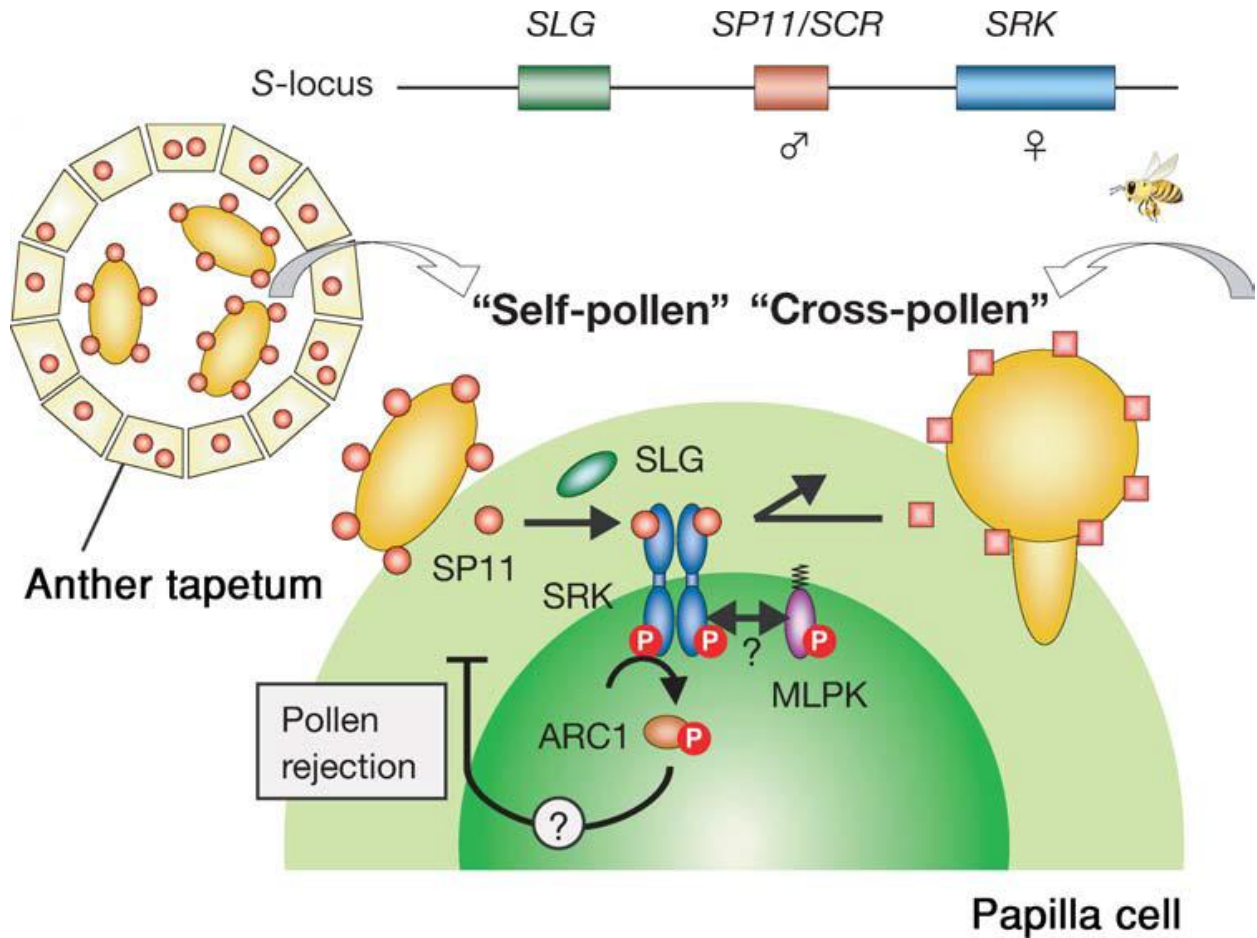
Figure 4.2 A working model of Solanaceae-type self-incompatibility (SI) as published in Takayama and Isogai (2005). S-RNase, the female determinant, penetrates the growing pollen tubes of both self- and cross-pollen in the style, but it is only allowed to degrade pollen RNA in self-pollen. This differentiation is regulated by the male determinant, SLF/SFB, which targets nonself S-RNases for degradation via ubiquitin signaling.
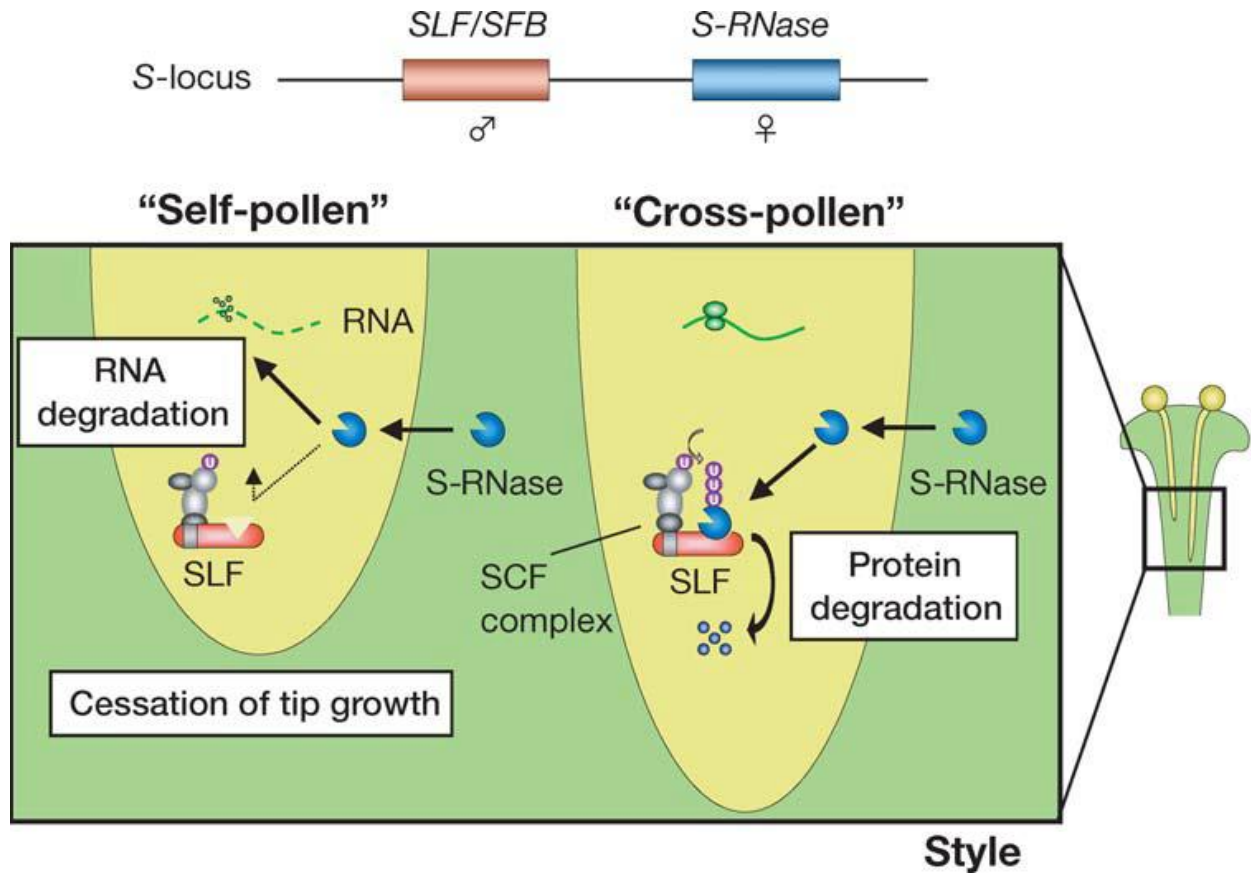
Figure 4.3 A working model of Papaveraceae-type self-incompatibility (SI) as published in Takayama and Isogai (2005). The male determinant is an integral membrane protein expressed exclusively in growing pollen tubes. At the time that this figure was created the male determinant had not been identified, so it is referred to as the S-receptor in the below figure although it is now called PrpS. The female determinant is expressed in and secreted by stigma cells. Again, in this figure it is referred to as the S-protein, but it is now commonly called PrsS. PrpS opens calcium ion channels upon recognition of the self-PrsS protein. The resulting calcium influx degrades the actin cytoskeleton, leading to the cessation of pollen tube growth. This step is reversible until programmed cell death is triggered.
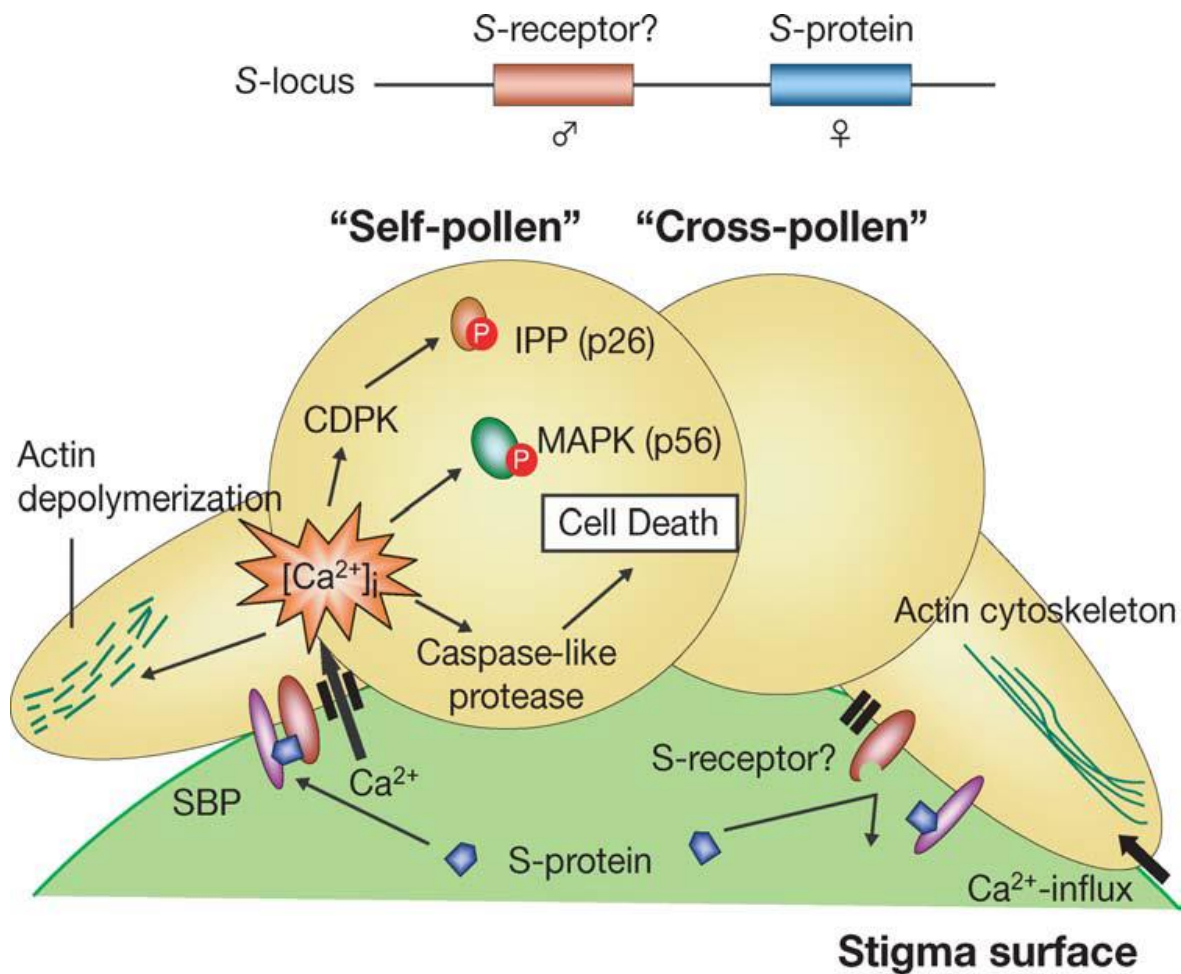
Figure 4.4 A phylogeny of the grasses created by Gaut (2002), including the estimated number of species, basic chromosome numbers, genome size range, and economic species of interest present in each subfamily.

Subfamily	No. species	Basic chrom. number	Genome size (pg)	Economic species
Anomochlooideae	3	9,11	–	–
Bambusoideae	970	12	–	–
Ehrhartoideae	105	10,11,12,15	0.70 – 2.45	Rice
Pooideae	3300	2,4,5,6,7,8, 9,10,13,19	2.25 – 17.9	Oats, turfgrasses, barley, wheat, rye
Chloridoideae	1350	7,8,9,10	0.50 – 3.35	Finger millet
Panicoideae	3240	5,6,7,8, 9,10,12	1.05 – 9.15	Maize, sorghum, pearl millet, sugar cane, foxtail millet

Figure 4.5 The *Miscanthus* inflorescence with important anatomical features labeled. Together the stigma, style and ovary make the pistil while the anther and filament make up the stamen. [Modified from Sun *et al*. (2010)]
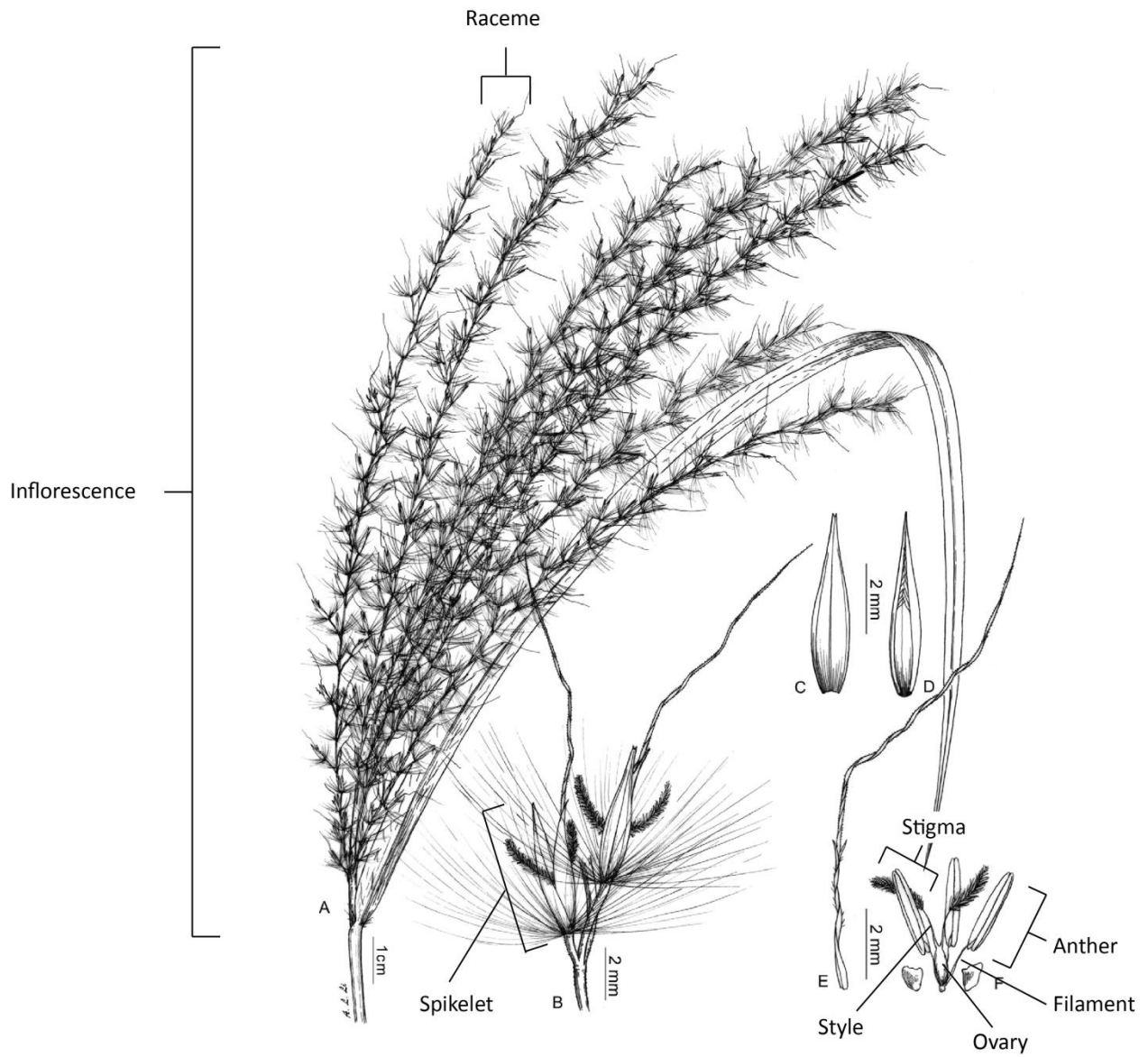
Figure 4.6 Photographs illustrating the process of imaging pollen tube growth: (A) An individual plant flowering (B) A close up of two racemes (C) Placing dissected pistils onto the boric acid plate (D) Pollinating the boric acid plate containing pistils (E) Evidence of a nice pollen cloud (F) Stigmas softening in 10 M NaOH (G) An individual microscope slide containing a single pistil soaking in 0.1% aniline blue dye (H) A microscope slide box containing 96 pistils ready for imaging
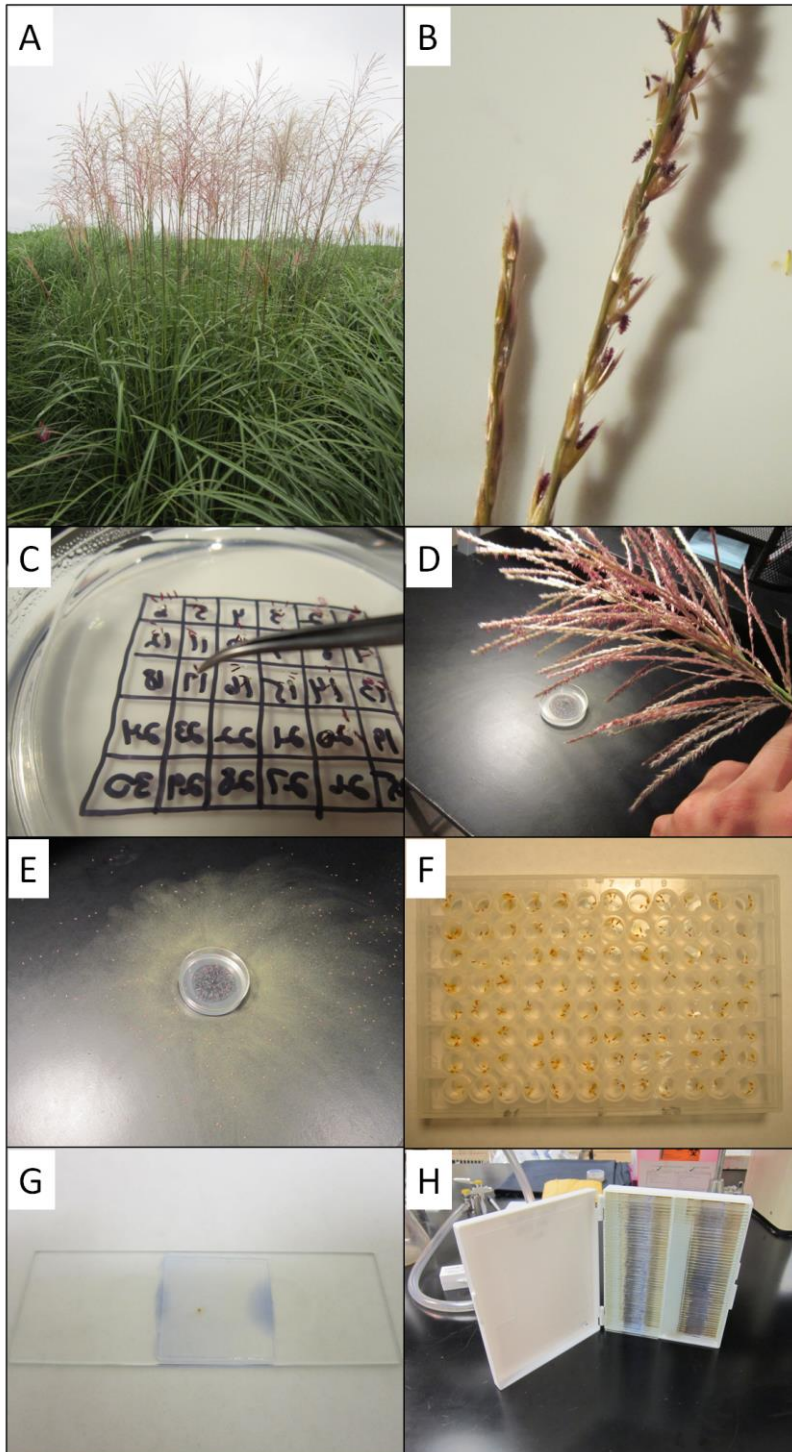
Figure 4.7 A series of Punnett squares depicting how the self-incompatibility (SI) genotype of the parents of a biparental population affect the number of SI genotypes and thus groups segregating in the population when (A) SI is controlled by a single locus and the population is developed from a single direction cross, (B) SI is controlled by a single locus and the population is developed from a reciprocal cross, (C) SI is controlled by two loci and the population is developed from a single direction cross, and (D) SI is controlled by two loci and the population is developed from a reciprocal cross. The number of SI genotypes resulting from each cross is provided below each Punnett square. The single horizontal dashes signify incompatibility. Identical genotypes are similarly colored to more easily depict the number of SI genotypes.
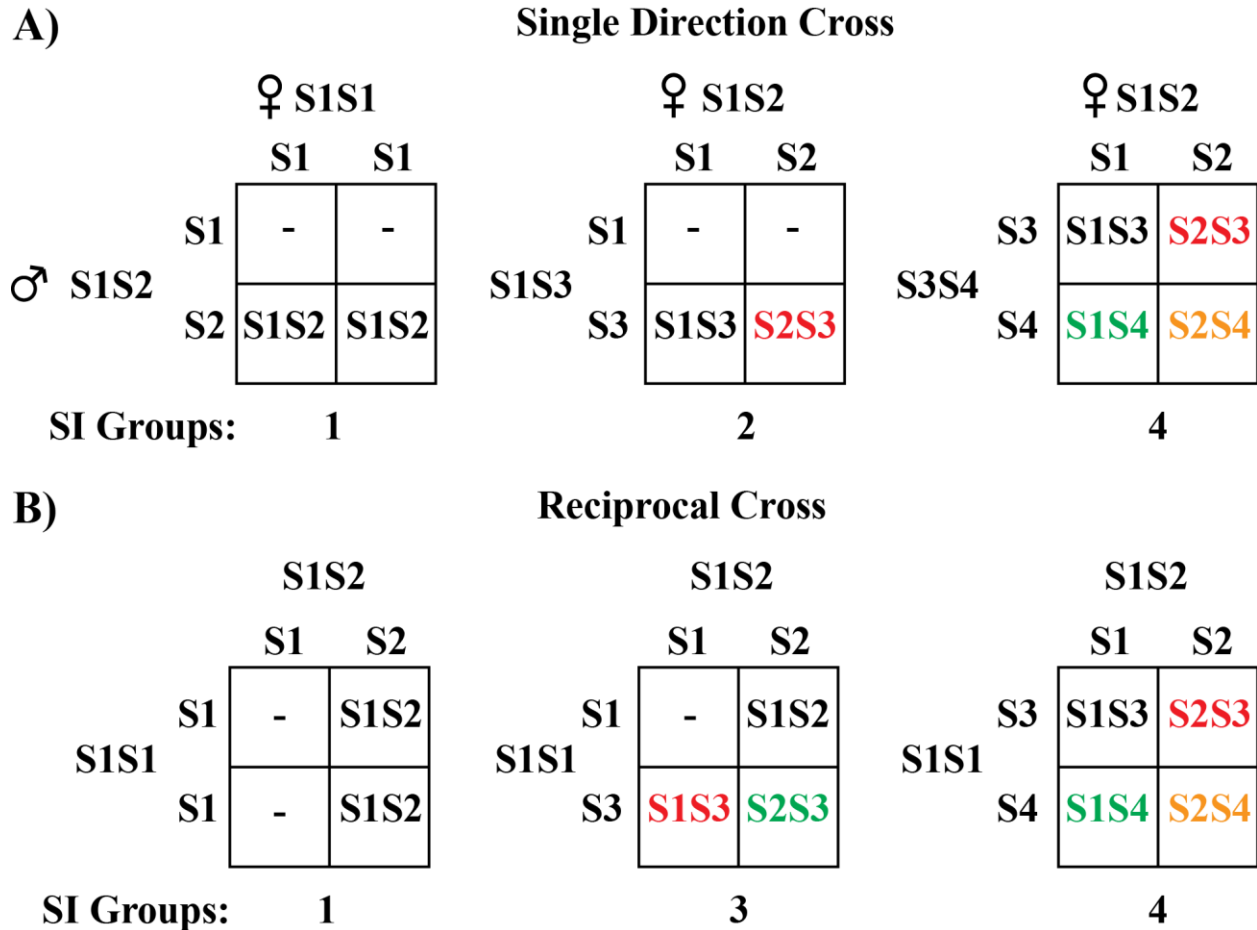
Figure 4.7 (Continued)

C) Single Direction Cross

♀ S1S1 / Z1Z1

	S1Z1	S1Z1	S1Z1	S1Z1
S2Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1
S2Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1
S2Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1
S2Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1

♂ S2S2 Z1Z1

SI Groups: **1**

♀ S1S2 / Z1Z1

	S1Z1	S1Z1	S2Z1	S2Z1
S1Z1	-	-	-	-
S1Z1	-	-	-	-
S3Z1	S1S3 Z1Z1	S1S3 Z1Z1	S2S3 Z1Z1	S2S3 Z1Z1
S3Z1	S1S3 Z1Z1	S1S3 Z1Z1	S2S3 Z1Z1	S2S3 Z1Z1

S1S3 Z1Z1

SI Groups: **2**

♀ S1S2 / Z1Z1

	S1Z1	S1Z1	S2Z1	S2Z1
S3Z1	S1S3 Z1Z1	S1S3 Z1Z1	S2S3 Z1Z1	S2S3 Z1Z1
S3Z1	S1S3 Z1Z1	S1S3 Z1Z1	S2S3 Z1Z1	S2S3 Z1Z1
S4Z1	S1S4 Z1Z1	S1S4 Z1Z1	S2S4 Z1Z1	S2S4 Z1Z1
S4Z1	S1S4 Z1Z1	S1S4 Z1Z1	S2S4 Z1Z1	S2S4 Z1Z1

S3S4 Z1Z1

SI Groups: **4**

S1S2 Z1Z2

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z1	-	-	-	-
S1Z2	-	-	-	-
S3Z1	S1S3 Z1Z1	S1S3 Z1Z2	S2S3 Z1Z4	S2S3 Z1Z2
S3Z2	S1S3 Z1Z2	S1S3 Z2Z2	S2S3 Z1Z2	S2S3 Z2Z2

♂ S1S3 Z1Z2

SI Groups: **6**

S1S2 Z1Z2

	S1Z1	S1Z2	S2Z1	S2Z2
S3Z1	S1S3 Z1Z3	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2
S3Z1	S1S3 Z1Z3	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2
S4Z1	S1S4 Z1Z1	S1S4 Z1Z2	S2S4 Z1Z1	S2S4 Z1Z2
S4Z1	S1S4 Z1Z1	S1S4 Z1Z2	S2S4 Z1Z1	S2S4 Z1Z2

S3S4 Z1Z1

SI Groups: **8**

S1S2 Z1Z2

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z1	-	-	-	-
S1Z3	S1S1 Z1Z3	S1S1 Z2Z3	S1S2 Z1Z3	S1S2 Z2Z3
S3Z1	S1S3 Z1Z1	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2
S3Z3	S1S3 Z1Z3	S1S3 Z2Z3	S2S3 Z1Z3	S2S3 Z2Z3

S1S3 Z1Z3

SI Groups: **12**

S1S2 Z1Z2

	S1Z1	S1Z2	S2Z1	S2Z2
S3Z3	S1S3 Z1Z3	S1S3 Z2Z3	S2S3 Z1Z3	S2S3 Z2Z3
S3Z4	S1S3 Z1Z4	S1S3 Z2Z4	S2S3 Z1Z4	S2S3 Z2Z4
S4Z3	S1S4 Z1Z3	S1S4 Z2Z3	S2S4 Z1Z3	S2S4 Z2Z3
S4Z4	S1S4 Z1Z4	S1S4 Z2Z4	S2S4 Z1Z4	S2S4 Z2Z4

♂ S3S4 Z3Z4

SI Groups: **16**

168

Figure 4.7 (Continued)

D)

Reciprocal Cross

		S1S1 Z2Z2				S1S3 Z2Z2				S1S3 Z1Z2

First grid (SI Group 1):

Column headers: S1Z2 S1Z2 S1Z2 S1Z2

Row label: **S1S3 Z2Z2**

	S1Z2	S1Z2	S1Z2	S1Z2
S1Z2	-	-	-	-
S1Z2	-	-	-	-
S3Z2	S1S3 Z2Z2	S1S3 Z2Z2	S1S3 Z2Z2	S1S3 Z2Z2
S3Z2	S1S3 Z2Z2	S1S3 Z2Z2	S1S3 Z2Z2	S1S3 Z2Z2

SI Groups: 1

Second grid (SI Group 3):

Row label: **S1S3 Z1Z1**

	S1Z2	S1Z2	S3Z2	S3Z2
S1Z1	S1S1 Z1Z2	S1S1 Z1Z2	S1S3 Z1Z2	S1S3 Z1Z2
S1Z1	S1S1 Z1Z2	S1S1 Z1Z2	S1S3 Z1Z2	S1S3 Z1Z2
S3Z1	S1S3 Z1Z2	S1S3 Z1Z2	S3S3 Z1Z2	S3S3 Z1Z2
S3Z1	S1S3 Z1Z2	S1S3 Z1Z2	S3S3 Z1Z2	S3S3 Z1Z2

SI Groups: 3

Third grid (SI Group 4):

Row label: **S1S3 Z1Z1**

	S1Z1	S1Z2	S3Z1	S3Z2
S1Z1	-	S1S1 Z1Z2	-	S1S2 Z1Z2
S1Z1	-	S1S1 Z1Z2	-	S1S2 Z1Z2
S3Z1	-	S1S3 Z1Z2	-	S2S3 Z1Z2
S3Z1	-	S1S3 Z1Z2	-	S2S3 Z1Z2

SI Groups: 4

Fourth grid (SI Group 6):

Headers: S1S2 Z1Z2 ; Row label: **S1S2 Z3Z3**

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z3	S1S1 Z1Z4	S1S1 Z2Z4	S1S2 Z1Z3	S1S2 Z2Z3
S1Z3	S1S1 Z1Z4	S1S1 Z2Z4	S1S2 Z1Z3	S1S2 Z2Z3
S2Z3	S1S2 Z1Z3	S1S2 Z2Z3	S2S2 Z1Z3	S2S2 Z2Z3
S2Z3	S1S2 Z1Z3	S1S2 Z2Z3	S2S2 Z1Z3	S2S2 Z2Z3

SI Groups: 6

Fifth grid (SI Group 7):

Headers: S1S2 Z1Z2 ; Row label: **S1S3 Z1Z1**

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z1	-	S1S1 Z1Z2	S1S2 Z1Z1	S1S2 Z1Z2
S1Z1	-	S1S1 Z1Z2	S1S2 Z1Z1	S1S2 Z1Z2
S3Z1	S1S3 Z1Z1	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2
S3Z1	S1S3 Z1Z1	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2

SI Groups: 7

Sixth grid (SI Group 8):

Headers: S1S2 Z1Z2 ; Row label: **S1S1 Z3Z4**

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z3	S1S1 Z1Z3	S1S1 Z2Z3	S1S2 Z1Z3	S1S2 Z2Z3
S1Z4	S1S1 Z1Z4	S1S1 Z2Z4	S1S2 Z1Z4	S1S2 Z2Z4
S1Z3	S1S1 Z1Z3	S1S1 Z2Z3	S1S2 Z1Z3	S1S2 Z2Z3
S1Z4	S1S1 Z1Z4	S1S1 Z2Z4	S1S2 Z1Z4	S1S2 Z2Z4

SI Groups: 8

Seventh grid (SI Group 9):

Headers: S1S2 Z1Z2 ; Row label: **S1S3 Z1Z2**

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z1	-	-	S1S2 Z1Z1	S1S2 Z1Z2
S1Z2	-	-	S1S2 Z1Z2	S1S2 Z2Z2
S3Z1	S1S3 Z1Z1	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2
S3Z2	S1S3 Z1Z2	S1S3 Z2Z2	S2S3 Z1Z2	S2S3 Z2Z2

SI Groups: 9

Eighth grid (SI Group 15):

Headers: S1S2 Z1Z2 ; Row label: **S1S3 Z1Z3**

	S1Z1	S1Z2	S2Z1	S2Z2
S1Z1	-	S1S1 Z1Z2	S1S2 Z1Z1	S1S2 Z1Z2
S1Z3	S1S1 Z1Z3	S1S1 Z2Z3	S1S2 Z1Z3	S1S2 Z2Z3
S3Z1	S1S3 Z1Z1	S1S3 Z1Z2	S2S3 Z1Z1	S2S3 Z1Z2
S3Z3	S1S3 Z1Z3	S1S3 Z2Z3	S2S3 Z1Z3	S2S3 Z2Z3

SI Groups: 15

Ninth grid (SI Group 16):

Headers: S1S2 Z1Z2 ; Row label: **S3S4 Z3Z4**

	S1Z1	S1Z2	S2Z1	S2Z2
S3Z3	S1S3 Z1Z3	S1S3 Z2Z3	S2S3 Z1Z3	S2S3 Z2Z3
S3Z4	S1S3 Z1Z4	S1S3 Z2Z4	S2S3 Z1Z4	S2S3 Z2Z4
S4Z3	S1S4 Z1Z3	S1S4 Z2Z3	S2S4 Z1Z3	S2S4 Z2Z3
S4Z4	S1S4 Z1Z4	S1S4 Z2Z4	S2S4 Z1Z4	S2S4 Z2Z4

SI Groups: 16

Figure 4.8 Representative photographs of (A) 0% pollen compatibility (B) 50% pollen compatibility (C) 100% pollen compatibility, and (D) 0% pollen growth on overly mature pistil.
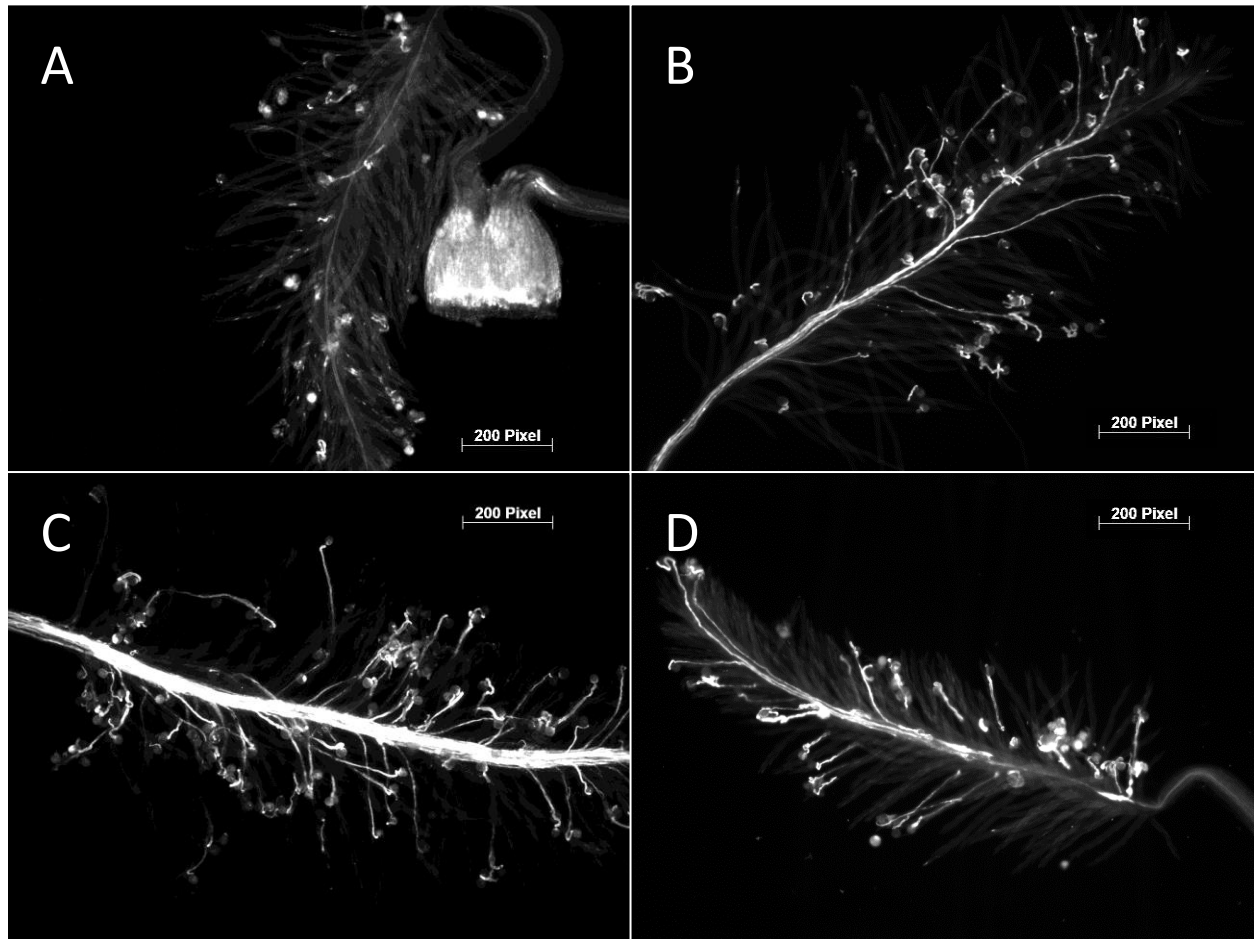
Figure 4.9 Punnett squares depicting the actual segregation found in the (A) *Miscanthus sinensis* 'Grosse Fountain' x *M. sinensis* 'Undine' population and (B) the *M. sinensis* 'Strictus' x *M. sinensis* 'Kaskade' mapping population as well as the self-incompatibility relationships between the genotypes of each population.

A
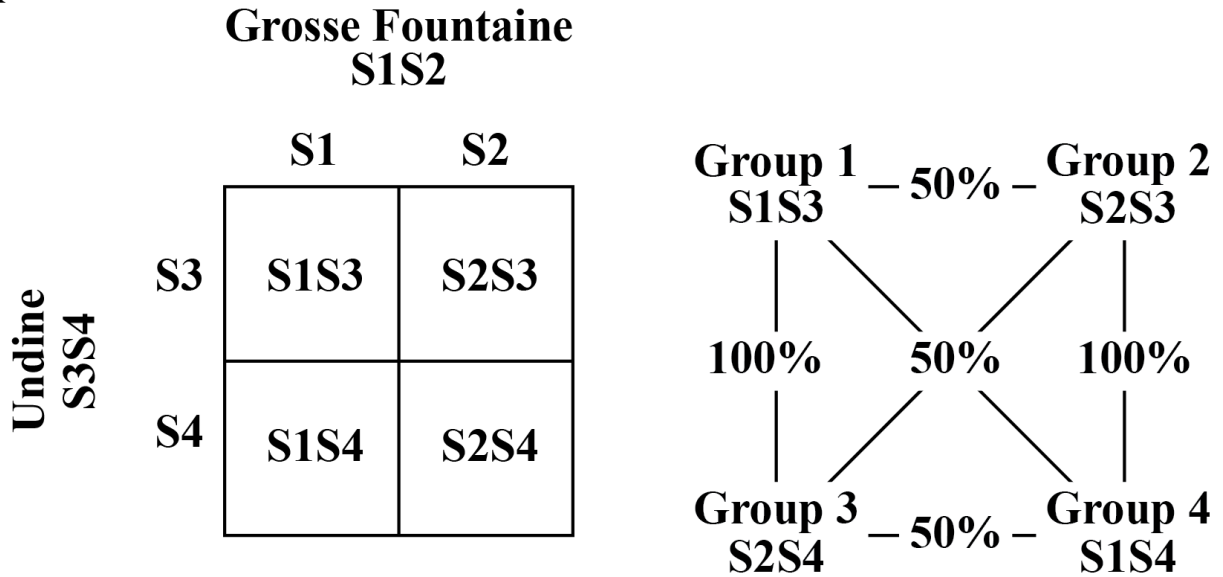
B
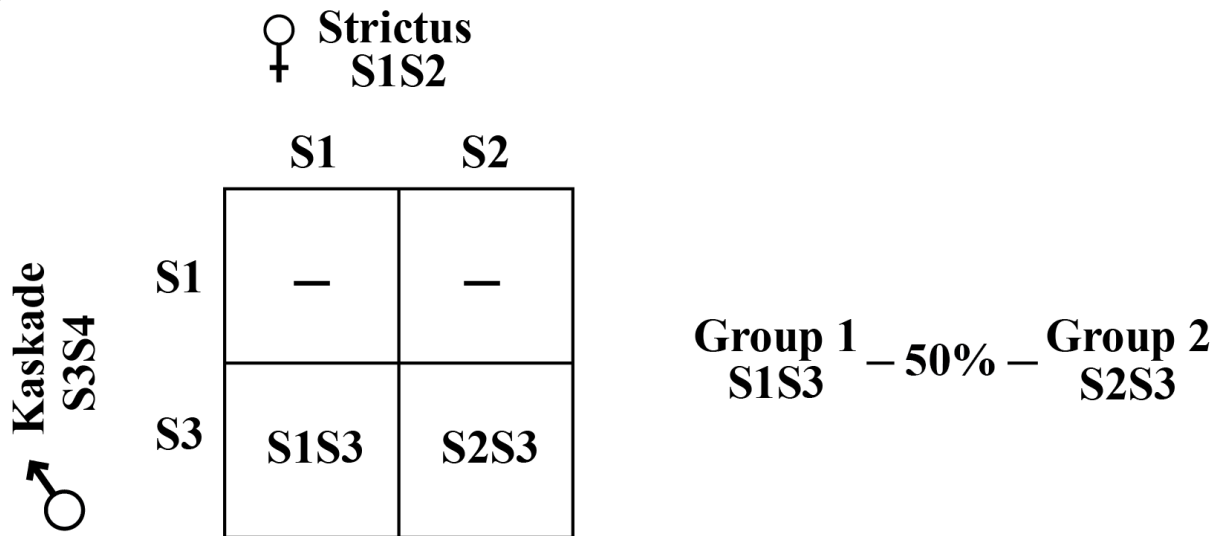
C

Figure 4.10 The self-incompatibility QTL LOD scores plotted on linkage group 15 in the (A) *Miscanthus sinensis* 'Grosse Fountain' x *M. sinensis* 'Undine' population, (B) the *M. sinensis* 'Strictus' x *M. sinensis* 'Kaskade' (STxKA) mapping population with a reduced set of markers, (C) the STxKA mapping population with markers specific to *M. sinensis* 'Strictus,' and (D) the STxKA mapping population with markers specific to *M. sinensis* 'Kaskade.' The number of individuals as well as the number of markers involved in each analysis is found in the top left corner of each plot. The horizontal line is the permuted significance threshold based on 1,000 permutations. The scale of the STxKA linkage group has been adjusted by markers common to both mapping populations.

Self-Incompatibility Locus on Linkage Group 15
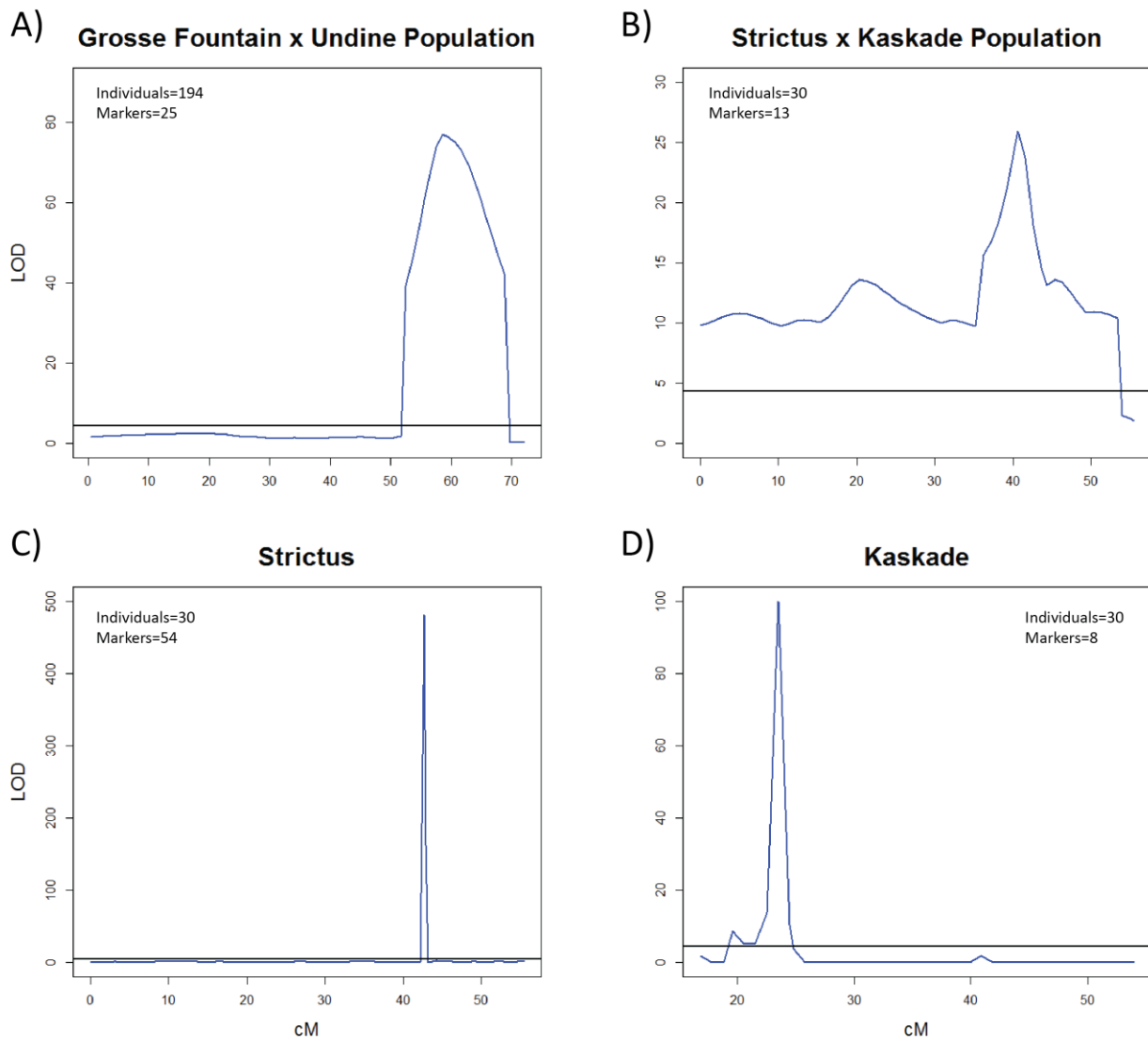
Figure 4.11 The number of consecutive markers that could not be uniquely placed onto the sorghum genome (observed) vs. the number of consecutive markers that would be expected if they were randomly distributed along the *Miscanthus sinensis* 'Grosse Fountaine' x *M. sinensis* 'Undine' genetic map (expected). The x-axis refers to the length of the run of consecutive markers while the y-axis refers to the number of times a run of that length occurred. The top portion of the figure shows the full scale while the bottom portion of the figure shows a close up.

Number of Consecutive Unplaced Markers
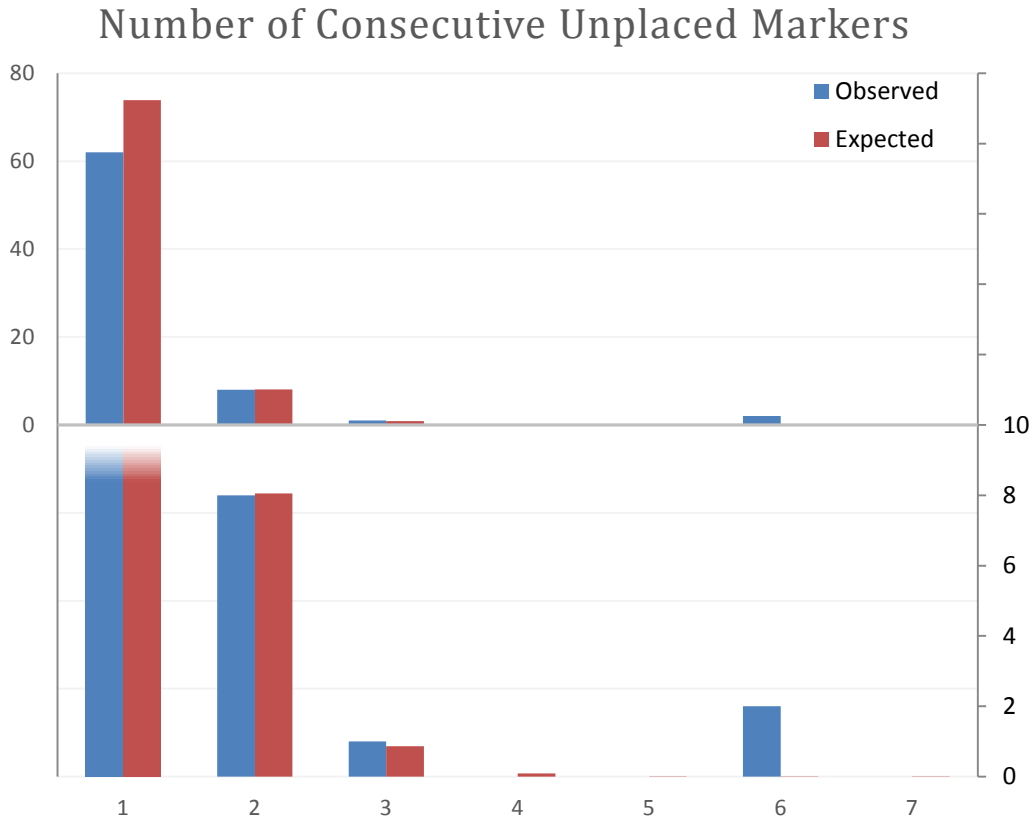
Figure 4.12 Plot showing the conserved synteny between rice and sorghum highlighting the region of the T-locus. The red arrow indicates a gap suggesting a possible deletion in sorghum relative to rice. The gap is 1.56 Mb in the x-direction and 0.61 Mb in the y-direction.
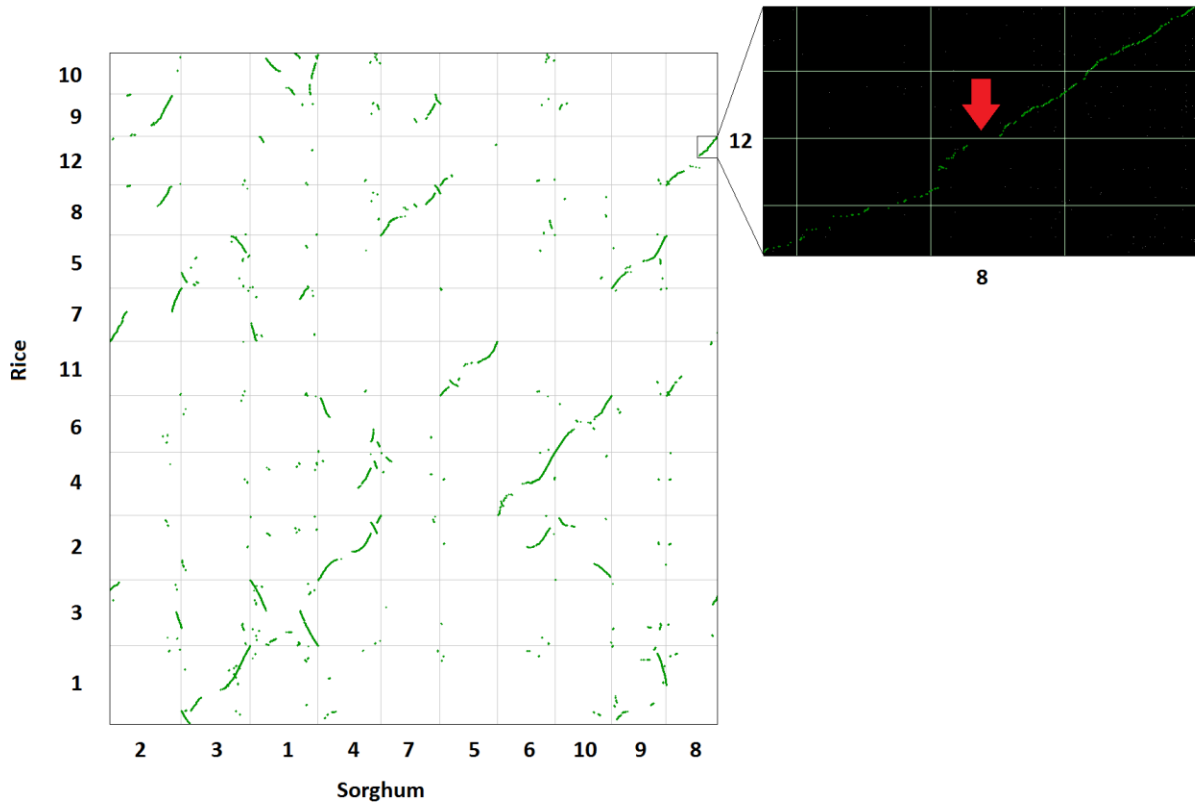
Figure 4.13 Plot showing the conserved synteny between maize and sorghum highlighting the region of the T-locus. The red arrow on chromosome ten of maize indicates a gap suggesting a possible deletion in sorghum relative to maize. The gap is 1.45 Mb in the x-direction and 0.67 Mb in the y-direction. The red arrow on chromosome three of maize indicates a gap suggesting the region shares no homology between the two species. The gap is 1.30 Mb in the x-direction and 1.24 Mb in the y-direction.
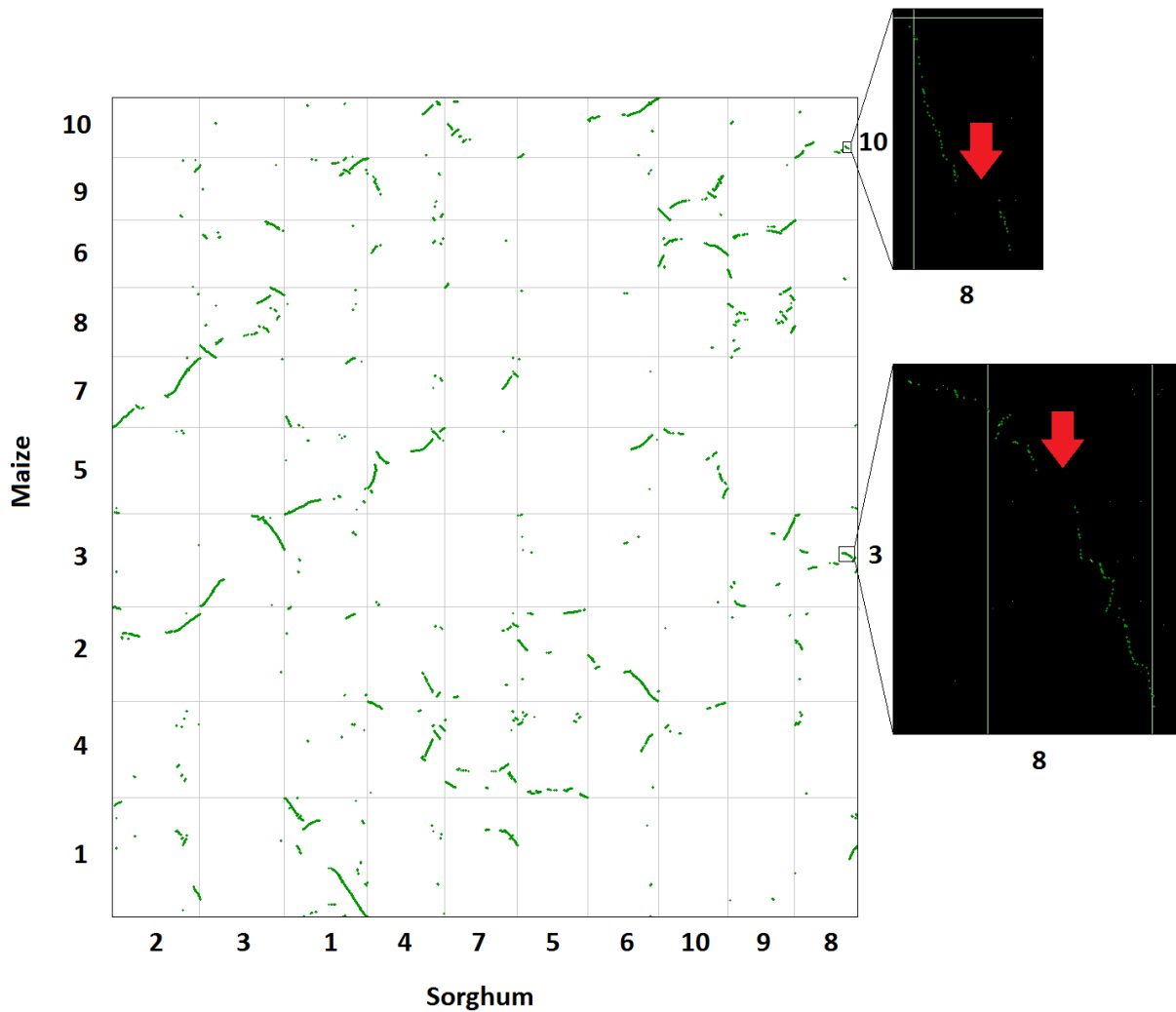
Figure 4.14 Plot showing the conserved synteny between rice and maize highlighting the region of the T-locus. The red arrow on chromosome ten of maize indicates a small gap. The gap is 0.82 Mb in the x-direction and 0.41 Mb in the y-direction. The red arrow on chromosome three of maize indicates a gap suggesting a possible deletion in maize compared to rice. The gap is 1.89 Mb in the x-direction and 0.55 Mb in the y-direction.
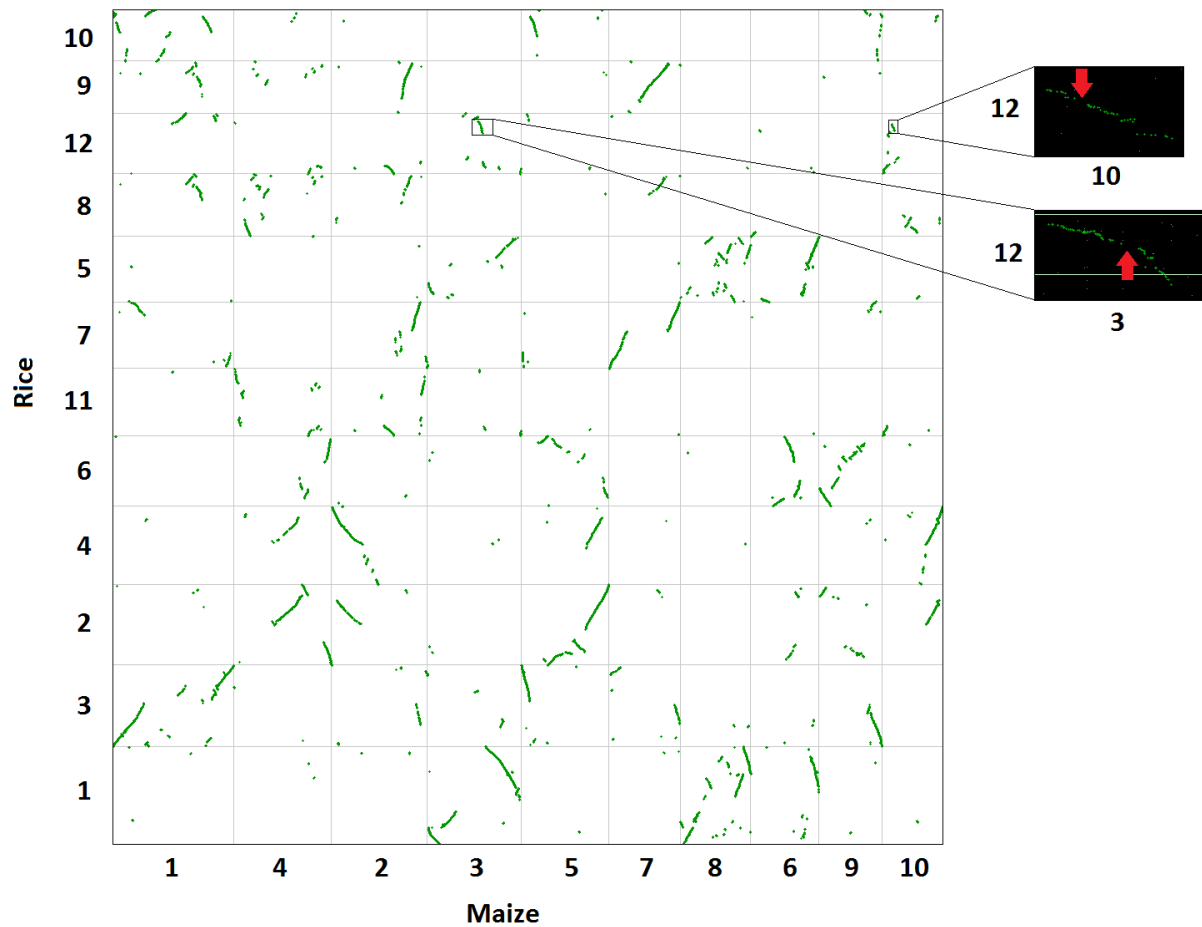
Figure 4.15 A series of Punnett squares depicting how to use *Miscanthus sinensis* 'Strictus' and the plants that resulted from the self-pollination of *M. sinensis* 'Strictus' to identify the plants that are homozygous for both self-incompatibility loci. The double homozygotes will be 50% compatible with *M. sinensis* 'Strictus'

A)

♀ S1S1
♂ Z1Z1

♂ S1S2 Z1Z1

	S1Z1	S1Z1	S1Z1	S1Z1
S1Z1	–	–	–	–
S1Z1	–	–	–	–
S2Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1
S2Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1

B)

♀ S2S2
♂ Z1Z1

♂ S1S2 Z1Z1

	S2Z1	S2Z1	S2Z1	S2Z1
S1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1
S1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1	S1S2 Z1Z1
S2Z1	–	–	–	–
S2Z1	–	–	–	–

C)

♀ S1S2
♂ Z1Z1

♂ S1S2 Z1Z1

	S1Z1	S1Z1	S1Z1	S1Z1
S1Z1	–	–	–	–
S1Z1	–	–	–	–
S2Z1	–	–	–	–
S2Z1	–	–	–	–

177

Table 4.1 The accessions involved in the diallel cross along with information regarding their origin. The accessions acquired from nurseries have since been shown to be of Japanese origin (Clark *et al*., 2014).

Accession	Species	Type	Nursery	Latitude	Longitude	Elevation	Country	State	City
10UI-009-059	*M. sacchariflorus* x *M. sinensis*	Cross*							
Andante	*M. sinensis*	Cultivar	Bluemel						
Giraffe	*M. sinensis*	Cultivar	Bluemel						
NC-2010-004-B(37)-d	*M. sinensis*	Wild plant		35.7576	-82.5412	665	USA	NC	
PMS-071	*M. sacchariflorus*	Wild plant		40.3906	117.0071	116	China	Beijing	Beijing
PMS-076	*M. sacchariflorus*	Wild plant		40.1642	116.0500	141	China	Beijing	Beijing
PMS-144	*M. sinensis*	Wild plant		34.2450	106.9331	1467	China	Shaanxi	Baoji
PMS-159	*M. sinensis*	Wild plant		34.1295	111.0440	757	China	Henan	Sanmenxia
PMS-164	*M. sinensis*	Wild plant		37.3400	114.2810	360	China	Hebei	Xingtai
Positano	*M. sinensis*	Cultivar	Earthly Pursuits						
Rotsilber	*M. sinensis*	Cultivar	Bluemel						
Sirene	*M. sinensis*	Cultivar	Bluemel						

* *M. sacchariflorus* 'Robustus' x *M. sinensis* 'Cosmopolitan Revert'

Table 4.2 The five closest markers to the peak of the SI QTL in the *Miscanthus sinensis* 'Grosse Fountaine' (GF) x *M. sinensis* 'Undine' (UN) mapping population along with their position on LG 15, segregation type, and phase. The segregation type indicates the genotype of the two parents at each locus. The genotype of GF is listed first, followed by the genotype UN. For example, GF is homozygous for allele 'n' at locus EBI 684 while UN is heterozygous; it has both the 'n' and 'p' allele. The phase indicator reveals what alleles are in coupling vs. in repulsion in each parent. The first digit of the phase indicator refers to the first allele (leftmost) of GF while the second digit refers to the first allele of UN. If the phase indicator is the same for two markers then they are in coupling in that parent. If the phase indicator is different, then the alleles are in repulsion for that parent. For example, in UN, allele 'n' at locus EBI 684 is in repulsion with allele 'n' at locus EBI 686, and in GF, allele 'l' at locus EBI 685 is in coupling with allele 'h' at locus EBI 687. The most common genotype for each group is also given as well as the number and frequency of recombinants. Recombinants are individuals that do not have the most common genotype of the group.

Markers	EBI 684	EBI 685	EBI 686	EBI 687	EBI 688
Position (cM)	50.0	51.6	52.5	59.7	69.7
Segregation	\<nnxnp\>	\<lmxll\>	\<nnxnp\>	\<hkxhk\>	\<nnxnp\>
Phase indicator	{-0}	{0-}	{-1}	{01}	{-1}
Group 1	np	lm	nn	hk	nn
Group 2	np	ll	nn	hh	nn
Group 3	nn	ll	np	hk	np
Group 4	nn	lm	np	kk	np
Number of Recombinant Individuals	18	10	15	13	12
Recombination Frequency (%)	9.3	5.2	7.7	6.7	6.2

Table 4.3 Chi-square tests for the number of self-incompatibility (SI) genotypes segregating in both mapping populations: *Miscanthus sinensis* 'Grosse Fountaine' x *M. sinensis* 'Undine' and reciprocal (GFxUN & UNxGF) and *M. sinensis* 'Strictus' x 'Kaskade' (STxKA). Only a single scenario for both population is statistically supported and shown in bold.

Population	SI Genotypes	Number of Loci	Individuals Tested	Chi-Square	P-value
GFxUN &	2	1	194	NA	NA
UNxGF	3	1	194	NA	NA
	4	**1**	**194**	**1.55**	**0.67**
	6	2	194	34.65	<0.001
	8	2	194	100.09	<0.001
	12	2	194	263.31	<0.001
	16	2	194	442.69	<0.001
STxKA	**2**	**1**	**30**	**0.13**	**0.72**
	3	1	30	NA	NA
	4	1	30	12.84	<0.001
	6	2	30	29.04	<0.001
	8	2	30	45.73	<0.001
	12	2	30	79.53	<0.001
	16	2	30	113.50	<0.001

Table 4.4 The percentage of compatible pollen for the portion of the diallel that was able to be completed. The crosses were placed into the 0%, 50%, 75%, or 100% categories based on the average of ten replications. Crosses known to be 50% and 100% compatible were used to test for significant deviations. The mean compatibility of the crosses known to be 50% compatible was 51.2% while the mean of the ten crosses known to be 100% was 85.3%. The two crosses, *M. sinensis* 'Giraffe' x *M. sinensis* 'Positano' and *M. sacchariflorus* 'PMS-071' x *M. sinensis* 'Positano,' that were deemed to fit into the 75% category resulted in statistically significant differences between 50% and 100% compatible reactions (P-values <0.001) and had means of 74.5 and 74.1, respectively. Crosses with NA were not able to be investigated.

	Pollen Donor			
Pistil	Giraffe	PMS-071	Positano	PMS-076
009_059	NA	100	100	100
PMS-164	100	100	100	100
Giraffe	0	100	75	100
PMS-071	NA	0	75	NA
Rotsilber	NA	100	100	100
Positano	100	100	0	100
NC-2010	NA	100	50	NA
Sirene	0	100	50	100
PMS-144	100	100	100	100
Andante	NA	100	100	100
PMS-159	100	100	100	100
PMS-076	NA	NA	NA	0

REFERENCES

Adati S, Shiotani I (1962) The cytotaxonomy of the genus *Miscanthus* and its phylogenic status. *Bulletin of Faculty of Agriculture, Mie University*, **25**, 1-24.

Al-Janabi SM, McClelland M, Petersen C, Sobral BWS (1994) Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. *Theoretical and Applied Genetics*, **88**, 933-944.

Alonso-Blanco C, El-Assal SE, Coupland G, Koornneef M (1998) Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics*, **149**,749-764.

Amalraj VA, Balasundaram N (2006) On the taxonomy of the members of 'Saccharum complex'. *Genet Resources and Crop Evolution*, **53**, 35-41.

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. *Proceedings of National Academy of Sciences*, 95, 13079-13084.

Anderson MA, Cornish EC, Mau S-L *et al*. (1986) Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. *Nature*, 321, 38-44.

Angelini LG, Ceccarini L, Nassi o Di Nasso N, Bonari E (2009) Comparison of *Arundo donax* L. and *Miscanthus × giganteus* in a long-term field experiment in Central Italy: analysis of productive characteristics and energy balance. *Biomass and Bioenergy*, **33**, 635-643.

Annamalai P, Rao ALN (2005) Replication-independent expression of genome components and capsid protein of *Brome mosaic virus* in planta: a functional role for viral replicase in RNA packaging. *Virology*, **338**, 96-111.

Annual Energy Outlook 2013 (2013) *U.S. Energy Information Administration*.

Arnoult S, Quillet MC, Brancourt-Hulmel M (2014) *Miscanthus* clones display large variation in floral biology and different environmental sensitivities useful for breeding. *BioEnergy Research*, **7**, 430-441.

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant molecular biology reporter*, **9**, 208-218.

Arundale RA, Dohleman FG, Heaton EA, Mcgrath JM, Voigt TB, Long SP (2014a) Yields of *Miscanthus× giganteus* and *Panicum virgatum* decline with stand age in the Midwestern USA. *Global Change Biology Bioenergy*, **6**, 1-13.

Arundale RA, Dohleman FG, Voigt TB, Long SP (2014b) Nitrogen fertilization does significantly increase yields of stands of *Miscanthus× giganteus* and *Panicum virgatum* in multiyear trials in Illinois. *BioEnergy Research*, **7**, 408-416.

Aslanidis C, DeJong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic acids research*, **18**, 6069-6074.

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003a) Identification of QTLs associated with yield and its components in *Miscanthus sinensis* Anderss. *Euphytica*, **132**, 353-361.

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003b) Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anderss. I. Total height, flag-leaf height and stem diameter. *Theoretical and Applied Genetics*, **107**, 123-129.

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003c) Identification of QTLs influencing combustion quality in *Miscanthus sinensis* Anderss. II. Chlorine and potassium content. *Theoretical and Applied Genetics*, **107**, 857-863.

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2003d) Influencing combustion quality in *Miscanthus sinensis* Anderss.: Identification of QTLs for calcium, phosphorus and sulphur content. *Plant Breeding*, **122**, 141-145.

Atwell WA, Hood LF, Lineback DR, Varriano-Marstonand E, Zobel HF (1988) The terminology and methodology associated with basic starch phenomena. *Cereal Foods World*, **33**, 306-311.

Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, Elledge SJ (1996) SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell*, **86**, 263-274.

Bassham JA, Calvin M (1957) The path of carbon in photosynthesis. Prentice Hall, Englewood Cliffs, NJ.

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. *Current Opinion in Plant Biology*, **2**, 109-113.

Baumann U (1995) Pollen mRNAs of *Phalaris coerulescens* and their possible role in self-incompatibility. PhD thesis. University of Adelaide.

Beale CV, Bint DA, Long SP (1996) Leaf photosynthesis in the C*4*-grass *Miscanthus × giganteus*, growing in the cool temperate climate of southern England. *Journal of Experimental Botany*, **47**, 267-273.

Beale CV, Long SP (1997) Seasonal dynamics of nutrient accumulation and partitioning in the perennial C_4-grasses *Miscanthus* × *giganteus* and *Spartina cynosuroides*. *Biomass and Bioenergy*, **12**, 419-428.

Beale CV, Morison JIL, Long SP (1999) Water use efficiency of C_4 perennial grasses in a temperate climate. *Agricultural and Forest Meteorology*, **96**, 103-115.

Beavis WD (1994) The power and deceit of QTL experiments: Lessons from comparitive QTL studies. In: 49th Annual Corn and Sorghum Resource Conference. American Seed Trade Association (ed. Wilkinson DB), pp. 250-266.

Beavis WD (1998) QTL analyses: Power, precision, and accuracy, pp. 145–162 in Molecular Dissection of Complex Traits, edited by A. H. Paterson. CRC Press, New York.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, **356**, 519-520.

Benavente LM, Scofield SR (2011) A new tool for functional genomics in maize. *New Phytologist*, **189**, 363-365.

Bennypaul HS, Mutti JS, Rustgi S, Kumar N, Okubara PA, Gill KS (2012) Virus-induced gene silencing (VIGS) of genes expressed in root, leaf, and meiotic tissues of wheat. *Functional & Integrated Genomics*, **12**, 143-156.

Bernardez TD, Lyford K, Hogsett DA, Lynd LR (1993) Adsorption of *Clostridium thermocellum* cellulase onto pretreated mixed hardwood, Avicel and Lignin. *Biotechnology and Bioengineering*, **43**, 899-907.

Bian XY, Friedrich A, Bai JR, Baumann U, Hayman DL, Barker SJ, Langridge P (2004) High-resolution mapping of the S and Z loci of *Phalaris coerulescens*. *Genome*, **47**, 918-930.

Björkman O, Pearcy RW, Harrison AT, Mooney H (1972) Photosynthetic adaptation to high temperatures: a field study in Death Valley, California. *Science*, **175**, 786-789.

Blevins T, Rajeswaran R, Shivaprasad PV *et al*. (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Research*, **34**, 6233-6246.

Bouché N, Lauressergues D, Gasciolli V, Vaucheret H (2006) An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *The EMBO Journal*, **25**, 3347-3356.

Bowers JE, Arias MA, Asher R *et al*. (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 13206-13211.

Boyko V, Hu Q, Seemanpillai M, Ashby J, Heinlein M, (2007) Validation of microtubule-associated *Tobacco mosaic virus* RNA movement and involvement of microtubule-aligned particle trafficking. *The Plant Journal*, **51**, 589-603.

BP Statistical review of world energy (2013) *British Petroleum* <http://www.bp.com/content/dam/bp/pdf/statistical-review/statistical_review_of_world_energy_2013.pdf>.

Bragg JN, Jackson AO (2004) The C-terminal region of the barley stripe mosaic virus γb protein participates in homologous interactions and is required for suppression of RNA silencing. *Molecular Plant Pathology*, **5**, 465-481.

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, **320**, 1185-1190.

Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. *Trends in Genetics*, **22**, 268-280.

Brosnan CA, Voinnet O (2011) Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications. *Current Opinion In Plant Biology*, **14**, 580-587.

Brown, RH (1978) A difference in N use efficiency in C_3 and C_4 plants and its implications in adaption and evolution. *Crop Science*, **18**, 93-98.

Brown RH, Wilson JR (1983) Nitrogen response of *Panicum* species differing in CO_2 fixation pathways. II. CO_2 exchange characteristics. *Crop Science*, **23**, 1154-1159.

Bruun-Rasmussen M, Madsen CT, Jessing S, Albrechtsen M (2007) Stability of *Barley stripe mosaic virus*-induced gene silencing in barley. *Molecular Plant-Microbe Interaction*, **20**, 1323-1331.

Buchanan BB, Gruissem W, Jones RL (2000) Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, MD.

Buckler ES, Holland JB, Bradbury PJ *et al*. (2009) The genetic architecture of maize flowering time. *Science*, **325**, 714-718.

Buleon A, Colonna P, Planchot V, Ball S (1998) Starch granules: structure and biosynthesis. *International Journal of Biological Macromolecules*, **23**, 85-112.

Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant Journal*, **39**, 734-746.

Burton RA, Gibeaut DM, Bacic A, Findlay K, Roberts K, Hamilton A, Baulcombe DC, Fincher GB (2000) Virus-induced silencing of a plant cellulose synthase gene. *The Plant Cell*, **12**, 691-705.

Cai W, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theoretical and Applied Genetics*, **104**, 1217-1228.

Carroll TW (1986) Hordeiviruses: biology and pathology. In: The *Plant Viruses, The Rod-shaped Plant Viruses*. (eds. Van Regenmortel MH, Fraenkel-Conrat H), pp. 373-395. Plenum, New York, NY.

Cellulosic biofuels begin to flow but in lower volumes than foreseen by statutory targets (2013) *U.S. Energy Information Administration* <http://www.eia.gov/todayinenergy/detail.cfm?id=10131 >.

Chae WB (2012) Cytogenetics and genome structure in genus *Miscanthus*, a potential source of bioenergy feedstocks. PhD dissertation. University of Illinois at Urbana-Champaign.

Chae WB, Hong SJ, Gifford JM, Rayburn AL, Sacks EJ, Juvik JA (2014) Plant morphology, genome size, and SSR markers differentiate five distinct taxonomic groups among accessions in the genus *Miscanthus*. *Global Change Biology Bioenergy*, **6**, 646-660.

Chae WB, Hong SJ, Gifford JM, Rayburn AL, Widholm JM, Juvik JA (2013) Synthetic polyploidy production of *Miscanthus sacchariflorus*, *Miscanthus sinensis*, and *Miscanthus* x *giganteus*. *Global Change Biology Bioenergy*, **5**, 338-350.

Chang MM, Chou TYC, Tsao GT (1981) Structure, pretreatment and hydrolysis of cellulose. *Advances in Biochemical Engineering*, **20**, 15-42.

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*, **11**, 113-116.

Chiko AW (1984) Increased virulence of *Barley stripe mosaic virus* for wild oats: Evidence of strain selection by host passage. *Phytopathology*, **74**, 595-599.

Chisholm ST, Mahajan SK, Whitham SA, Yamamoto ML, Carrington JC (2000) Cloning of the *Arabidopsis* RTM1 gene, which controls restriction of long-distance movement of *Tobacco etch virus*. *Proceedings of the National Academy of Sciences*, **97**, 489-494.

Christian DG, Haase E (2001) Agronomy of Miscanthus. In: *Miscanthus for energy and fibre* (eds. Jones M, Walsh M), pp. 21–45, James and James, London, UK.

Chu YE, Morishima H, Oka HI (1969) Parital self-incompatibility found in *Oryza perennis* subsp. *barthii*. *Japan Journal of Genetics*, **44**, 225-229.

Churchhill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963-971.

Clark LV (2015) Genome-wide association analysis of flowering time in *Miscanthus sinensis*. In *Plant and Animal Genome XXIII Conference*. Plant and Animal Genome.

Clark LV, Brummer JE, Głowacka K *et al*. (2014) A footprint of past global climate change on the diversity and population genetic structure of *Miscanthus sinensis*. *Annals of Botany*. doi: 10.1093/aob/mcu084.

Clark LV, Stewart JR, Nishiwaki A *et al*. (2015) Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *Journal of Experimental Botany*, 10.1093/jxb/eru511.

Clayton WD, Renvoize SA (1986) *Genera Graminum: Grasses of the World*. *Kew Bulletin Additional Series 13*, *Royal Botanic Gardens, Kew* H.M.S.O, London. ISBN: 0112500064.

Clifton-Brown JC, Chiang YC, Hodkinsn T (2008) *Miscanthus*: genetic resources and breeding potential to enhance bioenergy production. In: *Genetic Improvement of Bioenergy Crops* (ed. Vermerris W), pp. 273–294. Springer, New York.

Clifton-Brown JC, Jones MB (2001) Yield performance of *M*. ×*giganteus* during a 10 year field trial in Ireland. *Aspects of Applied Biology*, **65**, 153-160.

Clifton-Brown JC, Jones MB (2007) Carbon mitigation by the energy crop, *Miscanthus*. *Global Change Biology Bioenergy*, **13**, 2296-2307.

Clifton-Brown JC, Lewandowski I (2000a) Overwintering problems of newly established Miscanthus plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. *New Phytologist*, **148**,287-294.

Clifton-Brown JC, Lewandowski I (2000b) Water use efficiency and biomass partitioning of three different *Miscanthus* genotypes with limited and unlimited water supply. *Annals of Botany*, **86**, 191-200.

Clifton-Brown JC, Lewandowski I (2002) Screening Miscanthus genotypes in field trials to optimise biomass yield and quality in Southern Germany. *European Journal of Agronomy*, **16**, 97-110.

Clifton-Brown JC, Lewandowski I, Andersson B *et al*. (2001) Performance of 15 *Miscanthus* genotypes at five sites in Europe. *Agronomy Journal*, **93**, 1013-1019.

Clifton-Brown JC, Lewandowski I, Bangerth F, Jones MB (2002) Comparative responses to water stress in stay-green, rapid- and slow senescing genotypes of the biomass crop, Miscanthus. *New Phytologist*, **154**, 335–345.

Clifton-Brown JC, Stampfl PF, Jones MB (2004) *Miscanthus* biomass production for energy in Europe and its potential contribution to decreasing fossil fuel carbon emissions. *Global Change Biology*, **10**, 509–518.

Conley EC, Saunders VA, Saunders JR (1986) Deletion and rearrangement of plasmid DNA during transformation of *Escherichia coli* with linear plasmid molecules. *Nucleic acids research*, **14**, 8905-8917.

Connor HE (1979) Breeding systems in the grasses – a survey. *New Zealand Journal of Botany*, **17**, 547-574.

Cooper JP (1970) Potential production and energy conversion in temperate and tropical grasses. *Bureau of pastures and forage crops*, **40**, 1-15.

Cooper JP, Tainton NM (1968) Light and temperature requirements for the growth of tropical and temperate grasses. *Herbage Abstracts*, **38**, 167-176.

Corn background (2014) United States Department of Agriculture <http://www.ers.usda.gov/topics/crops/corn/background.aspx#.U2D7rlcw-7s>.

Cornish MA, Hayward MD, Lawrence MJ (1980) Self-incompatibility in ryegrass. III. The joint segregation of S and PGI-2 in *Lolium perenne* L. *Heredity*, **44**, 55-62.

Curaba J, Chen X (2008) Biochemical activities of *Arabidopsis* RNA-dependent RNA polymerase 6. *The Journal of Biological Chemistry*, **283**, 3059-3066.

Dai A (2010) Drought under global warming: A review. *Wiley Interdisciplinary Reviews: Climate Change*, **2**, 45-65.

Darke R (1994) A century of grasses. *Arnoldia*, 54, 3-11.

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics*, **134**, 943-951.

Darwin C (1876) The effects of cross- and self-fertilisation in the vegetable kingdom. John Murray, London, UK.

Decroocq V, Salvador B, Sicard O *et al.* (2009) The determinant of potyvirus ability to overcome the RTM resistance of *Arabidopsis thaliana* maps to the N-terminal region of the coat protein. *Molecular Plant-Microbe Interactions*, **22**, 1302-1311.

Deguchi S, Tsujii K, Horikoshi K (2006) Cooking cellulose in hot and compressed water. *Chemical Communications*, **31**, 3293-3295.

Demirjian K (2015) Thousands in Armenia protest steep hikes in electricity rates. *The Washington Post*. http://www.washingtonpost.com/world/europe/thousands-in-armenia-protest-steep-hikes-in-electricity-rates/2015/06/23/51377fa0-19bb-11e5-bed8-1093ee58dad0_story.html

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

de Nettancourt D (2001) Incompatibility and incongruity in wild and cultivated plants. Berlin/Heidelberg/New York: Springer-Verlag. 322 pp. 2nd ed.

Deuter M (2000) Breeding approaches to improvement of yield and quality in Miscanthus grown in Europe. In: *European Miscanthus improvement (FAIR3 CT-96–1392) Final Report.* (eds Lewandowski I, Clifton-Brown JC), pp. 28–52, Stuttgart, Germany.

Dhont C, Castonguay Y, Nadeau P, Belanger G, Chalifour F-P (2002) Alfalfa root carbohydrates and regrowth potential in response to fall harvests. *Crop Science*, **42**, 754-765.

Dhont C, Castonguay Y, Nadeau P, Belanger G, Drapeau R, Chalifour F-P (2004) Untimely fall harvest affects dry matter yield and root organic reserves in field-grown alfalfa. *Crop Science*, **44**, 144-157.

Díaz-Pendón JA, Ding SW (2008) Direct and indirect roles of viral suppressors of RNA silencing in pathogenesis. Annual Review of Phytopathology, **46**, 303-326.

Ding XS, Schneider WL, Chaluvadi SR, Mian MR, Nelson RS (2006) Characterization of a *Brome mosaic virus* strain and its use as a vector for gene silencing in monocotyledonous hosts. *Molecular Plant-Microbe Interactions*, **19**, 1229-1239.

DiStilio VS (2011) Empowering plant evo-devo: Virus induced gene silencing validates new and emerging model systems. *BioEssays*, **33**, 711-718.

Dohleman FG, Long SP (2009). More productive than maize in the Midwest: how does *Miscanthus* do it? *Plant Physiology*, **150**, 2104-2115.

Donaire L, Wang Y, Gonzalez-Ibeas D, Mayer KF, Aranda MA, Llave C (2009) Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology*, **392**, 203-214.

Downes RW (1969) Differences in transpiration rates between tropical and temperate grasses under controlled conditions. *Planta*, **88**, 261-273.

Downing M, Eaton LM, Graham RL, Langholtz MH, Perlack RD, Turhollow Jr AF, Brandt CC (2011) US billion-ton update: biomass supply for a bioenergy and bioproducts industry (No. ORNL/TM-2011/224). Oak Ridge National Laboratory: Oak Ridge, TN.

Eckebil JP, Ross WM, Gardner CO, Maranville JW (1977) Heritability estimates, genetic correlations, and predicted gains from S_1 progeny tests in three grain sorghum random-mating populations. *Crop Sciences*, **17**, 373-377.

Eckert B, Weber OB, Kirchhof G, Halbritter A, Stoffels M, Hartmann A (2001) *Azospirillum doebereinerae* sp. nov., a nitrogen-fixing bacterium associated with the C_4-grass *Miscanthus*. *International Journal of Systematic and Evolutionary Microbiology*, **51**, 17-26.

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW (1991) Climate change and the evolution of C_4 photosynthesis. *Trends in Ecology & Evolution*, **6**, 95-99.

Energy Act of 2000. Public Law 106-469, 114 Stat. 2029, codified as amended at 42 U.S.C. §§6201.

Energy Independence and Security Act of 2007. Public Law 110-140, 121 Stat. 1492, codified as amended at 42 U.S.C. §§17381.

Energy Policy Act of 2005. Public Law 109-58, 119 Stat. 594, codified as amended at 42 U.S.C. §§15801.

Entani T, Zwano M, Shiba H, Che FS, Isogai A, Takayama S (2003) Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes to Cells*, **8**, 203-213.

Entani T, Takayama S, Iwano M, Shiba H, Che F-S, Isogai A (1999) Relationship between polyploidy and pollen self-incompatibility phenotype in *Petunia hybrida* Vilm. *Bioscience, Biotechnology and Biochemistry*, **63**, 1882-1888.

EPA proposes renewable fuel standards for 2014, 2015, and 2016, and the biomass-based diesel volume for 2017 (2015) *The United States Environmental Protection Agency*. http://www.epa.gov/otaq/fuels/renewablefuels/documents/420f15028.pdf.

Flint-Garcia S, Thuillet AC, Yu J *et al*. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *The Plant Journal*, **44**, 1054-1064.

Foote HG, Ride JP, Franklin-Tong VE, Walker EA, Lawerence MJ, Franklin FC (1994) Cloning and expression of a distinctive class of self-incompatibility (S-) gene from *Papaver rhoeas*. *Proceedings of the National Academy of Science*, **91**, 2265-2269.

Forbes JC, Watson D (1992) Plants in Agriculture. Cambridge University Press, New York, NY.

Franklin-Tong VE, Franklin FCH (2003) Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *Trends in Plant Sciences*, **8**, 598-605.

Franklin-Tong VE, Holdaway-Clarke TL, Straatman KR, Kunkel JG, Hepler PK (2002) Involvement of extracellular calcium influx in the self-incompatibility response of *Papaver rhoeas*. *The Plant Journal*, **29**, 333-345.

Gagne JM, Downes BP, Shiu S-H, Durski A, Vierstra RD (2002) The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proceedings of the National Academy of Science*, **99**, 11519–24.

Gauder M, Graeff-Hönninger S, Lewandowski I, Claupein W (2012) Long-term yield and performance of 15 different Miscanthus genotypes in southwest Germany. *Annals of Applied Biology*, **160**, 126-136.

Gaut BS (2002) Evolutionary dynamics of grass genomes. *New phytologist*, **154**, 15-28.

Gaut B, Doebley J (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Science of the USA*, **94**, 6809-6814.

Geitmann A, Snowman BN, Emons AMC, Franklin-Tong VE (2000) Alterations in the actin cytoskeleton of pollen tubes are induced by the self-incompatibility reaction in *Papaver rhoeas*. *Plant Cell*, **12**, 1239-1252

Gerhardt KE, Wilson MI, Greenberg BM (1999) Tryptophan photolysis leads to a UVB-induced 66 kDa photoproduct of Ribulose-1,5 Bisphosphate Carboxylase/Oxygenase (Rubisco) in vitro and in vivo. *Photochemistry and photobiology*, **70**, 49-56.

Gertz A, Wricke G (1989) Linkage between the incompatibility locus Z and a beta-glucosidase locus in rye. *Plant Breeding*, **102**, 255-259.

Gifford JM (2012) Preliminary quantitative trait loci analysis for biomass traits in *Miscanthus sinensis*. Master's Thesis. University of Illinois at Urbana-Champaign.

Gifford JM, Chae WB, Swaminathan K, Moose SP, Juvik JA (2014) Mapping the genome of *Miscanthus sinensis* for QTL associated with biomass productivity. *Global Change Biology Bioenergy*. DOI: 10.1111/gcbb.12201.

Gillespie T, Boevink P, Haupt S, Roberts AG, Toth R, Valentine T, Chapman S, Oparka KJ (2002) Functional analysis of a DNA-shuffled movement protein reveals that microtubules are dispensable for the cell-to-cell movement of *Tobacco mosaic virus*. *The Plant Cell*, **14**, 1207-1222.

Głowacka K, Adhikari S, Peng J, Gifford J, Juvik JA, Long SP, Sacks EJ (2014a) Variation in chilling tolerance for photosynthesis and leaf extension growth among genotypes related to the C_4 grass *Miscanthus× giganteus*. *Journal of Experimental Botany*, doi: 10.1093/jxb/eru287.

Głowacka K, Clark LV, Adhikari S *et al*. (2014b) Genetic variation in *Miscanthus× giganteus* and the importance of estimating genetic distance thresholds for differentiating clones. *Global Change Biology Bioenergy*. doi: 10.1111/gcbb.12166.

Głowacka K, Jørgensen U, Kjeldsen JB, Kørup K, Spitz I, Sacks EJ, Long SP (2015) Can the exceptional chilling tolerance of C_4 photosynthesis found in *Miscanthus× giganteus* be exceeded? Screening of a novel *Miscanthus* Japanese germplasm collection. *Annals of botany*, **115**, 981-990.

Goldraij A, Kondo K, Lee CB, *et al*. (2006) Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. *Nature*, **439**, 805-810.

Goyal A, Ghosh B, Eveleigh D (1991) Characteristics of fungal cellulases. *Bioresource Technology*, **36**, 37-50.

Gray BN, Ahner BA, Hanson MR (2011) An efficient downstream box fusion allows high-level accumulation of active bacterial beta-glucosidase in tobacco chloroplasts. *Plant Molecular Biology*, **76**, 345-355.

Gritzali M, Brown RD Jr. (1978) The cellulase system of *Trichoderma*: Relationships between purified extracellular enzymes from induced or cellulose-grown cells. *Advances in Chemistry Series*, **18**, 237-260.

Gu T, Mazzurco M, Sulaman W, Matias DD, Goring DR (1998) Binding of an arm repeat protein to the kinase domain of the S-locus receptor kinase. *Proceedings of the National Academy of Sciences*, **95**, 382-387.

Gustafson G, Armour SL (1986) The complete nucleotide sequence of RNA β from the type strain of barley stripe mosaic virus. *Nucleic Acids Research*, **14**, 3895-3909.

Gustafson G, Hunter B, Hanau R, Armour SL, Jackson AO (1987) Nucleotide sequence and genetic organization of barley stripe mosaic virus RNA γ. *Virology*, **158**, 394-406.

Hackauf B, Wehling P (2005) Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. *Theoretical and Applied Genetics*, **110**, 832-845.

Harries P, Ding B (2011) Cellular factors in plant virus movement: At the leading edge of macromolecular trafficking in plants. *Virology*, **411**, 237–243.

Hatch MD, Osmond CB, Slatyer RO (1971) *Photosynthesis and photo-respiration. Wiley-Interscience*.

Haupt S, Duncan GH, Holzberg S, Oparka KJ (2001) Evidence for symplastic phloem unloading in sink leaves of barley. *Plant Physiology*, **125**, 209-218.

Hayman DL (1956) The genetic control of incompatibility in *Phalaris coerulescens* Desf. *Australian Journal of Biological Sciences*, **9**, 321-331.

Hayman DL (1992) The S-Z incompatibility system. In: *Grass evolution and domestication*. (ed. Chapman GP), pp. 117-137. Cambridge University Press, Cambridge, England.

Hayman DL, Richter J (1992) Mutations affecting self-incompatibility in *Phalaris coerulescens*. *Heredity*, **68**, 495-503.

Heaton EA, Clifton-Brown J, Voigt TB, Jones MB, Long SP (2004a) Miscanthus for renewable energy generation: European Union experience and projections for Illinois. *Mitigation and Adaptation Strategies for Global Change*, **9**, 21–30.

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: The potential of Miscanthus. *Global Change Biology Bioenergy*, **14**, 2000-2014.

Heaton E, Voigt T, Long SP (2004b) A quantitative review comparing the yields of two candidate C_4 perennial biomass crops in relation to nitrogen temperature and water. *Biomass Bioenergy*, **27**, 21-30.

Hershko A, Ciechanover A (1998) The ubiquitin system. *Annual Review of Biochemistry*, **67**, 425-479.

Himber C, Dunoyer P, Moissiard G, Ritzenthaler C, Voinnet O (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *The EMBO Journal*, **22**, 4523-4533.

Himken M, Lammel J, Neukirchen D, Czypionka-Krause U, Olfs HW (1997). Cultivation of *Miscanthus* under West European conditions: Seasonal changes in dry matter production, nutrient uptake and remobilization. *Plant and Soil*, **189** 117-126.

Hirayoshi I, Nishiwaki K, Kubono M, Murase T (1957) Cytogenetical studies on forage plants (VI): On the chromosome number of ogi (Miscanthus sacchariflorus). Research Bulletin: Faculty of Agriculture, Gifu University, **8**, 8-13.

Hiscock SJ, Allen AM (2008). Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus. *New Phytologist*, **179**, 286-317.

Hiscock SJ McInnis SM (2003) Pollen recognition and rejection during the sporophytic self-incompatibility response – *Brassica* and beyond. *Trends in Plant Science*, **8**, 606-613.

Hodkinson TR, Chase MW, Lledó MD, Salamin N, Renvoize SA (2002a) Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae Andropogoneae Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid trnL intron and trnL-F intergenic spacers. *Journal of Plant Research*, **115**, 381–392.

Hodkinson TR, Chase MW, Renvoize SA (2002b) Characterization of a genetic resource collection for *Miscanthus* (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. *Annals of Botany*, **89**, 627-636.

Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA (2002c) The use of DNA sequencing (ITS and trnL-F), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). *American Journal of Botany*, **89**, 279-286.

Hodkinson TR, Renvioze SA, Chase MW (1997) Systematics of *Miscanthus*. *Aspects of Biology*, **49**, 189-197.

Hofmann C, Niehl A, Sambade A, Steinmetz A, Heinlein M (2009) Inhibition of *Tobacco mosaic virus* movement by expression of an actin-binding protein. *Plant Physiology*, **149**, 1810-1823.

Holzberg S, Brosio P, Gross C, Pogue GP (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. *The Plant Journal*, **30**, 315-327.

Howe GT, Saruul P, Davis J, Chen THH (2000) Quantitative genetics of bud phenology, frost damage and winter survival in an F_2 family of hybrid poplars. *Theoretical and Applied Genetics*, **101**, 632-642.

How fossil fuels were formed (2013) *United States Department of Energy* <http://www.fe.doe.gov/education/energylessons/coal/gen_howformed.html>.

Hua Z, Kao T-h (2006) Identification and characterization of components of a putative *Petunia* S-locus F-box-containing E3 ligase complex involved in S-RNase-based self-incompatibility. *The Plant Cell*, **18**, 2531-2553.

Hua Z, Kao T-h (2008) Identification of major lysine residues of S3-RNase of *Petunia inflata* involved in ubiquitin-26S proteasome-mediated degradation *in vitro*. *The Plant Journal*, **54**, 1094-1104.

Hua Z, Meng XY, Kao T-h (2007) Comparison of *Petunia inflata* S-locus F-box protein (Pi SLF) with Pi SLF-like proteins reveals its unique function in S-RNase-based self-incompatibility. *The Plant Cell*, **19**, 3593-3609.

Huang CL, Ho CW, Chiang YC *et al*. (2014) Adaptive divergence with gene flow in incipient speciation of *Miscanthus floridulus/sinensis* complex (Poaceae). *The Plant Journal*, **80**, 834-847.

Huang S, Lee H-S, Karunanandaa B, Kao T-H (1994) Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen. *The Plant Cell*, **6**, 1021-1028.

Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, **9**, 22-32.

Hwang OJ, Cho MA, Han YJ *et al*. (2014). *Agrobacterium*-mediated genetic transformation of *Miscanthus sinensis*. *Plant Cell, Tissue and Organ Culture*, **117**, 51-63.

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23-28.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793-800.

Isemura T, Kaga A, Konishi S, Ando T, Tomooka N, Han OK, Vaughan DA (2007) Genome dissection of traits related to domestication in azuki bean (Vigna angularis) and comparison with other warm-season legumes. *Annals of Botany*, **100**, 1053-1071.

Ivanov R, Fobis-Loisy I, Gaude T (2010) When no means no: guide to Brassicaceae self-incompatibility. *Trends in Plant Sciences*, **15**, 387-394.

Iwano M, Shiba H, Funato M, Shimosato H, Takayama S, Isogai A (2003) Immunohistochemical studies on translocation of pollen S-haplotype determinant in self-incompatibility of *Brassica rapa*. *Plant Cell Physiology*, **44**, 428-436.

Jackson AO, Hunter BG, Gustafson GD (1989) Hordeivirus relationships and genome organization. *Annual Review of Phytopathology*, **27**, 95-121.

Jackson AO, Lane LC (1981) Hordeiviruses. In: *Handbook of Plant Virus Infections and Comparative Diagnosis*. (ed. Kurstak E), pp. 565-625. Elsevier, Amsterdam.

Jackson AO, Lim HS, Bragg J, Ganesan U, Lee MY (2009) Hordeivirus replication, movement, and pathogenesis. *Annual Review of Phytopathology*, **47**, 385-422.

James BT, Chen C, Rudolph A *et al*. (2011) Development of microsatellite markers in autopolyploid sugarcane and comparative analysis of conserved microsatellites in sorghum and sugarcane. *Molecular Breeding*, **30**, 661-669.

Janda M, French R, Ahlquist P (1987) High efficiency T7 polymerase synthesis of infectious RNA from cloned *Brome mosaic virus* cDNA and effects of 5'extensions on transcript infectivity. *Virology*, **158**, 259-262.

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *The Plant Journal*, **50**, 574-585.

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. *Genetics*, **135**, 205-211.

Jansen RC (1994) Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, **138**, 871-881.

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, **136**, 1447-1455.

Jensen E, Farrar K, Thomas-Jones S, Hastings A, Donnison I, Clifton-Brown J (2011) Characterization of flowering time diversity in *Miscanthus* species. *Global Change Biology Bioenergy*, **3**, 387-400.

Jenson E, Robson P, Norris J, Cookson A, Farrar K, Donnison I, Clifton-Brown J (2013) Flowering induction in the bioenergy grass *Miscanthus sacchariflorus* is a quantitative short-day response, whilst delayed flowering under long days increases biomass accumulation. *Journal of Experimental Botany*, **64**, 541-552.

Jeżowski S (2008) Yield traits of six clones of Miscanthus in the first 3 years following planting in Poland. *Industrial Crops and Products*, **27**, 65-68.

Jiang J, Zhu M, Ai X, Xiao L, Deng G, Yi Z (2013) Molecular evidence for a natural diploid hybrid between *Miscanthus sinensis* (Poaceae) and *M. sacchariflorus*. *Plant Systematics and Evolution*, **299**, 1367-1377.

John M, Schmidt J, Wandrey C, Sahm H (1982) Gel chromatography of oligosaccharides up to DP60. *Journal of* Chromatography, **247**, 281-288.

Jones RJ, Mansfield TA (1972) Effects of abscisic acid and its esters on stomatal aperture and the transpiration ratio. *Physiologia Plantarum*, **26**, 321-327.

Jørgensen U (1997) Genotypic variation in dry matter accumulation and content of N, K and Cl in *Miscanthus* in Denmark. *Biomass and Bioenergy*, **12**, 155-169.

Juhasz T, Egyhazi A, Reczey K (2005) β-glucosidase production by *Trichoderma reesei*. *Applied Biochemesry Biotechnology*, **121–124**, 243–254.

Jung S, Suyeon K, Bae H, Lim H-S, Bae H-J (2010) Expression of thermostable bacterial β-glucosidase (BglB) in transgenic tobacco plants. *Bioresource Technology*, **101**, 7144-7150.

Kachroo A, Nasrallah ME, Nasrallah JB (2002) Self-incompatibility in the Brassicaceae: receptor-ligand signaling and cell-to-cell communication. *Plant Cell*, **14**, S227-S238.

Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB (2001) Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. *Science*, **293**, 1824-1826.

Kaiser CM (2014) Characterizing phenotypic diversity, genotype by environment interactions, and optimizing selection efficiency of a *Miscanthus* germplasm collection. Master's Thesis. University of Illinois at Urbana-Champaign.

Kakeda K (2009) S locus–linked F-box gene expressed in anthers of *Hordeum bulbosum*. *Plant Cell Reports*, **28**, 1453-1460.

Kakeda K, Ibuki T, Suzuki J, Tadano H, Kurita Y, Hanai Y, Kowyama Y (2008) Molecular and genetic characterization of the S locus in *Hordeum bulbosum* L., a wild self-incompatible species related to cultivated barley. *Molecular Genetics and Genomics*, **280**, 509-519.

Kalantidis K, Schumacher HT, Alexiadis T, Helm JM (2008) RNA silencing movement in plants. *Biology of the Cell*, **100**, 13-26.

Kanai R, Edwards GE (1999) The biochemistry of C_4 photosynthesis. In: *C4 plant biology*. (eds. Sage RF and Monson RK), pp. 49-87. Academic Press, San Diego.

Kang BC, Yeam I, Jahn MM (2005) Genetics of plant virus resistance. *Annual Review of Phytopathology*, **43**, 581-621.

Kasschau KD, Carrington JC (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. *Cell*, **95**, 461-70.

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

Kato H, Namai H (1987) Floral characteristics and environmental factors for increasing natural outcrossing rate for F1 hybrid seed production of rice *Oryza sativa* L. *Japanese Journal of Breeding*, **37**, 318-330.

Kellogg EA (1998) Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Science*, **95**, 2005-2010.

Khavkin E, Coe E (1997) Mapped genomic locations for developmental functions and QTLs reflect concerted groups in maize (*Zea mays* L.). *Theoretical and Applied Genetics*, **95**, 343-352.

Kim C, Wang X, Lee TH, Jakob K, Lee GJ, Paterson AH (2014) Comparative analysis of *Miscanthus* and *Saccharum* reveals a shared whole-genome duplication but different evolutionary fates. *The Plant Cell*, **26**, 2420-2429.

Kim C, Zhang D, Auckland SA *et al*. (2012) SSR-based genetic maps of *Miscanthus sinensis* and *M. sacchariflorus*, and their comparison to sorghum. *Theoretical and Applied Genetics*, **124**, 1325-1338.

Kirkpatrick, JR (2013) Construction and analysis of the Miscanthus genespace. M.S. thesis, Univ. Illinois.

Klaas M, Yang B, Bosch M, Thorogood D, Manzanares C, Armstead IP, Franklin FCH, Barth S (2011) Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in *Lolium. Annals of Botany,* **108**, 677-685.

Klemm D, Heublein B, Fink H-P, Bohn A (2005) Cellulose: Fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition*, **44**, 3358-3393.

Klemm D, Philipp B, Heinze T, Heinze U, Wagenknecht W (1998) Comprehensive cellulose chemistry. I. Fundamentals and analytical methods. *Weinheim*: Wiley-VCH.

Kobayashi K, Zambryski P (2007) RNA silencing and its cell-to-cell spread during Arabidopsis embryogenesis. *The Plant Journal*, **50**, 597-604.

Körnicke F (1890) Über die autogenetische und heterogenetische Befrüchtung bei den Pflanzen. *Verhandlungen des naturhistorischen Vereines des preussischen Rheinlandes*, **5**, 84-99.

Kościańska E, Kalantidis K, Wypijewski K, Sadowski J, Tabler M (2005). Analysis of RNA silencing in agroinfiltrated leaves of *Nicotiana benthamiana* and *Nicotiana tabacum*. *Plant molecular biology*, **59**, 647-661.

Kuehl RO (2000) Design of experiments: Statistical principles of research design and analysis, Ed 2. pp 550-571. Duxbury Press, Pacific Grove, CA.

Kumagai MH, Donson J, Della-Cioppa G, Harvey D, Hanley K, Grill LK (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proceedings of the National Academy of Sciences*, **92**, 1679-1683.

Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah M (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell*, **13**, 627-643.

Lange M, Yellina AL, Orashakova S, Becker A (2013). Virus-induced gene silencing (VIGS) in plants: an overview of target species and the virus-derived vector systems. In: *Virus-Induced Gene Silencing: Methods and Protocols*. (ed. Becker A), pp. 1-14. Humana Press, New York, NY.

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185-199.

Lane LC (1974) The bromoviruses. *Advances in virus research*, **19**, 151.

Langridge P, Baumann U (2008) Self-incompatibility in the grasses. In*: Self-Incompatibility in Flowering Plants*. (ed. Franklin-Tong VE), pp. 275-287. Springer, Berlin-Heidelberg.

Langridge P, Baumann U, Juttner J (1999) Revisiting and revising the self-incompatibility genetics of *Phalaris coerulescens*. *The Plant Cell*, 11, 1826.

Lawrence DM, Jackson AO (1998) Hordeivirus isolation and RNA extraction. *In Plant Virology Protocols* (pp. 99-106). Humana Press.

Lawrence DM, Jackson AO (2001) Requirements for cell-to-cell movement of Barley stripe mosaic virus in monocot and dicot hosts. *Molecular Plant Pathology*, **2**, 65-75.

Le Thierry D'ennequin M, Toupance B, Robert T, Godelle B, Gouyon PH (1999) Plant domestication: A model for studying the selection of linkage. *Journal of Evolutionary Biology*, **12**, 1138-1147.

Lee H-S, Huang S, Kao T-H (1994) S proteins control rejection of incompatible pollen in *Petunia inflata*. *Nature*, 367, 560-563.

Lee WS, Hammond-Kosack KE, Kanyuka K (2012) *Barley stripe mosaic virus*-mediated tools for investigating gene function in cereal plants and their pathogens: virus-induced gene silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiology*, **160**, 582-590.

Lee YN (1993) Manual of the Korean grasses. Ewha Womans University Press, Seoul.

Lesur-Dumoulin C, Lorin M, Bazot M, Jeuffroy MH, Loyce C (2015) Analysis of young *Miscanthus× giganteus* yield variability: a survey of farmers' fields in east central France. *GCB Bioenergy*. DOI: 10.1111/gcbb.12247.

Llave C (2010) Virus-derived small interfering RNAs at the core of plant–virus interactions. *Trends in Plant Science*, **15**, 701-707.

Lewandowski I, Clifton-Brown JC, Scurlock JMO, Huisman W (2000) Miscanthus: European experience with a novel energy crop. *Biomass and Bioenergy*, *19*, 209-227.

Lewandowski I, Schmidt U (2006) Nitrogen, energy and land use efficiencies of miscanthus, reed canary grass and triticale as determined by the boundary line approach. *Agriculture, Ecosystems and Environment*, **112**, 335–346.

Li X, Nield J, Hayman D, Langridge P (1994) Cloning a putative self-incompatibility gene from the pollen of the grass *Phalaris coerulescens*. *The Plant Cell*, **6**, 1923-1924.

Ligation independent cloning (LIC) (2014) *New England Biolabs* <https://www.neb.com/applications/cloning-and-synthetic-biology/ligation-independent-cloning>.

Lin NS, Langenberg WG (1985) Peripheral vesicles in proplastids of barley stripe mosaic virus infected wheat cells contain double-stranded RNA. *Virology*, **142**, 291-298.

Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics*, **141**, 391-411.

Linde-Laursen IB (1993) Cytogenetic analysis of *Miscanthus* '*Giganteus*', an interspecific hybrid. *Hereditas*, **119**, 297-300.

Lindhout P, Korta W, Dijkstra J (1988) Use of detached tomato leaves to test for resistance to *Tomato mosaic virus*. *Netherlands Journal of Plant Pathology*, **94**, 307-310.

Liu B, Morse D, Cappadocia M (2009) Compatible pollinations in *Solanum chacoense* decrease both S-RNase and S-RNase mRNA. *PLoS ONE*, **4**, e5774.

Liu E, Page JE (2008) Optimized cDNA libraries for virus induced gene silencing (VIGS) using tobacco rattle virus. Plant Methods **4**, 5.

Liu JZ, Blancaflor EB, Nelson RS, (2005) The *Tobacco mosaic virus* 126-kilodalton protein, a constituent of the virus replication complex, alone or within the complex aligns with and traffics along microfilaments. *Plant Physiology*, **138**, 1853-1865.

Liu S (2015) High density genetic map of *Miscanthus sinensis* reveals inheritance of zebra stripe. Master's Thesis. University of Illinois at Urbana-Champaign.

Liu S, Clark LV, Swaminathan K, Gifford JM, Juvik JA, Sacks EJ (2015) High density genetic map of *Miscanthus sinensis* reveals inheritance of zebra stripe. *Global Change Biology Bioenergy*, DOI: 10.1111/gcbb.12275

Long SP (1983) C$_4$ photosynthesis at low temperatures. *Plant, Cell, and Environment*, **6**, 345-363.

Long SP (1999) Environmental responses. In: *C$_4$ Plant Biology*. (eds. Sage RF and Monson RK), pp. 215-249. Academic Press, San Diego, CA.

Lu R, Martin-Hernandez AM, Peart JR, Malcuit I, Baulcombe DC (2003). Virus-induced gene silencing in plants. *Methods*, **30**, 296-303.

Lundqvist A (1957) Self-incompatibility in rye. II. Genetic control in the autotetraploid. *Heriditas*, **43**, 467-511.

Lundqvist A (1954) Studies on self-sterility in rye, *Secale cereale* L. *Hereditas*, **40**, 278-294.

Lundqvist A (1968) The mode of origin of self-fertility in grasses. *Hereditas*, **59**, 413-426.

Luu D-T, Qin X, Morse D, Cappadocia M (2000) S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature*, **407**, 649-651.

Ma XF, Jensen E, Alexandrov N *et al*. (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS One*, **7**, e33821.

Mace ES, Jordan DR (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. *Theoretical and Applied Genetics*, **123**, 169-191.

Manzanares C (2013) Genetics of self-incompatibility in perennial ryegrass (*Lolium perenne* L.). PhD thesis. University of Birmingham.

Manzanares C, Studer B, Barth S, Asp T, Thorogood D (2012) Self-incompatibility in ryegrass: Homing in on the genes of the S-locus complex. Presented at the Proceedings of the 7th International Symposium on the Molecular Breeding of Forage and Turf, June 4-7, 2012, Salt Lake City, Utah.

Mao D, Liu T, Xu C, Li X, Xing Y (2011) Epistasis and complementary gene action adequately account for the genetic bases of transgressive segregation of kilo-grain weight in rice. *Euphytica*, **180**, 261-271.

Marathi B, Ramos J, Hechanova SL, Oane RH, Jena KK (2015) SNP genotyping and characterization of pistil traits revealing a distinct phylogenetic relationship among the species of *Oryza*. *Euphytica*, **201**, 131-148.

Martinez-Reyna JM, Vogel KP (2002) Incompatibility system in switchgrass. *Crop Science*, **42**, 1800-1805.

Massachusetts prohibited plants list (2013) *Massachusetts Department of Agricultural Resources* < http://www.mass.gov/eea/docs/agr/farmproducts/docs/prohibited-plant-list-sciname.pdf>.

Matthews REF (1991) *Plant Virology*. Academic Press, San Diego, CA.

Matton DP, Maes O, Laublin G, Xike Q, Bertrand C, Morse D, Cappadocia M (1997) Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. *The Plant Cell*, **9**, 1757-1766.

McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature*, **342**, 955-957.

McCubbin AG, Kao T-H (2000) Molecular recognition and response in pollen and pistil interactions. *Annual Review of Cell and Developmental Biology*, **16**, 333-364.

McKinney HH, Greeley LW (1965) Biological characteristics of *Barley stripe-mosaic virus* strains and their evolution (No. 1324). *US Department of Agriculture*.

McMullen CR, Gardner WS, Myers GA (1978) Aberrant plastids in barley leaf tissue infected with barley stripe mosaic virus. *Phytopathology*, **68**, 317-325.

Mehra PN, Sharma ML (1975) Cytological studies in some central and eastern Himalayan grasses: 1 the Andropogoneae. *Cytologia*, **40**, 61-74.

Meng XY, Hua ZH, Wang N, Fields AM, Dowd PE, Kao T-h (2009) Ectopic expression of S-RNase of *Petunia inflata* in pollen results in its sequestration and non-cytotoxic function. *Sexual Plant Reproduction*, **22**, 263-275.

Meng XY, Sun P, Kao TH (2011) S-RNase-based self-incompatibility in *Petunia inflata*. *Annals of Botany*, **108**, 637-646.

Miguez FE, Villamil MB, Long SP, Bollero GA (2008). Meta-analysis of the effects of management factors on *Miscanthus × giganteus* growth and biomass production. *Agricultural and Forest Meteorology*, **148**, 1280-1292.

Miki D, Itoh R, Shimamoto K (2005) RNA silencing of single and multiple members in a gene family of rice. *Plant Physiology*, **138**, 1903-1913.

Miller GL (1963) Cellodextrins. *Methods in Carbohydrate Chemistry*, **3**, 134-139.

Misawa N, Yamano S, Linden H, Felipe MR, Lucas M, Ikenaga H, Sandmann G (1993) Functional expression of the *Erwinia uredovora* carotenoid biosynthesis gene *crtl* in transgenic plants showing an increase of β-carotene biosynthesis activity and resistance to the bleaching herbicide norflurazon. *The Plant Journal*, **4**, 833-840.

Mi-Yeon LEE, Gorter F, Kim BY, Grindheim J, Ganesan U, Jackson AO (2012) The triple gene block movement protein of the ND18 strain of *Barley stripe mosaic virus* is an avirulence determinant that elicits *Brachypodium distachyon* Bd3-1 resistance. 中国植物病理学会年学术年会论文集.

Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC (2005) Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes & Development*, 19, 2837-2848.

Moss DN (1962) The limiting carbon dioxide concentration for photosynthesis. *Nature*, **173**, 587.

Murfett J, Atherton TL, Mou B, Gasser CS, McClure BA (1994) S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection. *Nature*, **367**, 563-566.

Muzzurco M, Sulaman W, Elina H, Cock JM, Goring DR (2001) Further analysis of the interactions between the *Brassica* S receptor kinase and three interacting proteins (ARC1, THL1 and THL2) in the yeast two-hybrid system. *Plant Molecular Biology*, **45**, 365-376.

Naidu SL, Moose SP, Al-Shoaibi AK, Raines CA, Long SP (2003). Cold tolerance of C_4 photosynthesis in *Miscanthus× giganteus*: adaptation in amounts and sequence of C_4 photosynthetic enzymes. *Plant Physiology*, **132**, 1688-1697.

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes *in trans*. *The Plant Cell*, **2**, 279-289.

Nasrallah JB, KaoT-h, Chen C-H, Goldberg ML, Nasrallah ME (1987) Amino-acid sequence of glycoproteins encoded by three alleles of the S locus of *Brassica oleracea*. *Nature*, **326**, 617-619.

Nayar NM (1967) Prevalence of self-incompatibility in *Oryza barthii* Cheval: Its bearing on the evolution of rice and related taxa. *Genetica*, **38**, 521-527.

Nayar NM (1958). Studies on the origin of cultivated rice *O. sativa* L.: Cytogenetical studies of related species and their hybrids. Thesis. Indian Agricultural Research Institute, New Delhi.

Nicolaides G (2015) South Africans cannot afford another electricity price hike. *Eyewitness News*. <http://ewn.co.za/2015/06/24/South-Africans-cannot-afford-another-electricity-price-hike1>.

Nishiwaki A, Mizuguti A, Kuwabara S *et al*. (2011) Discovery of natural *Miscanthus* (Poaceae) triploid plants in sympatric populations of *Miscanthus sacchariflorus* and *Miscanthus sinensis* in southern Japan. *American Journal of Botany*, **98**, 154-159.

Noor MAF, Cunningham AL, Larkin JC (2001) Consequences of recombination rate variation on quantitative trait locus mapping studies: Simulations based on the *Drosophila melanogaster* genome. *Genetics*, **159**, 581-588.

Oikawa A, Rahman A, Yamashita T, Taira H, Kidou S-I (2007) Virus-induced gene silencing of P23k in barley leaf reveals morphological changes involved in secondary wall formation. *Journal of Experimental Botany*, **58**, 2617-2625.

Oka HI, Morishima H (1967) Variations in the breeding systems of wild rice, *Oryza perennis*. *Evolution*, **21**, 249-258.

Onishi K, Horiuchi Y, Ishigoh-Oka N, Takagi K, Ichikawa N, Maruoka M, Sano Y (2007) A QTL cluster for plant architecture and its ecological significance in Asian wild rice. *Breeding Science*, **57**, 7-16.

Osmond CB, Winter K, Ziegler H (1982) Functional significance of different pathways of CO_2 fixation in photosynthesis. In *Encyclopedia of Plant Physiology (New Series), Vol 12B* (eds. Lange 0, Nobel P, Osmond C, and Ziegler H), pp. 479-547. Springer-Verlag, Berlin.

Paape T, Miyake T, Takebayashi N, Wolf D, Kohn JR (2011) Evolutionary genetics of an S-like polymorphism in Papaveraceae with putative function in self-incompatibility. *PLoS One*, **6**, e23635.

Pacak A, Geisler K, Jorgensen B *et al*. (2010) Investigations of barley stripe mosaic virus as a gene silencing vector in barley roots and in *Brachypodium distachyon* and oat. *Plant Methods*, **6**, 26.

Palauqui JC, Elmayan T, Pollien JM, Vaucheret H (1997) Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *The EMBO Journal*, **16**, 4738-4745.

Paterson AH, Bowers JE, Bruggmann R *et al*. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551-556.

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Science*, **101**, 9903-9908.

Peele C, Jordan CV, Muangsan N, Turnage M, Egelkrout E, Eagle P, Hanley-Bowdoin L, Robertson D (2001) Silencing of a meristematic gene using geminivirus-derived vectors. *The Plant Journal*, **27**, 357-366.

Perlack RD, Wright LL, Turhollow A, Graham RL, Stokes B, Erbach DC (2005) Biomass as feedstock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply. (DOE/GO-102995-2135). Oak Ridge National Laboratory: Oak Ridge, TN.

Petersen KK, Hagberg P, Kristiansen K, Forkmann G (2002). *In vitro* chromosome doubling of *Miscanthus sinensis*. *Plant Breeding*, **121**, 445-450.

Petty IT, Donald RG, Jackson AO (1994) Multiple genetic determinants of *Barley stripe mosaic virus* influence lesion phenotype on *Chenopodium amaranticolor*. *Virology*, **198**, 218-226.

Petty IT, Hunter BG, Wei N, Jackson AO (1989) Infectious *Barley stripe mosaic virus* RNA transcribed in Vitro from full-length genomic cDNA clones. *Virology*, **171**, 342-349.

Petty IT, Jackson AO (1990) Mutational analysis of barley stripe mosaic virus RNAβ. *Virology*, **179**, 712-718.

Pierce J (1998) Prospects for manipulating the substrate specificity of ribulose bisphosphate carboxylase/oxygenase. *Physiologia Plantarum*, **72**, 690-698.

Purkayastha A, Dasgupta I (2009) Virus induced gene silencing: a versatile tool for discovery of gene functions in plants. *Plant Physiology and Biochemistry*, **47**, 967-976.

Pyter R, Heaton E, Dohleman F, Voigt T, Long S (2009) Agronomic experiences with *Miscanthus* x *giganteus* in Illinois, USA. In: *Biofuels: Methods and Protocals*. (ed. Mielenz JR), pp. 41-52. Humana Press, New York.

Qiao H, Wang F, Zhao L, Zhou J, Lai Z, Zhang Y, Robbins TP, Xue, Y (2004a). The F-box protein AhSLF-S2 controls the pollen function of S-RNase–based self-incompatibility. *The Plant Cell Online*, **16**, 2307-2322.

Qiao H, Wang H, Zhao L, Zhou J, Huang J, Zhang Y, Xue Y (2004b) The F-box protein Ah SLF-S_2 physically interacts with S-RNases that may be inhibited by the ubiquitin/26S proteasome pathway of protein degradation during compatible pollination in Antirrhinum. *The Plant Cell*, **16**, 582-595.

Qi X, Bao FS, Xie Z (2009) Small RNA deep sequencing reveals role for *Arabidopsis thaliana* RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS One*, **4**, e4971.

Ramanna H, Ding XS, Nelson RS (2013) Rationale for developing new virus vectors to analyze gene function in grasses through virus-induced gene silencing. In: *Virus-Induced Gene Silencing: Methods and Protocols*. (ed. Becker A), pp. 15-32. Humana Press, New York, NY.

Rayburn A, Biradar D, Bullock D, McMurphy L (1993) Nuclear DNA content in F1 hybrids of maize. *Heredity*, **70**, 294-300.

Rayburn AL, Crawford J, Rayburn CM, Juvik JA (2009) Genome size of three *Miscanthus* species. *Plant Molecular Biology Reporter*, **27**, 184-188.

Renvoize SA (2003) The genus *Miscanthus*. *The Plantsman*, **2**, 207-211.

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity*, **83**, 363-372.

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2003) The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **358**, 1141-1147.

Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Structural & Molecular Biology*, **12**, 340-349.

Ruiz MT, Voinnet O, Baulcombe DC (1998) Initiation and maintenance of virus-induced gene silencing. *The Plant Cell*, **10**, 937-946.

Sacks EJ, Juvik JA, Lin Q, Stewart JR, Yamada T (2013) The gene pool of *Miscanthus* species and its improvement. In: *Genomics of the Saccharinae*. (ed. Paterson AH), pp. 73-101. Springer, New York.

Sage RF, Monson RK, eds. (1999) C$_4$ plant biology. Academic Press, San Diego, CA.

Sage RF, Pearcy RW (1987) The nitrogen use efficiency of C$_3$ and C$_4$ plants. *Plant Physiology*, **84**, 954-958.

Schell DJ, Hinman ND, Wyman CE, Werdene PJ (1990) Whole broth cellulase production for use in simultaneous saccharification and fermentation. *Applied Biochemistry and Biotechnology*, **24-25,** 287–297.

Schmidt MR, Edwards GE (1981) Photosynthetic capacity and nitrogen use efficiency of maize, wheat, and rice: a comparison between C$_3$ and C$_4$ photosynthesis. *Journal of Experimental Botany*, **32**, 459-466.

Schnable PS, Ware D, Fulton RS *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112-1115.

Schon CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics*, **167**, 485-498.

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in *Brassica*. *Science*, **286**, 1697-1700.

Schulein M (2000) Protein engineering of cellulases. *Biochimica et Biophysica Acta*, **1543**, 239-252.

Schwarz H, Liebhard P, Ehrendorfer K, Ruckenbauer P (1994a). The effect of fertilization on yield and quality of *Miscanthus sinensis* 'Giganteus'. *Industrial Crops and Products*, **2**, 153-159.

Schwarz KU, Murphy DPL, Schnug E (1994b) Studies of the growth and yield of *Miscanthus* x *giganteus* in Germany. *Aspects of Applied Biology*, **40**, 533-540.

Scofield SR, Huang L, Brandt AS, Gill BS (2005) Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. *Plant Physiology*, **138**, 2165-2173.

Scofield SR, Nelson RS (2009) Resources for virus-induced gene silencing in the grasses. *Plant Physiology*, **149**, 152-157.

Senthil-Kumar M, Mysore KS (2011) New dimensions for VIGS in plant functional genomics. *Trends in Plant Science*, **16**, 656-665.

Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. *Theoretical and Applied Genetics*, **101**, 1273-1281.

Shakun JD, Clark PU, He F *et al*. (2012) Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, **484**, 49-54.

Sharkey TD (1988) Estimating the rate of photorespiration in leaves. *Physiologia Plantarum*, **73**, 147-152.

Shiba H, Iwano M, Entani T *et al*. (2002) The dominance of alleles controlling self-incompatibility in *Brassica* pollen is regulated at the RNA level. *The Plant Cell*, **14**, 491-504.

Shiba H, Takayama S, Iwano M *et al*. (2001) A pollen coat protein, SP_{11}/SCR, determines the pollen S-specificity in the self-incompatibility of *Brassica* species. *Plant Physiology*, **125**, 2095-2103.

Shield IF, Barraclough TJ, Riche AB, Yates NE (2014) The yield and quality response of the energy grass *Miscanthus× giganteus* to fertiliser applications of nitrogen, potassium and sulphur. *Biomass and Bioenergy*, **68**, 185-194.

Shinozuka H, Cogan N, Smith K, Forster J (2007) Fine-structure genetic and physical mapping of the perennial ryegrass [*Lolium perenne* L.] self-incompatibility loci. *Molecular breeding of forage 2007*, Sapporo, Japan, pp. 98.

Shinozuka H, Cogan N, Smith K, Spangenberg GC, Forster J (2010) Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). *Plant molecular biology*, **72**, 343-355.

Shivanna KR, Heslop-Harrison Y, Heslop-Harrison J (1982) The pollen–stigma interaction in the grasses. III. Features of the self-incompatibility response. *Acta Botanica Neerlandica*, **31**, 307-319.

Shouliang C, Renvoize SA (2006) *Miscanthus*. *Flora of China*, **22**, 581-583.

Sijacic P,Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao T-h (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature*, **429**, 302-305.

Skowyra D, Craig KL, Tyers M, Elledge SJ, Harper JW (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell*, **91**, 209-219.

Slack SA, Shepherd RJ, Hall DH (1975) Spread of seed-borne barley stripe mosaic virus and effects of the virus on barley in California. *Phytopathology*, **65**, 1218-1223.

Sobhy H, Colson P (2012) Gemi: PCR primers prediction from multiple alignments. *Comparative and functional genomics*, **2012**, doi:10.1155/2012/783138.

Spindler DD, Wyman CE, Grohmann K, Mohagheghi A (1989) Simultaneous saccharification and fermentation of pretreated wheat straw to ethanol with selected yeast strains and β-glucosidase supplementation. *Applied Biochemistry and Biotechnology*, **20-21**, 529-540.

Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB (1991) Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Sciences*, **88**, 8816-8820.

Stephenson AG, Doughty J, Dixon S, Elleman C, Hiscock S, Dickinson HG (1997) The male determinant of self-incompatibility in *Brassica oleracea* is located in the pollen coating. *The Plant Journal*, **12**, 1351-1359.

Stireman JO, Dyer LA, Janzen DH *et al*. (2005) Climatic unpredictability and parasitism of caterpillars: implications of global warming. *Proceedings of the National Academy of Science*, **102**, 17384-17387.

Stone SL, Anderson EM, Mullen RT, Goring DR (2003) ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. *The Plant Cell*, **15**, 885-898.

Stone SL, Arnoldo M, Goring DR (1999) A breakdown of *Brassica* self-incompatibility in ARC1 antisense transgenic plants. *Science*, 286, 1729-1731.

Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, **28**, 1102-1104.

Sun Q, Lin Q, Yi ZL, Yang ZR, Zhou F (2010) A taxonomic revision of *Miscanthus* Andersson s.l. (Poaceae) from China. Botanical *Journal of the Linnean Society*, **164**, 178-220.

Suzuki G, Kai N, Hirose T *et al*. (1999) Genomic organization of the S locus: identification and characterization of genes in *SLG*/*SRK* region of S_9 haplotype of *Brassica campestris* (syn. *rapa*). *Genetics*, **153**, 391-400.

Suzuki G, Kakizaki T, Takada Y, Shiba H, Takayama S, Isogai A, Watanabe M (2003) The S haplotypes lacking *SLG* in the genome of *Brassica rapa*. *Plant Cell Reports*, **21**, 911-915.

Suzuki T, Kusaba M, Matsushita M, Okazaki K, Nishio T (2000) Characterization of *Brassica* S-haplotypes lacking S-locus glycoprotein. FEBS Letter, 482, 102-108.

Swaminathan K, Alabady MS, Varala K *et al*. (2010) Genomic and small RNA sequencing of *Miscanthus × giganteus* shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses. *Genome Biology*, **11**, R12.

Swaminathan K, Chae WB, Mitros T *et al*. (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics*, **13**, 142.

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004a) Close split of sorghum and maize genome progenitors. *Genome Research*, **14**, 1916-1923.

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004b) On the tetraploid origin of the maize genome. *Comparative and functional genomics*, **5**, 281-284.

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K (2000) The S receptor kinase determines self-incompatibility in *Brassica* stigma. Nature, **403**, 913-916.

Takayama S, Isogai A (2003) Molecular mechanism of self-recognition in *Brassica* self-incompatibility. *Journal of Experimental Botany*, **54**, 149-156.

Takayama S, Isogai A (2005) Self-incompatibility in plants. *Annual Review of Plant Biology*, **56**, 467-489.

Takayama S, Isogai A, Tsukamoto C, Ueda Y, Hinata K, Okazaki K, Suzuki A (1987) Sequences of S-glycoproteins, products of *Brassica campestris* self-incompatibility locus. *Nature*, 326, 102-105.

Takayama S, Shiba H, Iwano M *et al*. (2000) The pollen determinant of self-incompatibility in *Brassica campestris*. *Proceedings of the National Academy of Sciences*, **97**, 1920-1925.

Takayama S, Shimosato H, Shiba H, Funato M, Che F-S, Watanabe M, Iwano M, Isogai A (2001) Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. *Nature*, **413**, 534-538.

Tan LW, Jackson JF (1988) Stigma proteins of the two loci self-incompatible grass *Phalaris coerulescens*. *Sexual Plant Reproduction*, **1**, 25-27.

Tanksley SD, Ganal MW, Prince JP *et al*. (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics*, **132**, 1141-1160.

Tans P, Keeling R (2014) Trends in atmospheric carbon dioxide. *National Oceanic and Atmospheric Administration* <http://www.esrl.noaa.gov/gmd/ccgg/trends/>.

Terrestrial invasive species: Amur silver grass (*Miscanthus sacchariflorus*) (2014) *Minnesota Department of Natural Resources* <http://www.dnr.state.mn.us/invasives/terrestrialplants/grasses/amursilvergrass.html>.

Thomas CL, Jones L, Baulcombe DC, Maule AJ (2001) Size constraints for targeting post-transcriptional gene silencing and for RNA-directed methylation in *Nicotiana benthamiana* using a potato virus X vector. *The Plant Journal*, **25**, 417-425.

Thomas SG, Franklin-Tong VE (2004) Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature*, **429**, 305-309.

Thomson LJ, Macfadyen S, Hoffman AA (2010) Predicting the effects of climate change on natural enemies of agricultural pests. *Biological Control*, **52**, 296-306.

Thorogood D, Armstead IP, Turner LB, Humphreys MO, Hayward MD (2005) Identification and mode of action of self-incompatibility loci in *Lolium perenne* L. *Heredity*, **94**, 356-363.

Thorogood D, Hayward MD (1991) The genetic control of self-compatibility in an inbred line of *Lolium perenne* L. *Heredity*, **67**, 175-181.

Tomar SS, Sivakumar S, Ganesamurthy K (2012) Genetic variability and heritability studies for different quantitative triats in sweet sorghum [*Sorghum bicolor* (L.) Moench] genotypes. *Electronic Journal of Plant Breeding*, **3**, 806-810.

Toth RL, Pogue GP, Chapman S (2002) Improvement of the movement and host range properties of a plant virus vector through DNA shuffling. *The Plant Journal*, **30**, 593-600.

Tournier B, Tabler M, Kalantidis K (2006) Phloem flow strongly influences the systemic spread of silencing in GFP *Nicotiana benthamiana* plants. *The Plant Journal*, **47**, 383-394.

Turpen TH, Turpen AM, Weinzettl N, Kumagai MH, Dawson WO (1993) Transfection of whole plants from wounds inoculated with *Agrobacterium tumefaciens* containing cDNA of tobacco mosaic virus. *Journal of Virological Methods*, **42**, 227-239.

Tyers M, Jorgensen P (2000) Proteolysis and the cell cycle: with this RING I do thee destroy. *Current Opinion in Genetics and Development*, **10**, 54-64.

Uga Y, Fukuta Y, Ohsawa R, Fujimura T (2003) Variations of floral traits in Asian cultivated rice (*Oryza sativa* L.) and its wild relatives (*O. rufipogon* Griff). *Breeding Science*, **53**, 345-352.

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *The Plant Cell*, **15**, 771-781.

Vales MI, Schon CC, Capettini F *et al*. (2005) Effect of population size on the estimation of QTL: A test using resistance to barley stripe rust. *Theoretical and Applied Genetics*, **111**, 1260-1270.

Valjamae P, Sild V, Pettersson G, Johansson G (1998) The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. *European Journal of Biochemistry*, **253**, 469-475.

VanLoocke A, Twine TE, Zeri M, Bernacchi CJ (2012). A regional comparison of water use efficiency for miscanthus, switchgrass and maize. *Agricultural and Forest Meteorology*, **164**, 82-95.

Van Ooijen JW (2004) MapQTL 5^{TM}, Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, the Netherlands.

Van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics Research*, **93**, 343-349.

Vaistij FE, Jones L, Baulcombe DC (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *The Plant Cell*, **14**, 857-867.

Vanoosthuyse V, Miege C, Dumas C, Cock JM (2001) Two large *Arabidopsis thaliana* gene families are homologous to the Brassica gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. *Plant Molecular Biology*, **16**, 17-34.

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development*, **20**, 759-771.

Vidal J, Chollet R (1997) Regulatory phosphorylation of C$_4$ PEP carboxylase. *Trends in Plant Sciences*, **2**, 230-237.

Vinzant TB, Adney WS, Decker SR, Baker JO, Kinter MT, Sherman NE, Fox JW, Himmel ME (2001) Fingerprinting *Trichoderma reesei* hydrolases in a commercial cellulase preparation. *Applied Biochemical Biotechnology*, **91-93**, 99-107.

Virmani SS, Athwal DS (1973) Genetic variability in floral characters influencing outcrossing in *Oryza sativa* L. *Crop Science*, **13**, 66-67.

Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics*, **143**, 1013-1020.

Voinnet O, Baulcombe DC (1997) Systemic signalling in gene silencing. *Nature*, **389**, 553.

Voinnet O, Vain P, Angell S, Baulcombe DC (1998) Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell*, **95**, 177-187.

Voylokov AV, Fuong FT, Smirnov VG (1993) Genetic studies of self-fertility in rye (*Secale cereal* L.) 1. The identification of genotypes of self-fertile lines for the *Sf* alleles of self-incompatibility genes. *Theoretical and Applied Genetics*, **83**, 616-618.

Wang X, Cao A, Yu C, Wang D, Wang X, Chen P (2010) Establishment of an effective virus induced gene silencing system with BSMV in *Haynaldia villosa*. *Molecular Biology Reports*, **37**, 967-972.

Wang XUN, Yamada T, Kong FJ *et al*. (2011) Establishment of an efficient *in vitro* culture and particle bombardment-mediated transformation systems in *Miscanthus sinensis* Anderss., a potential bioenergy crop. *Global Change Biology Bioenergy*, **3**, 322-332.

Wassenegger M, Krczal G (2006) Nomenclature and functions of RNA-directed RNA polymerases. *Trends in Plant Science*, **11**, 142-151.

Watson L (1990) The grass family, Poaceae. In: *Reproductive versatility in the grasses*. (ed. Chapman GP), pp. 1-31. Cambridge University Press, Cambridge.

Watson L, Dallwitz MJ (1992) The grass genera of the world. CAB International, Wallingford, UK.

Wheeler MJ, de Graaf BHJ, Hadjiosif N *et al*. (2009) Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature*, **459**, 992-995.

Wehling P, Hackauf B, Wricke G (1994) Phosphorylation of pollen proteins in relation to self-incompatibility in rye (*Secale cereale* L.). *Sexual Plant Reproduction*, **7**, 67-75.

Wei F, Coe ED, Nelson W *et al*. (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genetics*, **3**, e123.

Wei S, Marton I, Dekel M, Shalitin D, Lewinsohn E, Bravdo BA, Shoseyov O (2004) Manipulating volatile emission in tobacco leaves by expressing *Aspergillus niger* beta-glucosidase in different subcellular compartments. *Plant Biotechnololgy Journal*, **2**, 341-350.

White LM (1973) Carbohydrate reserves of grasses: A review. *Journal of Range Management*, **26**, 13-18.

Whitehouse HLK (1950) Multiple allelomorph incompatibility of pollen and style in the evolution of angiosperms. *Annals of Botany*, **14**, 198–216.

Wilkins PW, Thorogood D (1992) Breakdown of self-incompatibility in perennial ryegrass at high temperature and its uses in breeding. *Euphytica*, **64**, 65-69.

Wolf S, Deom CM, Beachy RN, Lucas WJ (1989) Movement protein of *Tobacco mosaic virus* modifies plasmodesmatal size exclusion limit. *Science*, **240**, 377-379.

Wong SC (1979) Elevated atmospheric partial pressure of CO_2 and plant growth. I. Interactions of nitrogen nutrition and photosynthetic capacity in C3 and C4 plants. *Oecologia*, **44**, 68-74.

World energy outlook 2013 factsheet (2013) *International Energy Agency*.

World energy outlook: Renewable energy outlook (2013) *International Energy Agency*.

Wricke G (1978) Pseudo-self-compatibility in rye and its utilization in breeding. *Zeitschrift für Pflanzenzüchtung*, **81**, 140-148.

Wricke G, Wehling P (1985) Linkage between an incompatibility locus and a peroxidase isozyme locus (*Prx*7) in rye. *Theoretical and Applied Genetics*, **71**, 289-291.

Xi Q (2000) Investigation on the distribution and potential of giant grasses in China – *Triarrhena*, *Miscanthus*, *Arundo*, *Phragmites* and *Neyraudia* . Cuvillier, Goettingen.

Xu P, Zhang Y, Kang L, Roossinck MJ, Mysore KS (2006) Computational estimation and experimental verification of off-target silencing during posttranscriptional gene silencing in plants. *Plant Physiology*, **142**, 429-440.

Xu Y, McCouch SR, Shen Z (1998) Transgressive segregation of tiller angle in rice caused by complementary gene action. *Crop Sciences*, **38**, 12-19.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. *PLOS Biology*, **2**, E104.

Yamane H, Ikeda K, Ushijima K, Sassa H, Tao R (2003) A pollen-expressed gene for a novel protein with an F-box motif that is very tightly linked to a gene for S-RNase in two species of cherry, *Prunus cerasus* and *P. avium*. *Plant and Cell Physiology*, **44,** 764-769.

Yang B, Thorogood D, Armstead I, Barth S (2008). How far are we from unravelling self-incompatibility in grasses? *New Phytologist*, **178**, 740-753.

Yang B, Thorogood D, Armstead I, Franklin FCH, Barth S (2009) Identification of genes expressed during the self-incompatibility (SI) response in perennial ryegrass (*Lolium perenne* L.). *Plant Molecular Biology*, **70**, 709-723.

Yano M, Harushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theoretical and Applied Genetics*, **95**, 1025-1032.

Yelina NE, Savenkov EI, Solovyev AG, Morozov SY, Valkonen JPT (2002) Long-distance movement, virulence, and RNA silencing suppression controlled by a single protein in hordei- and potyviruses: complementary functions between virus families. *Journal of Virology*, **76**, 12981-12991.

Yu CY, Kim HS, Rayburn AL, Widholm JM, Juvik JA (2009) Chromosome doubling of the bioenergy crop, *Miscanthus* x *giganteus*. *Global Change Biology Bioenergy*, **1**,404-412.

Yuan C, Li C, Yan L, Jackson AO, Liu Z, Han C, Yu J, Li D (2011). A high throughput *Barley stripe mosaic virus* vector for virus induced gene silencing in monocots and dicots. *PLoS One*, **6**, e26468.

Zhang C, Bradshaw JD, Whitham SA, Hill JH (2010) The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. *Plant Physiology*, **153**, 52-65.

Zhang QX, Shen YK, Shao RX *et al*. (2013) Genetic diversity of natural *Miscanthus sinensis* populations in China revealed by ISSR markers. *Biochemical Systematics and Ecology*, **48**, 248-256.

Zhang T, Wyman CE, Jakob K, Yang B (2012) Rapid selection and identification of *Miscanthus* genotypes with enhanced glucan and xylan yields from hydrothermal pretreatment followed by enzymatic hydrolysis. *Biotechnology for Biofuels*, **5**, 56.

Zhang Y-H P, Lynd LR (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and bioengineering*, **88**, 797-824.

Zhao L, Huang J, Zhao Z, Li Q, Sim T, Xue Y (2010) The Skp1-like protein SSK1 is required for cross-pollen compatibility in S-RNase-based self-incompatibility. *The Plant Journal*, **62**, 52-63.

Zimmermann J, Styles D, Hastings A, Dauber J, Jones MB (2014) Assessing the impact of within crop heterogeneity ('patchiness') in young *Miscanthus× giganteus* fields on economic feasibility and soil carbon sequestration. *Global Change Biology Bioenergy*, **6**, 566-576.

Table A.1 Pearson correlation coefficients between all traits are below the diagonal while genetic correlations are above the diagonal. All Pearson correlations are significant with a P-value <0.05. The 10, 11, 12, and 13 refer to 2010, 2011, 2012, and 2013 and denote the year in which the data was collected.

	A10	A11	A12	A13	BC11	BC12	BC13	CC11	CC12	CC13	CCBC11	CCBC12	CCBC13	E11	H11
Anthesis (A10)	1	0.74	0.71	0.65	0.09	0.07	0.11	-0.02	-0.07	-0.03	-0.16	-0.24	-0.18	-0.10	0.05
Anthesis (A11)	0.31	1	0.88	0.80	-0.05	-0.01	-0.07	-0.06	-0.05	-0.02	0.01	-0.12	0.01	-0.02	-0.09
Anthesis (A12)	0.40	0.47	1	0.83	-0.05	-0.05	-0.05	0.03	0.01	-0.04	0.12	0.09	0.00	0.05	-0.14
Anthesis (A13)	0.45	0.43	0.69	1	0.05	0.04	0.08	0.16	0.10	0.10	0.22	0.12	0.09	0.06	0.00
Basal Circ. (BC11)	-0.14	NS	NS	NS	1	0.91	0.91	0.81	0.73	0.71	0.05	-0.22	0.04	-0.78	0.25
Basal Circ. (BC12)	-0.09	NS	-0.10	NS	0.87	1	0.93	0.73	0.82	0.80	0.02	-0.17	0.17	-0.77	0.25
Basal Circ. (BC13)	NS	NS	-0.11	NS	0.83	0.84	1	0.78	0.72	0.80	-0.05	-0.26	0.13	-0.68	0.35
Comp. Circ. (CC11)	-0.14	NS	NS	0.09	0.68	0.65	0.61	1	0.87	0.80	0.53	0.29	0.39	-0.62	0.23
Comp. Circ. (CC12)	-0.14	NS	-0.17	NS	0.60	0.67	0.63	0.68	1	0.86	0.56	0.42	0.51	-0.63	0.15
Comp. Circ. (CC13)	-0.10	NS	NS	NS	0.50	0.58	0.61	0.48	0.51	1	0.30	0.19	0.69	-0.71	0.34
CC to BC Ratio (CCBC11)	NS	NS	NS	0.09	NS	0.14	0.09	0.75	0.41	0.19	1	0.93	0.57	0.04	-0.07
CC to BC Ratio (CCBC12)	-0.11	NS	-0.11	NS	-0.10	-0.15	NS	0.21	0.62	0.08	0.39	1	0.61	0.06	-0.18
CC to BC Ratio (CCBC13)	NS	NS	NS	NS	NS	0.09	NS	0.15	0.17	0.80	0.17	0.14	1	-0.38	0.16
Emergence (E11)	0.11	0.11	NS	NS	-0.52	-0.44	-0.40	-0.45	-0.35	-0.29	-0.18	NS	NS	1	-0.40
Height (H11)	NS	-0.10	-0.22	NS	0.38	0.39	0.40	0.38	0.39	0.24	0.16	0.10	NS	-0.35	1

Table A.1 (Continued—below diagonal)

	A10	A11	A12	A13	BC11	BC12	BC13	CC11	CC12	CC13	CCBC11	CCBC12	CCBC13	E11	H11
Height (H12)	0.14	0.08	NS	0.22	0.33	0.37	0.35	0.35	0.31	0.29	0.17	NS	0.12	-0.25	0.69
Height (H13)	0.11	0.09	-0.04	0.16	0.17	0.26	0.26	0.26	0.31	0.28	0.19	0.14	0.17	-0.15	0.66
Heading Date (HD10)	0.87	0.32	0.40	0.44	-0.20	-0.13	-0.11	-0.20	-0.20	-0.13	-0.11	-0.14	-0.08	0.16	-0.09
Heading Date (HD11)	0.51	0.59	0.73	0.74	NS	NS	NS	-0.08	-0.08	NS	NS	NS	NS	0.17	-0.21
Heading Date (HD12)	0.34	0.44	0.90	0.61	NS	-0.10	-0.10	-0.09	-0.18	NS	NS	-0.13	NS	NS	-0.22
Heading Date (HD13)	0.48	0.49	0.72	0.86	0.10	0.08	0.08	0.09	NS	NS	NS	NS	NS	NS	NS
Leaf Area (LA11)	NS	NS	NS	NS	0.33	0.31	0.32	0.36	0.28	0.29	0.20	NS	0.13	-0.22	0.33
Leaf Area (LA12)	0.09	NS	NS	0.09	0.34	0.36	0.35	0.34	0.29	0.28	0.17	NS	0.09	-0.22	0.43
Leaf Area (LA13)	0.14	NS	NS	0.18	0.28	0.31	0.33	0.34	0.32	0.37	0.22	0.10	0.21	-0.16	0.42
Leaf Length (LL11)	NS	NS	-0.12	NS	0.26	0.29	0.27	0.35	0.30	0.26	0.25	0.09	0.12	-0.18	0.37
Leaf Length (LL12)	NS	NS	NS	0.09	0.26	0.33	0.28	0.31	0.29	0.21	0.19	NS	NS	-0.21	0.50
Leaf Length (LL13)	0.10	NS	NS	0.13	NS	0.14	0.14	0.25	0.26	0.28	0.26	0.20	0.23	NS	0.39
Leaf Rolling (LR12)	NS	0.12	0.29	NS	-0.17	-0.21	-0.26	-0.24	-0.29	-0.10	-0.17	-0.17	NS	NS	-0.40
Leaf Width (LW11)	NS	NS	0.09	NS	0.29	0.26	0.27	0.29	0.22	0.24	0.14	NS	0.11	-0.18	0.24
Leaf Width (LW12	0.09	NS	NS	NS	0.31	0.30	0.31	0.29	0.23	0.26	0.13	NS	0.10	-0.19	0.31
Leaf Width (LW13)	0.12	NS	NS	0.15	0.30	0.31	0.34	0.30	0.29	0.33	0.16	NS	0.16	-0.16	0.34

Table A.1 (Continued—below diagonal)

	A10	A11	A12	A13	BC11	BC12	BC13	CC11	CC12	CC13	CCBC11	CCBC12	CCBC13	E11	H11
Number of Tillers (NT10)	-0.20	NS	NS	0.08	0.34	0.34	0.30	0.33	0.33	0.11	0.14	NS	-0.08	NS	0.13
Number of Tillers (NT11)	-0.15	-0.12	-0.15	NS	0.50	0.50	0.46	0.79	0.51	0.37	0.61	0.14	0.12	-0.33	0.19
Number of Tillers (NT12)	-0.14	NS	-0.16	NS	0.64	0.75	0.64	0.61	0.78	0.46	0.26	0.24	0.10	-0.31	0.33
Number of Tillers (NT13)	NS	NS	NS	0.14	0.65	0.70	0.75	0.50	0.54	0.68	0.10	NS	0.30	-0.30	0.27
Tiller Diameter (TD11)	0.10	0.11	0.17	0.20	0.21	0.17	0.18	0.25	0.20	0.12	0.15	0.09	NS	NS	0.23
Tiller Diameter (TD12)	0.16	NS	NS	0.14	0.12	0.15	0.17	0.19	0.20	0.08	0.14	0.11	NS	-0.11	0.42
Tiller Diameter (TD13)	NS	NS	NS	NS	0.08	0.13	0.15	0.14	0.17	0.26	0.11	NS	0.23	NS	0.29
Vigor Rating (VR10)	-0.13	NS	NS	NS	0.22	0.18	0.17	0.16	0.18	NS	NS	NS	NS	-0.11	0.15
Yield (Y11)	-0.11	NS	-0.10	0.14	0.66	0.66	0.63	0.81	0.76	0.44	0.52	0.30	0.08	-0.45	0.55
Yield (Y12)	NS	NS	-0.13	0.18	0.60	0.68	0.62	0.68	0.86	0.47	0.39	0.42	0.13	-0.31	0.55
Yield (Y13)	NS	NS	NS	0.18	0.61	0.68	0.69	0.56	0.59	0.81	0.20	NS	0.51	-0.32	0.38

Table A.1 (Continued—above diagonal)

	H12	H13	HD10	HD11	HD12	HD13	LA11	LA12	LA13	LL11	LL12	LL13	LR12	LW11	LW12
Anthesis (A10)	0.27	0.20	0.95	0.68	0.61	0.66	0.13	0.17	0.28	-0.01	0.15	0.29	-0.04	0.15	0.17
Anthesis (A11)	0.29	0.16	0.72	0.89	0.87	0.87	-0.04	0.00	-0.06	0.01	0.11	0.04	0.21	-0.04	-0.02
Anthesis (A12)	0.22	0.12	0.70	0.90	0.92	0.87	0.07	0.09	0.16	-0.11	0.04	0.07	0.15	0.09	0.09
Anthesis (A13)	0.30	0.22	0.65	0.81	0.78	0.91	0.07	0.10	0.21	0.00	0.15	0.23	-0.06	0.07	0.08
Basal Circ. (BC11)	0.33	0.19	-0.01	0.11	-0.05	0.12	0.35	0.34	0.36	0.30	0.26	0.07	-0.33	0.31	0.32
Basal Circ. (BC12)	0.33	0.25	0.01	0.11	-0.04	0.11	0.33	0.37	0.40	0.30	0.35	0.19	-0.26	0.29	0.33
Basal Circ. (BC13)	0.37	0.29	0.03	0.08	-0.05	0.12	0.32	0.36	0.34	0.27	0.33	0.05	-0.36	0.28	0.33
Comp. Circ. (CC11)	0.37	0.20	-0.15	0.11	0.01	0.17	0.39	0.37	0.34	0.36	0.27	0.14	-0.42	0.33	0.34
Comp. Circ. (CC12)	0.25	0.20	-0.14	0.07	0.02	0.10	0.35	0.34	0.35	0.37	0.32	0.26	-0.25	0.29	0.31
Comp. Circ. (CC13)	0.46	0.32	-0.09	0.03	-0.04	0.09	0.39	0.41	0.34	0.36	0.39	0.20	-0.26	0.33	0.37
CC to BC Ratio (CCBC11)	0.15	0.03	-0.26	0.06	0.09	0.15	0.23	0.21	0.18	0.26	0.15	0.26	-0.25	0.19	0.19
CC to BC Ratio (CCBC12)	-0.13	-0.07	-0.29	-0.05	0.10	0.02	0.07	-0.01	0.01	0.15	-0.03	0.22	0.00	0.03	0.00
CC to BC Ratio (CCBC13)	0.33	0.22	-0.19	-0.05	-0.01	0.02	0.24	0.23	0.18	0.25	0.24	0.31	0.02	0.20	0.21
Emergence (E11)	-0.44	-0.35	0.05	-0.07	0.00	-0.08	-0.36	-0.42	-0.44	-0.35	-0.46	-0.30	-0.08	-0.30	-0.37
Height (H11)	0.87	0.84	0.02	-0.07	-0.11	-0.01	0.37	0.45	0.46	0.41	0.61	0.54	-0.46	0.30	0.34

Table A.1 (Continued)

	H12	H13	HD10	HD11	HD12	HD13	LA11	LA12	LA13	LL11	LL12	LL13	LR12	LW11	LW12
Height (H12)	1	0.92	0.26	0.23	0.22	0.33	0.52	0.55	0.55	0.60	0.69	0.62	-0.36	0.41	0.43
Height (H13)	0.75	1	0.18	0.10	0.14	0.24	0.49	0.53	0.52	0.58	0.68	0.55	-0.36	0.38	0.41
Heading Date (HD10)	0.11	0.07	1	0.69	0.62	0.65	0.10	0.16	0.25	-0.08	0.09	0.23	-0.02	0.12	0.16
Heading Date (HD11)	0.10	NS	0.53	1	0.87	0.88	-0.02	0.01	0.07	-0.10	0.06	0.09	0.15	0.00	0.00
Heading Date (HD12)	NS	NS	0.35	0.71	1	0.84	0.08	0.09	0.15	-0.11	-0.02	0.01	0.16	0.11	0.10
Heading Date (HD13)	0.24	0.16	0.46	0.81	0.67	1	0.04	0.07	0.14	-0.02	0.13	0.14	0.05	0.05	0.05
Leaf Area (LA11)	0.40	0.40	NS	NS	NS	NS	1	0.99	0.98	0.30	0.40	0.33	-0.44	0.98	0.97
Leaf Area (LA12)	0.50	0.47	NS	NS	NS	NS	0.84	1	1.00	0.27	0.43	0.39	-0.49	0.96	0.98
Leaf Area (LA13)	0.50	0.55	0.12	NS	NS	0.12	0.79	0.82	1	0.23	0.37	0.43	-0.49	0.95	0.98
Leaf Length (LL11)	0.44	0.45	NS	-0.11	-0.13	NS	0.39	0.24	0.24	1	0.89	0.76	-0.08	0.10	0.10
Leaf Length (LL12)	0.52	0.52	NS	NS	-0.12	0.09	0.26	0.52	0.35	0.50	1	0.79	-0.22	0.22	0.24
Leaf Length (LL13)	0.44	0.51	NS	NS	NS	NS	0.21	0.26	0.58	0.43	0.47	1	-0.27	0.16	0.21
Leaf Rolling (LR12)	-0.29	-0.30	NS	0.15	0.31	NS	-0.26	-0.33	-0.31	-0.10	-0.16	-0.14	1	-0.43	-0.47
Leaf Width (LW11)	0.30	0.30	NS	NS	0.10	NS	0.96	0.83	0.77	0.14	0.13	NS	-0.25	1	0.99
Leaf Width (LW12	0.38	0.35	0.08	NS	NS	NS	0.87	0.95	0.81	0.10	0.25	0.13	-0.31	0.92	1
Leaf Width (LW13)	0.41	0.43	0.10	NS	0.10	0.11	0.85	0.87	0.93	0.11	0.22	0.27	-0.31	0.88	0.92

Table A.1 (Continued—below diagonal)

	H12	H13	HD10	HD11	HD12	HD13	LA11	LA12	LA13	LL11	LL12	LL13	LR12	LW11	LW12
Number of Tillers (NT10)	NS	NS	-0.27	NS	NS	0.11	NS	NS	NS	0.12	0.10	NS	-0.09	-0.11	-0.12
Number of Tillers (NT11)	0.10	NS	-0.19	-0.16	-0.15	NS	NS	0.08	NS	0.17	0.16	0.13	-0.16	NS	NS
Number of Tillers (NT12)	0.27	0.18	-0.18	NS	-0.16	0.10	NS	NS	NS	0.28	0.23	0.17	-0.21	-0.12	-0.12
Number of Tillers (NT13)	0.32	0.21	-0.09	NS	NS	0.17	NS	NS	NS	0.24	0.24	0.15	-0.16	NS	NS
Tiller Diameter (TD11)	0.34	0.29	0.08	0.19	0.15	0.22	0.47	0.40	0.42	0.27	0.19	0.18	-0.10	0.44	0.40
Tiller Diameter (TD12)	0.48	0.44	0.14	NS	-0.09	0.11	0.43	0.51	0.47	0.26	0.41	0.30	-0.31	0.39	0.44
Tiller Diameter (TD13)	0.38	0.48	NS	NS	NS	NS	0.36	0.41	0.46	0.22	0.27	0.30	-0.12	0.33	0.37
Vigor Rating (VR10)	NS	NS	-0.14	NS	NS	NS	NS	NS	NS	NS	0.09	NS	-0.11	NS	NS
Yield (Y11)	0.47	0.39	-0.16	NS	-0.13	0.13	0.38	0.38	0.40	0.41	0.36	0.30	-0.36	0.30	0.31
Yield (Y12)	0.56	0.44	-0.09	NS	-0.15	0.17	0.35	0.40	0.38	0.37	0.42	0.32	-0.42	0.26	0.30
Yield (Y13)	0.50	0.44	NS	NS	NS	0.19	0.32	0.34	0.39	0.33	0.34	0.32	-0.20	0.25	0.27

Table A.1 (Continued—above diagonal)

	LW13	NT10	NT11	NT12	NT13	TD11	TD12	TD13	VR10	Y11	Y12	Y13
Anthesis (A10)	0.24	-0.11	-0.23	-0.08	0.08	0.32	0.33	0.13	-0.23	0.06	0.09	0.11
Anthesis (A11)	-0.10	0.03	-0.27	0.01	-0.11	0.39	0.22	0.37	-0.36	0.06	0.10	0.18
Anthesis (A12)	0.14	0.03	-0.21	-0.03	0.00	0.44	0.26	0.12	-0.27	0.08	0.15	0.11
Anthesis (A13)	0.17	0.16	-0.05	0.10	0.22	0.35	0.23	0.15	0.06	0.23	0.28	0.28
Basal Circ. (BC11)	0.37	0.46	0.69	0.67	0.73	0.21	0.06	0.17	0.64	0.71	0.65	0.69
Basal Circ. (BC12)	0.41	0.43	0.73	0.77	0.83	0.13	0.06	0.20	0.51	0.77	0.76	0.78
Basal Circ. (BC13)	0.36	0.43	0.67	0.68	0.68	0.18	0.14	0.24	0.66	0.67	0.70	0.76
Comp. Circ. (CC11)	0.33	0.54	0.79	0.65	0.49	0.28	0.12	0.20	0.90	0.85	0.79	0.76
Comp. Circ. (CC12)	0.32	0.49	0.85	0.80	0.65	0.19	-0.03	0.26	0.74	0.87	0.88	0.79
Comp. Circ. (CC13)	0.33	0.34	0.60	0.64	0.63	0.29	0.30	0.35	0.45	0.74	0.83	0.92
CC to BC Ratio (CCBC11)	0.14	0.27	0.38	0.20	-0.10	0.20	0.14	0.09	0.62	0.56	0.50	0.27
CC to BC Ratio (CCBC12)	-0.04	0.12	0.29	0.13	-0.06	0.12	-0.13	0.12	0.40	0.26	0.30	0.12
CC to BC Ratio (CCBC13)	0.14	0.02	0.21	0.19	0.21	0.25	0.31	0.33	-0.13	0.38	0.48	0.60
Emergence (E11)	-0.38	-0.36	-0.43	-0.54	-0.64	-0.16	-0.05	-0.22	-0.87	-0.59	-0.54	-0.53
Height (H11)	0.35	0.05	-0.06	0.04	0.11	0.37	0.58	0.61	0.68	0.31	0.36	0.42

Table A.1 (Continued—above diagonal)

	LW13	NT10	NT11	NT12	NT13	TD11	TD12	TD13	VR10	Y11	Y12	Y13
Height (H12)	0.43	0.09	-0.05	0.12	0.21	0.60	0.66	0.69	0.52	0.56	0.54	0.60
Height (H13)	0.42	0.01	-0.18	0.03	0.04	0.51	0.64	0.77	0.28	0.33	0.42	0.46
Heading Date (HD10)	0.22	-0.25	-0.37	-0.11	0.05	0.30	0.29	0.07	-0.37	0.00	0.05	0.04
Heading Date (HD11)	0.04	0.11	-0.09	0.15	0.25	0.35	0.11	0.08	-0.11	0.19	0.21	0.25
Heading Date (HD12)	0.14	0.04	-0.21	-0.02	-0.01	0.39	0.17	0.14	-0.27	0.07	0.14	0.10
Heading Date (HD13)	0.10	0.17	-0.06	0.15	0.26	0.39	0.17	0.15	0.07	0.22	0.27	0.29
Leaf Area (LA11)	0.98	-0.17	-0.01	-0.17	-0.05	0.66	0.68	0.61	0.15	0.47	0.45	0.37
Leaf Area (LA12)	0.98	-0.22	-0.02	-0.15	-0.05	0.65	0.72	0.65	0.20	0.45	0.47	0.38
Leaf Area (LA13)	0.97	-0.11	-0.07	-0.16	-0.07	0.67	0.66	0.73	-0.29	0.48	0.44	0.36
Leaf Length (LL11)	0.06	0.19	0.20	0.29	0.31	0.28	0.33	0.47	0.40	0.45	0.45	0.44
Leaf Length (LL12)	0.19	0.05	0.08	0.23	0.13	0.35	0.54	0.50	0.58	0.36	0.50	0.48
Leaf Length (LL13)	0.20	0.09	-0.04	0.10	0.09	0.29	0.41	0.66	-0.37	0.40	0.41	0.31
Leaf Rolling (LR12)	-0.47	-0.17	-0.23	-0.08	-0.20	-0.22	-0.43	-0.31	-1.02	-0.48	-0.39	-0.36
Leaf Width (LW11)	1.00	-0.21	-0.05	-0.24	-0.11	0.65	0.65	0.54	0.04	0.40	0.37	0.29
Leaf Width (LW12	1.01	-0.26	-0.04	-0.21	-0.08	0.64	0.67	0.58	0.03	0.41	0.40	0.31

Table A.1 (Continued)

	LW13	NT10	NT11	NT12	NT13	TD11	TD12	TD13	VR10	Y11	Y12	Y13
Leaf Width (LW13)	1	-0.13	-0.08	-0.19	-0.08	0.66	0.60	0.63	-0.36	0.43	0.37	0.32
Number of Tillers (NT10)	NS	1	0.66	0.67	0.62	-0.15	-0.31	-0.13	1.05	0.53	0.43	0.40
Number of Tillers (NT11)	NS	0.27	1	0.87	0.51	-0.29	-0.40	-0.15	0.81	0.66	0.63	0.51
Number of Tillers (NT12)	NS	0.42	0.54	1	0.86	-0.23	-0.38	-0.12	0.82	0.72	0.73	0.69
Number of Tillers (NT13)	NS	0.29	0.46	0.73	1	-0.01	-0.17	-0.13	-0.09	0.66	0.68	0.73
Tiller Diameter (TD11)	0.43	NS	-0.31	NS	NS	1	0.81	0.58	-0.01	0.37	0.37	0.41
Tiller Diameter (TD12)	0.44	NS	NS	-0.13	NS	0.43	1	0.81	0.15	0.18	0.25	0.36
Tiller Diameter (TD13)	0.42	-0.10	NS	NS	-0.12	0.21	0.34	1	-0.09	0.29	0.41	0.42
Vigor Rating (VR10)	NS	0.33	0.09	0.19	0.10	0.08	NS	NS	1	0.76	0.89	0.55
Yield (Y11)	0.35	0.36	0.56	0.70	0.53	0.28	0.27	0.13	0.19	1	0.94	0.79
Yield (Y12)	0.33	0.31	0.46	0.79	0.59	0.28	0.38	0.20	0.19	0.83	1	0.93
Yield (Y13)	0.34	0.18	0.38	0.59	0.85	0.20	0.22	0.30	0.11	0.59	0.66	1

Table A.2 Directions to make 1 liter of lysogeny broth (A) and SOB media (B). To make LB or SOB plates, add 15 g/L agarose prior to autoclaving.

A)

10 g tryptone
5 g yeast extract
10 g NaCl
Fill to 1 liter with ddH$_2$O
Adjust pH to 7 with NaOH
Autoclave

B)

20 g tryptone
5 g yeast extract
0.5 g NaCl
10 ml 250 mM KCl
Fill to 1 L with ddH$_2$O
Adjust pH to 7 with NaOH
Autoclave
Add 5 ml filter sterilized 2 M MgCl$_2$ prior to use

*Add 5 ml 1 M filter sterilized glucose to 245 ml SOB media to make 250 ml of SOC media

Table A.3 The sequence of the CRTISO, IspF, IspH, and PDS *Miscanthus* gene fragments.

Gene	Sequence
CRTISO	NTGGGANATTTCCGATNGTGGCGTTGGTGGCTATTGCAATATCCCTGGCAGATGGCCTTGCTTGAAAAGGGCAGAT GAAATNCGTTACAAGGCGAATGTAACCAATGTTATTCTTGAAAATGGAAAGGCTGTTGGGGTGAGGTTATCAAATG GGAAGGAGTTCTTTGCTAAAACGGTGATATCAAATGCCACTAGATGGGACACGTTTGGAAAACTCTTGAAAGAAAA AGAACTTCCAGAAGAGGAGAAAAACTTTCAAAAGAATTATGTTAAGGCACCATCATTCCTTTCAATTCATATGGGTG TCAAAGCGTCAGTTCTGCCTGCTGGCACTGACTGCCACCATTTTGTACTTGAGGATAATTGGAATAATTTGGAAAAG TCTTATGGAAGCATATTTTTAAGTATCCCTACAGTTCTTGATCCATCATTAGCTCCTGAAGGACATCATATACTTCATGTA TTTACTACTGCAGGCATAGAAGACTGGGAGGGTCTTTCTAGGAAGGAGTATGAAGAGAAAAAAGAGGCGGTGGCC AATGAGATTATAAGAAGGCTTGAAAAGAAGTTGTTTCCTGGTCTTCAAGATTCAATAGTCCTCAAGGAGGTAGNCCT CACCAAAA
IspF	CTANGGCGGCATCACATCCCCCACGAACCGCGGCTGCGAGGCTCACTCTGACGGCGACGTGCTGCTGCACTGCGT GGTGGACGCGATTCTCGGCGCGCTGGGGCTGCCAGACATCGGGCAGATTTTCCCGGACTCCGACCCCCGTTGGAA GGGCGCCGATTCTTCTGTGTTCATGAAGGAAGCTGTGAAATTGATGCATGAAGCAGGCTATGAGCTGGGCAACCTT GATGCTACGCTGATCTTGCAAAGACCAAAAATTAGCCCATTCAAGGAGACAA
IspH	ACCACGTTCCCGAGGACGCATCTGGCTCACCAACGAAATCATTCACAACCCCACCGTCAACAAGAGGTTGGATGAG ATGGGTGTAGAAATCATTCCTGTTGATGCGGGTATCAAGGATTTCAATGTNGTCGGACAAGGTGATGTTGTTGTGTT GCCTGCATTTGGAGCTGCTGTGGAGGAGATGTACACGCTAAATGAGAAGAAAGTGCAGATTGTTGATACGACCTGC CCTTGGGTTTCAAAGGTCTGGAACATGGTCGAAAAGCACAAGAAGAGTGAATATACTTCAATTATTCATGGAAAATA TTCCCACGAAGAAACTGTTGCNACTGCTTCTTTTGCAGGAAAGTACATCATTGTGAAGAATATGGCAGAGGCAACCT ATGTGTGTGATTATATACTTGGTGGCCNACTTGATGGGTCTNTNTCAACAAAANANGANTTCCTTGANANNTTCNN NANANNTGNNTCNCCNNGNTTT
PDS	CCNTNACACAGGCATACNGATTTCAAATGGATGGAAANAGCAGGGTGCTTCCTGNATCGGGNGGAACGCANGNA GGNCNTTTAATGCAATGCTNCAAGGCACTCAAGTTTCATAAATCCTGATGAGNTATCCATGCAGTGCATTTTGATTGC TTTGAACAGATTTCTTCAGGAGAAGCATGGTTCCAAAATGGCATTCTTGGATGGTAGTCCACCTGAAAGGCTATGCA TGCCTATTGTTGATCACATTCGGTCTAGGGGTGGAGAGGTCCGCTTGAATTCTCGTATTAAGAAGATAGAGCTGAAT CCTGATGGAACTGTAAAACACTTTGCACTTAGCGATGGAACTCTAATAACTGGAGATGCTTATGTTTGTGCAACACCA GTTGATATCTTCAAGCTTCTTGTACCTCAAGAGTGGAGTGAAATTACTTACTTCAAGAAGCTGGAGAAGCTNGTGG GAGTTCCTGA

Table A.4 Grouping of individuals into self-incompatibility phenotypes within population: *Miscanthus sinensis* 'Grosse Fountaine' x *M. sinensis* 'Undine' and reciprocal (GFxUN/ UNxGF), *M. sinensis* 'Strictus' x 'Kaskade' (STxKA), and *M. sacchariflorus* '♀' x *M. sinensis* '♂'. Individuals in the same group are 100% incompatible with each other.

Population	Group	Individuals								
GFxUN/ UNxGF	1	GU008	GU011	GU034	GU067	GU069	GU070	GU087	GU090	GU099
		GU108	GU112	GU115	GU120	GU121	GU129	GU173	GU183	UG025
		UG033	UG042	UG046	UG061	UG063	UG077	UG079	UG080	UG082
		UG088	UG090	UG102	UG112	UG129	UG131	UG132	UG133	UG138
		UG143	UG146	UG153	UG159	UG162	UG164	UG166	UG170	UG179
		UG184	UG192							
	2	GU009	GU010	GU014	GU030	GU037	GU039	GU045	GU052	GU055
		GU056	GU057	GU066	GU096	GU106	GU139	GU140	GU152	GU159
		GU160	GU166	GU175	GU178	GU189	UG049	UG057	UG059	UG072
		UG075	UG076	UG078	UG081	UG091	UG098	UG099	UG105	UG106
		UG114	UG115	UG116	UG121	UG127	UG128	UG134	UG148	UG152
		UG154	UG173	UG175	UG191	UG050	UG120	GU170	GU042	GU147
		GU109								
	3	GU001	GU006	GU021	GU040	GU047	GU059	GU060	GU082	GU085
		GU086	GU094	GU103	GU113	GU122	GU132	GU133	GU141	GU142
		GU149	GU151	GU156	GU158	GU163	GU176	GU181	GU182	GU185
		GU188	GU192	UG005	UG008	UG018	UG023	UG058	UG083	UG097
		UG117	UG130	UG139	UG142	UG147	UG149	UG150	UG169	UG187
		UG188	UG190	GU118	GU136					
	4	GU018	GU023	GU025	GU035	GU041	GU053	GU061	GU075	GU077
		GU084	GU091	GU100	GU111	GU131	GU137	GU165	GU169	GU172
		UG011	UG016	UG020	UG022	UG024	UG034	UG037	UG038	UG045
		UG069	UG070	UG071	UG073	UG084	UG089	UG093	UG140	UG141
		UG155	UG163	UG167	UG168	UG174	UG178	UG180		
STxKA	1	5	60	66	70	85	96	102	103	127
		188	192	218	236	252				
	2	28	30	54	56	63	110	122	123	140
		155	166	176	203	211	243	260		