CYTOGENETICS AND GENOME STRUCTURE IN GENUS *MISCANTHUS*, A POTENTIAL SOURCE OF BIOENERGY FEEDSTOCKS

BY

WON BYOUNG CHAE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor John A. Juvik, Chair
Associate Professor Stephen P. Moose
Professor A. Lane Rayburn
Professor Ray Ming
Assistant Professor Erik J. Sacks

ABSTRACT

Miscanthus (subtribe Saccharinae, tribe Andropogoneae, family Poaceae) is a genus of temperate perennial C_4 grasses. Accessions in the genus *Miscanthus* are potential crop candidates of lignocellulosic biomass for energy production, whose high biomass productivity is attractive as a biofuel feedstock. *Miscanthus* is not native to the United States and has been introduced as ornamental accessions by private nurseries and research institutes. Comprehensive taxonomic studes have not been conducted on these U.S. accessions. Previous taxonomic studies have been conducted on the genus *Miscanthus* using morphology and DNA sequence variation. This study (chaper 2) combines information on genome size and ploidy levels and DNA sequence variation to classify *Miscanthus* accessions to aid potential biomass crop improvement programs and to investigate the evolution of the genus. We observed that *Miscanthus* accessions fell into 4 groups, including section *Miscanthus*, section *Triarrhena* and two groups intermediate between two sections, based on morphology and genome size. Sixteen simple sequence repeat (SSR) primer pairs were selected based on amplification and polymorphism across three genera, *Miscanthus*, *Saccharum* and *Erianthus*. Morphology, genome size and SSR genotyping of 42 accessions including diploid and triploid interspecific hybrid progeny suggested that there are three *Miscanthus* species (*M. sinensis*, *M. sacchariflorus* and *M. x giganteus*), one *M. sacchariflorus* variety and one putative hybrid among *Miscanthus* accessions, which were clearly separated from the other two genera. The species status of *M. floridulus* remains in question. The evolution of *Miscanthus* and related genera is discussed based on genome size, ploidy level, cluster analysis and geographical distribution.

Based on genome size and chromosome number comparision between *Miscanthus* and other related genera, we hypothesize large-scale duplications have occurred in recent ancestors of *Miscanthus*. Owing to the complexity of the *Miscanthus* genome and the complications of self-incompatibility, a complete genetic map with a high density of markers has not yet been developed. As described in chapter 3, a cross between two *M. sinensis* accessions, 'Grosse fontaine' and 'Undine' was made to produce 221 segrating progeny as a mapping population. Simple sequence repeat (SSR) markers from sugarcane expressed sequence tags (EST) and genomic sequences were screened in the two parental *M. sinensi*s accessions. Single nucleotide polymorphism (SNP) markers from deep transcriptome sequencing (RNAseq) were also used for map construction. A total of 210 SSR markers and 658 single nucleotide polymorphism (SNP) markers were validated via segregation in the full sib F_1 mapping population. A genetic map for *M. sinensis* was constructed that was resolved into 19 linkage groups, the haploid chromosome number expected from cytological evidence. Comparative genomic analysis revealed genome-wide duplication in *Miscanthus* relative to *S. bicolor*, with subsequent insertional fusion of a pair of chromosomes. The utility of the map is confirmed by the identification of two paralogous C_4-pyruvate, phosphate dikinase (C_4-PPDK) loci in *Miscanthus*, at positions syntenic to the single orthologous gene in sorghum. The *M. sinensis* map and comparative mapping with sorghum suggests that the genus *Miscanthus* experienced an ancestral tetraploidy and chromosome fusion prior to its diversification, but after its divergence from the closely related sugarcane clade. The genetic map for *Miscanthus* is useful in biological discovery and breeding efforts to improve this emerging biofuel crop, and also provide a valuable resource for understanding genomic responses to tetraploidy and chromosome fusion.

In chapter 4, artificial genome doubling of various *Miscanthus* accessions with antimototic agents was used to understand the phenotypic responses to whole genome duplication in *Miscanthus*. Interspecific manipulation of ploidy levels is also a potential strategy for *Miscanthus* crop improvement to generate superior germplasm and to circumvent reproductive barriers for the introduction of new genetic variation into core germplasm. Therefore, synthetic autotetraploid lines of *M. sacchariflorus* and *M. sinensis*, and autoallohexaploid *M. x giganteus* were produced in tissue culture from oryzalin treatments to seed- and immature inflorescence-derived callus lines. Genome doubling of diploid *M. sinensis*, *M. sacchariflorus,* and triploid *M. x giganteus* to generate tetraploid and hexaploid lines was confirmed by stomata size, nuclear DNA content, and chromosome counts. A putative pentaploid line was also identified among the *M. x giganteus* synthetic polyploid lines by nuclear DNA content and chromosome counts. Comparisons of phenotypic performance of synthetic polyploid lines with their diploid and triploid progenitors in the greenhouse found species-specific differences in plant tiller number, height, and flowering time among the doubled lines. Stem diameter tended to increase after polyploidization but there were no significant improvement in biomass traits. Under field conditions, *M. x giganteus* synthetic polyploid lines showed greater phenotypic variation, in terms of plant height, stem diameter and tiller number, than their progenitor lines. Production of synthetic autopolyploid lines displaying significant phenotypic variation suggests that ploidy manipulation can introduce genetic diversity in the limited *Miscanthus* germplasm currently available in the United States. The role of polyploidization in the evolution and breeding of the genus *Miscanthus* is discussed.

DEDICATION

I dedicate my dissertation work to my family. I would like to give special thanks to my parents, Soo Yeol Chae and Seong Ja Kim, my brother, Min Byoung Chae and his wife, Soo Hyun Hwang, and my two nephews Seung Heon and Jun Ik Chae, for their support and love.

ACKNOWLEDGMENT

I wish to thank my committee members. My thesis reseach would not have been possible without their support and advice. Foremost, I would like to express my sincere gratitude to my advisor, Dr. John Juvik, for his support and guidance throughout my graduate work. Special thanks to Dr. Stephen Moose, for giving me the opportunity to perform research in EBI. I would like to thank Dr. A. Lane Raybrun for offering me the opportunity to work in his Lab and Dr. Erik Sacks for providing me with his valuable *Miscanthus* accessions. My sincere thanks also goes to Dr. Ray Ming for serving on my thesis committee and for helpful advice.

I am very thankful to my fellow labmates in Dr. Juvik's Lab: Dr. Se Jin Hong who was sabbatical professor from Gangneung-Wonju National University in Korea, Hyoung Seok Kim, Justin Gifford, Kang Mo Ku and Talon Becker for their friendship and assistance. Thanks also to Feedstock genomics group in EBI: Dr. Kankshita Swaminathan, Dr. Bosola Oladeinde, Adam Barling, Brandon James, Liang Xie, Ornella Wa Ngamboma, Dr. Ashley Spence and Dr. Katarzyna Glowacka in UIUC and Dr. Daniel Rokhsar, Dr. Therese Mitros and Dr. Megan Hall in UC Berkeley.

TABLE OF CONTENTS

CHAPTER 1

Literature Review

1.1. Bioenergy and *Miscanthus*

The demand for energy is estimated to grow by more than 50% by 2025, largely due to the increasing energy demands from developing countries (Ragauskas *et al*., 2006). For over a century, our main source of energy has been fossil fuels but reserves are currently limited with imminent depletion. In addition, the increased use of fossil fuels releases large amounts of sequestered carbon in the form of carbon dioxide into the atmosphere and enhances the greenhouse effect and global warming. Ways to meet the increasing demand for energy are needed from carbon neutral sources. Biofuels generated from cellulosic and lignocellulosic biomass represents one of the best alternative energy sources. Ethanol from corn, sorghum and sugarcane and biodiesel from soybean are being utilized in automobiles today. Lignocellulosic biomass will serve as the feedstock for the next generation of biofuels. In addition to liquid fuel, biomass can be used to generate heat and electricity as is currently practiced in Europe.

Perennial C_4 grasses are excellent candidates for generating heat and producing biofuels due to their capacity to produce high biomass while reducing greenhouse gas emissions by sequestering carbon in the soil (McLaughlin & Walsh, 1998). *Miscanthus x giganteus* and switchgrass have been identified as promising candidates for low input bioenergy production in the United States and Europe (Lewandowski *et al*., 2003). The perennial C_4 grass, *M. x giganteus* has high biomass yields in trial plots in the Midwest and can be grown in a range of environments using conventional farming practices (Khanna *et al*., 2008).

M. x giganteus (2n=3x=57) is a triploid and thought to be a hybrid of *M. sinensis* (2n=2x=38) and a tetraploid accession of *M. sacchariflorus* (2n=4x=76) (Greef *et al*., 1997; Hodkinson *et al*., 2002b and c; Rayburn *et al*., 2009). It has been evaluated and successfully cultivated as a biofuel crop for several years in Europe and has performed equally well in the United States (Heaton *et al*., 2008). Its ability to perform cold tolerant C_4 photosynthesis, high water-use efficiency, nutrient reallocation, capability for carbon sequestration and high yields in a variety of different climates and environmental regions make *M. x giganteus* a very promising biofuel crop. (Clifton-Brown *et al*., 2001; Clifton-Brown *et al*., 2002; Clifton-Brown *et al*., 2007; Lewandowski *et al*, 2003; Naidu *et al*., 2003).

The triploid nature of *M. x giganteus* makes it sterile, producing no viable seed (Linde-Laursen, 1993). Therefore, *M. x giganteus* must be clonally propagated via rhizomes or tissue culture. Although the sterility of this accession reduces risk of invasiveness, it poses limitations for both propagation and breeding. It also makes it impossible to introduce desirable traits into *M. x giganteus* via conventional breeding and hybridization. Large-scale monoculture using a single or a few genotypes increases the risk of crop susceptibility to diseases (Clifton-Brown *et al*., 2001) and pests (Prasifka *et al*., 2009). So far, no diseases have been discovered that substantially affect *M. x giganteus* yield (Christian and Haase, 2001) and few insect pests have been discovered that pose serious threats to the plant (Bradshaw *et al*., 2010; Christian *et al*., 1997; Spencer & Raghu, 2009).

For any crop, diverse germplasm collections with extensive genetic variation are needed to drive crop improvement. Prime *Miscanthus* species that are candidate feedstocks have a relatively broad geographic distribution, ranging from eastern Asia to the Pacific islands (Hodkinson *et al*., 1997). In the United States, the number of different *Miscanthus* lines available

in research institutes, botanical gardens and private nurseries is limited, consisting mostly of ornamental *M. sinensis* accessions. However, a larger pool of germplasm exists in the wild. Efforts are underway to add to the current germplasm pool in the United States with accessions from Korea, Japan and China but export and import restrictions have been a major barrier and have slowed germplasm exchange (Jakob *et al*., 2009).

In the United States, *M. sinensis* was first introduced from Japan in the late 1800s (Quinn *et. al*., 2010) and various *Miscanthus* accessions have been introduced thereafter, mainly as ornamentals. *Miscanthus* accessions have been acquired independently by nurseries and are commercially available. However, accurate genotype or species identification is problematic, possibly due to the lack of proper taxonomic knowledge, absence of informative molecular markers and or mislabeling of plants in plots or greenhouses. Currently at the University of Illinois, more than 100 accessions of *Miscanthus* have been collected (mostly ornamentals) and planted both in the field and in the greenhouse. Within this collection from various nurseries and the National Plant Germplasm System of the USDA (USDA-NPGS), errors in identification both at the species level and within each species are apparent as some genotypes fail to exhibit the basic morphological traits associated with the species taxonomy.

1.2. Species determination and genome size evolution

The taxonomic determination of plant species has been a longstanding controversy (Soltis & Soltis, 2009). There are several concepts in speciation. The morphology-based taxonomic species concept (Grant, 1981), involving the grouping of individuals with morphologically similar traits that are distinct from other groups, is still popular among plant taxonomists. This concept is practical for taxonomic purposes but can be subjective since

different taxonomists have different criteria and emphasis on morphological characters. The biological species concept (Mayr, 1942), which groups interbreeding populations that are reproductively isolated from other groups, is common in animals but its application in plants is difficult because of frequent hybridization between species and asexual reproduction (Soltis & Soltis, 2009). The evolutionary species concept distinguishes species by determining ancestral-descendant relationships of populations that have evolved independently from other lineages. This concept requires knowledge of ecological niches, evolutionary tendencies and historical fates (Soltis & Soltis, 2009). Most phylogenetic species concepts that define a species as a group having a shared and unique evolutionary history fall under the evolutionary concept (Boggs, 2001).

Hybridization facilitates speciation and adaptive radiation in both animals and plants. Hybridization can increase genetic variation (Buerkle *et al*., 2000; Riesberg *et al*., 2003) which allows colonization of unexploited niches (Mallet, 2007). Hybrid speciation can occur where two species or populations are sympatric (Mallet, 2007). Hybrid speciation via allopolyploidization is a well known mechanism which leads to immediate reproductive isolation (Nolte & Tautz, 2010). Homoploid hybrid speciation, the hybrid speciation without genome doubling events, is harder to observe and define if hybrids are only weakly reproductively isolated and the genetic components from each parent are not 50% due to backcrossing (Mallet, 2007). About 20 homoploid hybrid species have been reported in plants (Rieseberg, 1997; Gross & Rieseberg, 2005) but the number would be underestimated due to the difficulty of homoploid hybrid detection.

The genus *Miscanthus* belongs to the subtribe Saccharinae Griseb., tribe Andropogoneae Dumort. of the family Poaceae and has been divided into 11-20 species (Clifton-Brown *et al*.,

2008; Hodkinson *et al*., 2002b). Speciation in *Miscanthus* has mainly been determined by morphology-based taxonomic studies and limited molecular marker analsysis. There is little consensus on the definition of the genus *Miscanthus* and the number of species (Clifton-Brown *et al*., 2008; Sun *et al*., 2010). The inclusion of quantitative traits like floral organ sizes (spikelet and inflorescence axis) and leaf anatomy as traits that can distinguish members of the genus *Miscanthus* remain uncertain (Hodkinson *et al*., 2002b). However, there are some key traits that distinguish the two sections (secondary rank between two primary ranks, genus and species in botany), *Miscanthus* and *Triarrhena*. These traits include the presence of rhizomes, the awn in spikelets, buds at the nodes, pilose or glabrous leaves on the abaxial surfaces and adventitious roots (Lee, 1993) (Fig. 1). A recent morphology-based taxonomic study in China, home to many *Miscanthus* species and one of the plausible centers for diversification of *Miscanthus*, reported that the genus *Miscanthus* comprises 6 species, 2 sub-species and 4 varieties (Sun *et al*., 2010).

Molecular marker analysis using Amplified Fragment Length Polymorphisms (AFLP) and Inter Simple Sequence Repeat (ISSR) PCR DNA marker revealed that there are six distinct taxonomic groups within *Miscanthus* including *M. sinensis*, *M. x giganteus*, *M. sacchariflorus*, *M. oligostachyus*, *M. transmorrisonensis* and *M. nepalensis* in the United Kingdom (Hodkinson *et al*., 1997; Hodkinson *et al*., 2002b). These results were generally similar to the morphology-based taxonomic treatment of Lee (1964a, b, c) but *M. sinensis* spp. *condensatus* and *M. floridulus* were embedded in *M. sinensis* groups, which made their species status questionable (Hodkinson *et al*., 2002b). *M. oligostachyus* was included in the *Miscanthus* group although it was separated from other *Miscanthus* species while *M. nepalensis* was grouped more closely with other genera of Saccharinae.

The genome size of a species represents "the DNA content of the monoploid genome or chromosome set" and the DNA C-value or 2C-value stands for "the DNA content of the whole chromosome complement of the karyotype irrespective of the degree of generative polyploidy of the organism" (Greilhuber *et al*., 2005). The C-value means "the DNA amount of the unreplicated haploid chromosome complement" (Swift, 1950; Bennett & Smith, 1976). It is practical for describing even numbered ploidy plant species but difficulties arise from odd-numbered polyploid plant species. Thus, Greilhuber *et al.* (2005) suggested C-value standing for "DNA content of a holoploid genome with chromosome number n" and Cx-value for "DNA content of a monoploid genome with chromosome base number x". "Holoploid" and "monoploid" genomes represent "the whole chromosome set of an organism with chromosome number n" and "one chromosome set of an organism and its DNA having the chromosome base number x", respectively.

Plants have a wide range of genome sizes from the smallest genome of 63 Megabase pairs (Mbp) found in *Genlisea margaretae* (Lentibulariaceae) to the largest of 127 Gigabase pairs (Gbp) in *Fritillaria assyriaca* (Liliaceae) (Dolezel *et al*., 2007). Genome size information of species is a useful measure that can aid taxonomic and evolutionary studies. The estimation of nuclear DNA content is efficient both in defining infrageneric division in a number of taxa (Ohri, 1998) and in determining how genomes grow or shrink during evolution (Bennetzen *et al*., 2005; Leitch *et al*., 2005). It is also useful for determining the ploidy level. The nuclear DNA content is also useful in identifying synthetic or natural hybrids since the genome sizes of hybrids usually follow mid-parent values (Bennett & Litch, 1995; Dolezel, 1997; Rayburn *et al*., 2005).

Two major species in *Miscanthus, M. sinensis* and *M. sacchariflorus*, which are representative of sections *Miscanthus* and *Triarrhena*, respectively, are different in terms of their

2C-value (Rayburn *et al*., 2009). These differences in 2C-value between two species infer that hybrid species can be determined by estimating genome size. The ploidy levels of *Miscanthus* species also vary from diploid to hexaploid (Hodkinson & Renvioze, 2001; Clifton-Brown *et al*., 2008). A recent study revealed natural triploid hybrid plants in sympatric populations of *M. sacchariflorus* and *M. sinensis* in Japan by measuring DNA 2C- values using flow cytometry (Nishiwaki *et al*. 2011).

1.3. Molecular markers and linkage mapping

Molecular markers identify neutral sites of variation at the DNA sequence level, which does not typically result in phenotypic differences. The advantages of using DNA molecular markers are that they can detect neutral sites of variation which do not result in phenotypic differences and, therefore, are much more abundant than morphological markers (Jones *et al*., 1997). Various molecular markers are now available such as restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP), diversity array technology polymorphisms (DArT), simple sequence repeats (SSR), and single-nucleotide polymorphisms (SNP). For *Miscanthus* species, AFLP and Inter-Simple Sequence Repeat (ISSR) PCR DNA markers were developed and used for characterizing germplasm (Greef *et al*. 1997; Hodkinson *et al*., 2002b).

SSR markers have been widely used due to codominant inheritance, multi-allelic nature and easy detection. SSR markers can be the marker of choice for studying genetics in *Miscanthus* species because of inter-species or -genera transferability of SSR markers. Large numbers of SSR markers have been developed in related species including sorghum, sugarcane, corn, wheat, barley and rice and they can be applied in the genus *Miscanthus* (Hernández *et al*. 2001).

Sugarcane and sorghum are taxonomically more closely related to *Miscanthus* than others (Hodkinson *et al*. 2002a) and, thus, a higher success rate is predictable (Jakob *et al*., 2009). Besides, SSRs are PCR based markers and easy and reliable to use. Therefore, multiplexed and automated SSR assays can be applied using capillary-based DNA analyzers in combination with various color florescence-labeled primers to provide large throughput genotyping (Jakob *et al*., 2009).

SNP markers promise higher map resolution due to the large number of polymorphisms present in the plant genome. They also provide higher throughput, lower cost and a lower error rate compared to SSR markers (Gupta *et al*. 2001; Rafalski, 2002). Mining expressed sequence tags (ESTs) is a routine method for SNP discovery (Jakob *et al*., 2009). ESTs can be easily generated by sequencing expressed mRNA in various plant tissues of core accessions. The sequencing data is then aligned and compared among genotypes for SNP discovery. Next generation sequencing technologies such as Illumina, 454 or SOLiD facilitate the identification of many SNPs from different genotypes at much lower cost and considerably reduce the complexity of the sequenced DNA to obtain the necessary sequence redundancy for reliable SNP calling (Ganal *et al*., 2009). This approach has been successfully applied to various plant species including maize and Eucalyptus with relatively low false discovery rates of 20% through the sequencing of a large number of ESTs (Imelfort *et al*., 2009). In *Miscanthus*, 1,596 SNPs from ESTs have been produced from Illumina and 454 sequencing technologies which were collected from two individual lines of *M. sinensis* and applied to the creation of a genetic linkage map of the *M. sinensis* genome (Swaminathan *et al*., 2012) and will be applied to the genotyping of *Miscanthus* germplasm.

Linkage mapping is arraying polymorphic DNA markers in order, calculating the genetic distances among them and assigning them to their linkage groups based on recombination frequency from all pairwise combinations (Jones *et al*., 1997). Linkage maps are powerful tools for accelerating breeding programs through marker assisted selection and recombination. Various mapping populations from two inbred parental lines have been used for genetic studies such as F_2, backcross (BC), doubled haploid (DH) and recombinant inbred lines (RIL). However, most *Miscanthus* species are obligate outcrossing plants and, thus, not amenable to selfing to generate inbred lines for the construction of genetic linkage maps. The pseudo-testcross mapping strategy, which was first put in practice in *Eucalyptus* (Grattapaglia & Sederoff, 1994), employs dominant markers that follow testcross configuration in heterozygous individuals and has been widely used for outcrossing plants. The strategy only requires first generation-progeny populations (F_1) from heterozygous parental lines and calculates dominant markers following testcross configuration and, thus, it is fast and easy to construct genetic linkage maps in outcrossing species.

High density linkage maps are essential for whole genome sequencing together with bacterial artificial chromosome (BAC) libraries, physical maps and an integrated genetic and physical map. A preliminary genetic linkage map was published for *M. sinensis* (Atienza *et al*. 2002) using RAPD markers. However, the map was incomplete and generated more linkage groups (LGs) than the expected 19 LGs (haploid number of chromosomes in *M. sinensis*), and map resolution was low. In addition, the markers are not very informative since the reproducibility of RAPD marker data is often questionable. Complete linkage maps with higher map resolution in *Miscanthus* species are required to understand genetic composition, conduct QTL analysis, facilitate marker associated breeding programs and assemble sequencing data in

order to sequence the genomes. This would be achieved by using SSR and SNP markers, which are plentiful in the genome and informative.

1.4. Whole genome duplication in Saccharinae

Whole genome duplication or polyploidy is common in all eukaryotes and particularly prominent in plants. The grass subtribe Saccharinae genera (*Erianthus*, *Miscanthus*, *Narenga*, *Saccharum* and *Sclerostachya*) also contain various levels of polyploids that arose recently and independently (Daniels & Roach, 1987). *Sorghum bicolor* and sugarcane carry even number of the basic chromosome sets (x = 10 and x = 8 or 10, respectively) and the latter display polysomic inheritance that indicates autopolyploidy (Brandes E, 1956; Sreenivasan *et al*., 1987; D'Hont *et al*., 1998). A recent and independent whole genome duplication which arose from a common diploid progenitor in Saccharinae can be assumed from the similar monoploid genome size between sugarcane (approximately 843 million base pairs (Mbp) for *Saccharum spontaneum* and 985 Mbp for *S. officinarum*, Zhang *et al*., 2012) and eusorghum (745−818 Mbp, Price *et al*, 2005). The basic chromosome number of the common ancestor of Saccharinae was likely the same as diploid *S. bicolor*, x=10.

Chromosome counting and bivalent pairing of chromosomes during meiosis indicate that *M. sinensis* is 'diploid' with 2n = 2x = 38 chromosomes (Burner, 1991; Lafferty & Lelley, 1994). Therefore, the genus *Miscanthus* has a basic chromosome number of x = 19. This is possibly the result of the whole genome duplication of an ancestor possessing x = 10 pairs of chromosomes (Paterson *et al*., 2010) although this has not been demonstrated. Direct comparisons of the monoploid genome size of *Miscanthus* to sorghum and sugarcane implicate that the genome of *Miscanthus* has duplicated or triplicated, as the x = 19 monoploid genome

size of *Miscanthus* is 2150−2650 Mbp (Rayburn *et al*., 2009), more than three times larger than that of eusorghum (745−818 Mbp, Price *et al*, 2005). This is possibly due to the content of highly repetitive DNA in the *Miscanthus* genome, recently shown by sample sequencing to be ~95% in *M. x giganteus* (Swaminathan *et al*., 2010). How the odd basic chromosome number in *Miscanthus* evolved is still unknown although additional chromosome-scale events such as loss or fusion can be hypothesized.

1.5. Ploidy manipulation

Polyploidy occurs repeatedly in flowering plants and serves as a barrier for gene flow from diploid parents to polyploidy progenies (Leitch *et al*., 2005) and, therefore, is commonly associated with speciation. Polyploidization increases cell volume via increasing genome size which can result in phenological changes such as delayed development and flowering of polyploidy plants (Ramsey & Schemske, 2002). The genomes of neopolyploid plants usually undergo extensive change as they are usually unstable and experience rapid repatterning (Wendel, 2000). These rapid and extensive genome rearrangements are reproducible and usually occur in early generations after polyploidization (Levy & Feldman, 2004). Studies using synthetic polyploids by either anti-mitotic agent treatment or spontaneous chromosome doubling have revealed that extensive and rapid genomic rearrangements are probably due to sequence rearrangements, homoeologous recombination, and sequence elimination (Adams & Wendel, 2005; Otto, 2007). These cytogenetic and genetic changes can generate phenotypic variation among neopolyploids. Unique phenotypic variation associated with polyploidization can be a target for natural selection (Otto, 2007) and one of major driving forces in plant speciation creating an instant reproductive barrier with their progenitors (Rieseberg & John, 2007).

Artificial genome doubling can be the method for introducing genetic variation into germplasm collections, especially when the creation of new accessions from conventional hybridization is difficult. *M. x giganteus* is sterile due to its triploid nature. There are few *M. x giganteus* accessions in Europe and the United States with limited genetic diversity among these genotypes (Greef *et al*., 1997). Neopolyploid plants generally exhibit sturdier foliage, thicker stems, and enlarged reproductive structures (Ramsey & Schemske, 2002). In colchicine-induced polyploid *Miscanthus* species, an increase in biomass was observed, via increased stem diameter and tuft weight (Glowacka *et al*., 2010).

The other potential strategies for *Miscanthus* crop improvement can be re-synthesis of new triploid *M. x giganteus* genotypes by conventional hybridization (Sacks *et al*., 2012). The production of triploid plants also circumvents invasive issues as observed in the two parental fertile species (Quinn et al., 2010). However, few tetraploid accessions of *M. sacchariflorus* have been reported in the United States so far, which is assumed to be one of the parents of *M. x giganteus*. Therefore, producing various tetraploid *M. sacchariflorus* plants is necessary for re-synthesizing triploid *M. x giganteus* germplasm in the United States. Genome doubling techniques by colchicine or oryzalin treatment via in vitro tissue culture have been established for *M. sinensis* and *M. x giganteus* (Peterson *et al*., 2002 and 2003; Yu *et al*., 2009; Głowacka *et al*., 2010) but not for *M. sacchariflorus*.

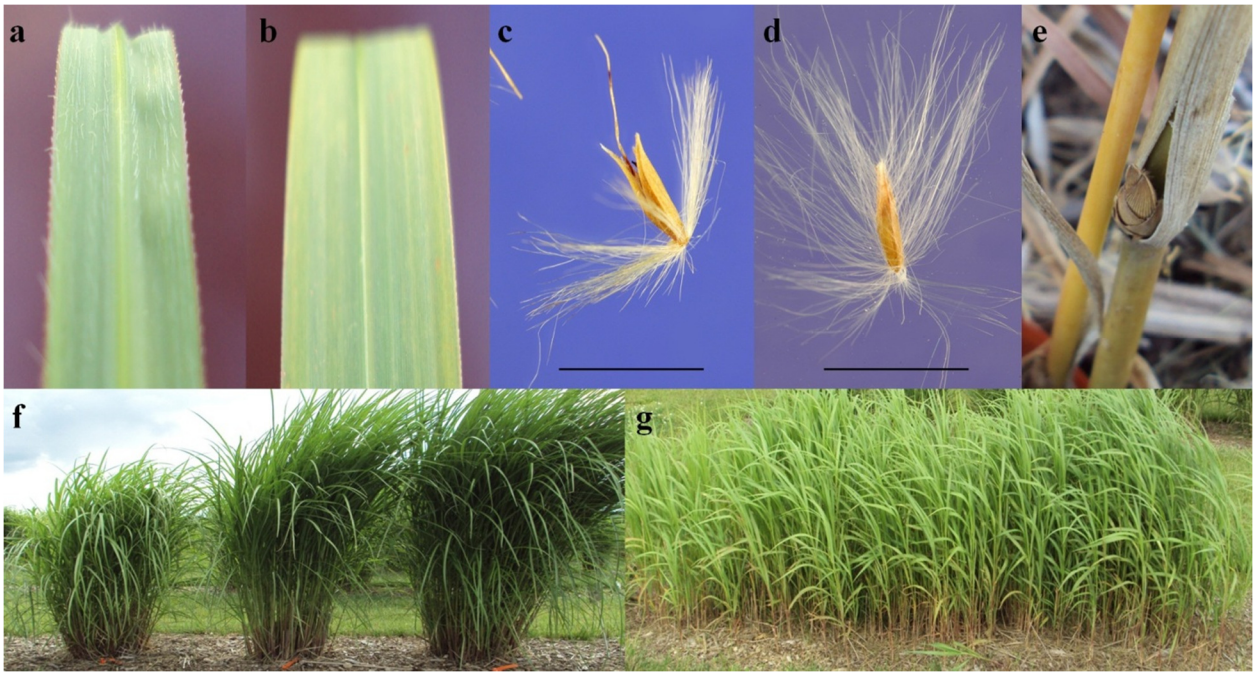

Figure 1.1. Key traits for morphology-based taxonomy in *Miscanthus* genus. a) A pilose abaxial leaf of *M. sinensis*, b) a glabrous abaxial leaf of *M. sacchariflorus*, c) awns in spikelets of *M. sinensis*, d) awnless spikelets of *M. sacchariflorus* and e) a bud at the node of *M. sacchariflorus*, f) tufted (clumped) growth habit of *M. sinensis* and g) rhizomatous (spreading) growth habits of *M. sacchariflorus*. *Miscanthus* accessions were panted at 0.9 m spacing and have grown for 5 years. Bars represent 5mm. Pictures were taken from May to October 2010 in the Energy Bioscience Institute Farm at UIUC. Images of c and d were provided courtesy of Justin M. Gifford.

CHAPTER 2

Taxonomy of the genus *Miscanthus* based on morphology, genome size and SSR markers

2.1. Abstract

While taxonomic studies have been conducted in the genus *Miscanthus* based on morphology and limited DNA fingerprinting, the number of species and their intra- and interspecific relationship is unclear. In this study, we provide important information on genome size and ploidy levels for classifying *Miscanthus* accessions and, thus, to assist potential biomass crop improvement programs and to understand the evolution of the genus *Miscanthus*. Combined analyses using plant morphology, genome size and molecular marker systems were applied to characterize *Miscanthus* accessions. Morphology of 101 horticultural *Miscanthus* accessions was investigated for five phenotypic traits including presence of elongated rhizomes, awns in spikelets, buds at the nodes, rhizomatous (spreading) or tufted (clumped) growth habit, and pilose or glabrous abaxial leaf surfaces. Genome size of *Miscanthus* accessions and related genera was estimated based on chromosome numbers and total nuclear DNA content. *Miscanthus* accessions divided into 4 groups, including section *Miscanthus*, section *Triarrhena* and a third and fourth group intermediate between the two sections, based on morphological traits and monoploid genome size. Simple sequence repeats (SSRs) were generated and applied to 31 *Miscanthus*, 4 *Saccharum*, 6 *Erianthus* and 1 *Sorghum bicolor* accessions (outgroup) to conduct cluster analysis and Principal coordinate analysis (PCoA) to determine genetic relationships. Sixteen SSR primer pairs were selected based on amplification and polymorphism across genera. Morphology, genome size and SSR genotyping of 42 accessions including diploid and triploid artificial interspecific *Miscanthus* hybrids, suggested that there are 3 species (*M.*

sinensis, *M. sacchariflorus* and *M. x giganteus*), one variety (tetraploid *M. sacchariflorus* and *M.sacchariflorus var. lutarioriparius*) and a putative hybrid, *Miscanthus* 'Purpurascens' among our *Miscanthus* accessions, which were clearly separated from the genera *Erianthus* and *Saccharum*. The evolution of *Miscanthus* and related genera is discussed based on genome size, ploidy level, cluster analysis and geographical distribution. Mechanisms changing genome size in *Miscanthus* and its phenotypic consequences influencing taxonomic and geographical distribution are also discussed.

2.2. Introduction

Tribe Andropogoneae (Poaceae) contains many important C_4 grasses such as *Zea mays* L. (maize), *Sorghum bicolor* (L.) Moench (sorghum) and *Saccharum officinarum* L. (sugarcane). The subtribe, Saccharinae is defined as possessing bisexual paired spikelets and a distribution in tropical and subtropical regions as features that the subtribe commonly shares (synapomorphy) (Amalraj & N. Blalsundaram, 2006; Clayton & Renvoize, 1986). The term "*Saccharum* complex" has been used to describe a subset of the Saccharinae (*Erianthus*, *Miscanthus*, *Narenga*, *Saccharum* and *Sclerostachya*) characterized by both high levels of polyploidy and aneuploidy and with frequent hybridizations among these genera, resulting in a lack of morphological traits distinct for characterizing the taxonomy of this subtribe (Daniels & Roach, 1987).

The genus *Miscanthus* is native to a relatively broad geographic distribution ranging from eastern Asia (Korea, China, Japan and neighboring regions) to the Pacific islands; from about 50° North in southern Siberia to 22° South (Hodkinson et al. 1997; 2002a). Centers of *Miscanthus* genetic diversity are in the northern temperate latitudes, where their cold tolerance is a key distinctive feature of *Miscanthus* relative to other members of the Saccharinae (Sacks *et al.*,

2012). Some *Miscanthus* genotypes maintain high photosynthetic rates at temperatures below 12 °C, in contrast to sugarcane, maize, and sorghum, which shows greatly reduced CO_2 assimilation at temperatures below 14 °C (Beale *et al*. 1996; Naidu *et al*. 2003; Wang *et al*. 2008). The adaptation to temperate climates makes the genus especially attractive for development of a perennial biomass crop adapted to North America and Europe and combining both high yield potentials with low nutrient input requirements (Clifton-Brown *et al*. 2004; Heaton *et al*. 2004; 2008) and perenniality.

Miscanthus accessions in the United States have been introduced as ornamental plants by private nurseries and research institutes since the late 1800's (Quinn et *al*., 2010). These accessions may have been selected for ornamental purposes and, thus, do not represent a comprehensive collection of the genus *Miscanthus*, and show limited genetic and phenotypic variation compared to populations at centers of diversity in Asia. Taxonomic information on these accessions is limited since comprehensive and systematic studies on the taxonomy of U.S. accessions have not been conducted. Currently at the University of Illinois at Urbana-Champaign (UIUC), more than 100 accessions of *Miscanthus* have been collected (mostly ornamentals) from various nurseries and from the National Plant Germplasm System of the USDA (USDA-NPGS). Errors in identification both at the species level and within each species are apparent, possibly due to the lack of proper taxonomic knowledge, absence of informative molecular markers and/or mislabeling of plants in nurseries. Also, the species names were assigned by private nurseries, except for the *Miscanthus* accessions from USDA-NPGS.

The genus *Miscanthus* is closely related to *S. officinarum* and its probable wild progenitor, *S. robustum* (Amalraj & Balasundaram 2006; Hodkinson *et al*. 2002a). The genus is divided into two sections, *Miscanthus* and *Triarrhena* and a subsection *Kariyasua* (Lee, 1993;

Hodkinson 2002a). Approximately 20 species were first reported in the genus *Miscanthus* (Clayton & Renvoize, 1986) but recent studies suggest 11-12 species (Hodkinson *et al*., 2002b; Clifton-Brown *et al.*2008). Sun *et al*. (2010) reported that the genus *Miscanthus* comprises six species, two sub-species and four varieties in China, one of the centers of diversity of *Miscanthus*. *Miscanthus* species from Himalayas and from Southern Africa should be classified differently from the genus *Miscanthus*, as they are more closely related to *Saccharum* and *Sorghum*, respectively (Amalraj & Balasundaram 2006; Hodkinson *et al*., 1997; Shouliang & Renvoize, 2006). They also have different basic chromosome numbers as the former has x=5 or 10 (Mehra & Sharma, 1975) and the latter has x=15 (Hodkinson *et al*., 2002a), in contrast to other *Miscanthus* species where x=19 (Adati & Shiotani, 1962).

Determining species, synthetic or natural hybrids and ploidy levels of accessions is critical for a breeding program to develop new *Miscanthus* cultivars. The determination of plant species has been a longstanding debate (Soltis & Soltis, 2009) and has usually been determined by investigating morphological or phenotypic variation among populations (Grant, 1981), plant reproductive isolation (Mayr, 1942), and lineage divergence or phylogeny (Boggs, 2001). Determining hybrids or hybrid species requires the detection of merging divergent genes and genomes, which is somewhat different from species determination that focuses on how lineages diverge (Rieseberg & Willis, 2007).

Taxonomic study of *Miscanthus* using morphology began in 1885, first described by Andersson (Scally *et al*., 2001). Morphological traits of inflorescence have been widely used for delimiting sections and species of the genus *Miscanthus*, such as inflorescence axis, the length of the racemes, the disposition of the spikelets on the axis, nerves of glume, dorsal hairs of glume and the presence of awns (Lee 1964 a, b and c; Clayton & Renvoize, 1986; Hodkinson *et al*.,

1997). The infloresecence-derived morphological taxonomy is, however, difficult to apply as observed with many grasses (Scally *et al*., 2001). Recently, Sun *et al*. (2010) classified *Miscanthus* germplasm not only by floral but also by vegetative traits which were both qualitative and quantitative in nature. The inclusion of quantitative traits to distinguish members of the genus *Miscanthus* is problematic (Hodkinson *et al*., 2002b). Some key qualitative traits for delimiting the two sections *Miscanthus* and *Triarrhena* have been used in *Miscanthus* germplasm including the presence of elongated rhizomes, awns in spikelets, buds at the nodes, rhizomatous (spreading) or tufted (clumped) growth habit, and pilose or glabrous abaxial leaf surfaces (Lee, 1993) (Fig. 2.1).

DNA markers have also been applied to assess genetic diversity of the genus *Miscanthus*. Amplified fragment length polymorphism (AFLP) fingerprinting was used to assess variation in three *Miscanthus* species (Greef *et al*., 1997). Later, the use of AFLPs and inter-simple sequence repeats (ISSR) and DNA sequencing of the internal transcribed spacer (ITS) region of ribosomal DNA and plastid trnL-F region were proven to be suitable for assessing genetic diversity in European *Miscanthus* accessions (Hodkinson *et al*., 2002 a, b and c). These studies suggested that there are six distinct taxonomic groups which include *M. sinensis*, *M. x giganteus*, *M. sacchariflorus*, *M. oligostachyus*, *M. transmorrisonensis* and *M. nepalensis* in *Miscanthus* germplasm of Royal Botanic Gardens Kew, UK and ADAS Arthur Rickwood Research Station (Hodkinson *et al*., 2002b).

Comparative genomics has been conducted among grass species (Gale & Devos 1998; Paterson *et al*., 2010) revealing significant sequence conservation across taxa. Simple sequence repeats (SSRs) have been widely used for germplasm characterization and can be diagnostic for different species and even different genera. In Saccharinae, early studies revealed the utility of

SSR markers from sugarcane Expressed Sequence Taqs (EST) to generate informative amplicons in *Erianthus* and sorghum (Cordeiro *et al*., 2001). Recently, SSR markers were generated from sugarcane EST and genomic sequences (James *et al*., 2011) and applied to create a *M. sinensis* genetic map and for comparative mapping with sorghum (Swamminathan *et al*., 2012). This suggests that SSR markers showing sequence homology with sorghum can be used for assessing taxonomic relationships among various genera of Saccharinae.

Nuclear DNA C-value (Dolezel *et al*., 2007) provides information that is a useful indicator for taxonomic and evolutionary studies. It is very efficient in both defining infrageneric division in a number of taxa (Ohri, 1998) and determining how genomes grow or shrink during evolution (Bennetzen *et al*., 2005; Leitch *et al*., 2005). It is also useful for detecting polyploids in core germplasm. Flow cytometry is one of the most efficient methods for verifying the DNA C-value of organisms, ploidy levels (De Laat *et al*., 1987) and identifying synthetic hybrids or hybrid species (Bennett & Litch, 1995; Dolezel, 1997; Rayburn *et al*., 2005). It can be also applied to studies in population biology, for crop breeding and for quality control in commercial seed production (Dolezel *et al*., 2007).

Total nDNA content, chromosome number and molecular marker data are complementary for taxonomic and evolutionary studies since the former two provide information on genome size and ploidy levels and the latter facilitate fine classification and genetic relationships among accessions. Furthermore, the detection of hybrid species or accessions can be improved by including genome size information in addition to molecular marker data since hybrids usually show mid-parent values of genome size (Bennett and Litch, 1995; Dolezel, 1997; Rayburn *et al*., 2005). Cluster analysis combined with genome size and geographic distribution data can be also useful to understand evolution in *Miscanthus* and related genera.

Accurate classification and characterization of accessions is a prerequisite for a successful breeding program and for evolutionary studies among *Miscanthus* species. Therefore, the five key morphological traits were used for initial classification of *Miscanthus* accessions. Chromosome numbers in a subset of *Miscanthus* accessions and related genera were counted or obtained from published literature since the primary feature distinguishing *Miscanthus* from other related genera is the basic chromosome number of x=19 (Adati & Shiotani, 1962). Total nDNA content of all accessions was measured by flow cytometry to estimate genome size. A subset of accessions from 31 *Miscanthus*, four *Saccharum* and six *Erianthus* accessions and one *Sorghum bicolor* (Table 2.1 and 2.2) were genotyped with SSR markers that are polymorphic between and within genus and species. Then, the genome size of *Miscanthus* and related genera was interpreted based on UPGMA (Unweighted pair-group method using arithmatic means) cluster analysis. The mechanisms to change genome size and its phenotypic consequences influencing taxonomic and geographical distribution are discussed.

2.3. Materials and Methods

Plant materials and phenotypic data collection

Morphological traits of one hundred one horticultural *Miscanthus* accessions (Supplementary Table 2.1) were investigated. These accessions were collected from various nurseries in the U.S. and clonal propagules transplanted into plots on the Energy Bioscience Institute (EBI) farm, SoyFACE plots and Plant Science Laboratory (PSL) greenhouse at UIUC. Five morphological traits were recorded for accessions including presence of elongated rhizomes, awns in spikelets, buds at the nodes, rhizomatous (spreading) or tufted (clumped)

growth habit, and pilose or glabrous abaxial leaf surfaces to assign to species (Figure 2.1) (Lee, 1993).

Chromosome counts of root tips

Chromosome numbers were counted in mitotic cells of 12 genotypes, which represent each species, variety and ploidy level (Table 2.1). Root tips 1–2 cm in length were excised and soaked in 0.05% 8-hydroxyquinoline for mitotic inhibition. After three hours, the root tips were rinsed in ddH_2O for 5 min and stored in 3:1 (v/v) 100% ethanol/acetic acid. The roots were stored at room temperature for four days and then stored at 4°C until use. Fixed root tips were rinsed in ddH_2O, hydrolyzed in 5 N HCl for 45 min and placed in Feulgen's stain for 2 h. Root tips were then rinsed in ddH_2O and a drop of 1% acetocarmine was added to the root tip. A cover slip was placed over the tissue and gently tapped with a dissecting needle to disperse the tissue. The slide was then flamed over an alcohol burner, and direct pressure was applied to the slide. The slides were then viewed using an Olympus BX61 microscope (Olympus America Inc., Melville, NY, USA). Photographs of chromosome spreads were taken using an Olympus U-CMAD3 camera and chromosome counts conducted on the clearest preparations. The number of chromosomes was determined based on at least three root tip squashes per accession.

Genome size estimation using flow cytometry

Nuclear DNA contents of 37 *Miscanthus*, eight *Saccharum* and six *Erianthus* accessions (Table 2.1) were measured by flow cytometry using a protocol modified from Rayburn *et al*. (2009). To determine the nDNA content and genome sizes of interspecific hybrid progeny, crosses were made between diploid *M. sinensis* and diploid *M. sacchariflorus* and

between diploid *M. sinensis* and tetraploid *M. sacchariflorus* to produce diploid and triploid interspecific hybrid progenies, respectively (Table 2.2). Five diploid and three triploid interspecific hybrid progeny plants were subjected to flow cytometric analysis. Young leaf tissue (1 cm^2) from newly emerging shoots of *Miscanthus* or related genera were co-chopped with sorghum '*S. bicolor* cv. Pioneer 8695' or maize 'W-22', as the internal standard (Table 2.1), in a petri-dish containing 10 ml extraction buffer consisting of 13% (v/v) hexylene glycol, 10 mM Tris–HCl (pH 8.0), and 10 mM $MgCl_2$ with 200 µl of 25% Triton X. The samples were filtered through a 50-µm nylon mesh (Partec GmbH, Gorlitz, Germany) into a labeled test tube and kept on ice throughout. Following filtration, samples were centrifuged for 25 min at 300 xg at 4°C. The supernatant was then aspirated, and nuclei are resuspended in 300 µl of propidium iodide (PI, Sigma-Aldrich, St Louis, MO, USA). Then the solution was transferred to a 1.5 mL micro-centrifuge tube and incubated for 20 min at 37°C. After incubation, 300 µL of PI salt was added to each sample. Samples are then briefly vortexed, placed on ice, and stored at 4°C for at least 1 h. Nuclei were analyzed using a flow cytometer Model LSRII (BD Biosciences, Flow Cytometry Facility at the University of Illinois-Keck Biotechnology Center). The excitation wavelength was set at 488 nm and a 570 nm emission filter was used. A minimum of 20,000 nuclei per sample were analyzed. Mean fluorescence of the *Miscanthus* G1 peak is divided by the fluorescence reading of the G0/G1 peak of sorghum or maize, multiplied by 1.74 or 5.35 pg/2C (McMurphy & Rayburn, 1991), respectively, and expressed in pg/2C nucleus. Samples with sorghum or maize G2/G1 peak ratios outside the range of 1.93–2.03 was excluded since the G2/G1 ratio of normal diploid cells is slightly less than 2.0 due to S phase cells contaminating the G0/G1 and the G2/M peaks (http://www.cyto.purdue.edu/cdroms/cyto3/15/data/dna.htm). A target G0/G1 peak in a DNA

histogram that was not symmetrical and exceeded the coefficient of variation (CV, a normalized standard deviation defined as CV = 100 × standard deviation / mean of peak) by more than 5% were also excluded. For each accession, three-five samples were examined and the samples that did not meet the criteria above were discarded until three acceptable replications (one leaf per replication) were recorded. The monoploid genome size (in megabase pairs, Mbp) was calculated according to the formulae by Lysak and Dolezel (1998) with conversion of 1 pg equal to 980Mbp (Dolezel *et al*. 2003).

Amplification of SSR markers

Genomic DNA of 42 accessions (Table 2.1 and 2.2) was extracted from young leaves of each accession using the Puregene protocol (Qiagen, Valencia, California, USA). SSR primers from a *M. sinensis* genetic map (Swamminathan *et al*. 2012) developed from EST and genomic sequences of *S. officinarum* (James *et al*. 2011) were first screened by investigating sequence homology with sorghum (Swamminathan *et al*. 2012). Selected 38 SSR markers having sequence homology with sorghum were then used to screen to 18 accessions including *S. bicolor*, *S. officinarum* 'LA purple', *Erianthus arundinaceus*, *M. sacchariflorus var. lutarioriparius*, tetraploid and diploid *M. sacchariflorus*, *M. x giganteus*, *M. floridulus*, *Miscanthus* 'Purpuracens', *M. sinensis*, *M. transmorrisonensis*, and *M. tinctorius*. Products were amplified in 10 µl PCR reactions containing 1 µl of genomic DNA (5–10 ng), 0.1 µl of forward and reverse primers (100 µM stock each), 3.8 µl of ddH2O and 5 µl of 2X GoTaq Green Master Mix (Promega, Madison, Wisconsin, USA). Touchdown PCR was used to amplify the SSRs: denaturation at 94°C for 3 min followed by 2 cycles of 94°C for 30 sec, 65°C for 30 sec, and 72°C for 45 sec. The annealing temperature was decreased every 2 cycles by 2°C until 57°C.

The amplification was finished with 26 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 45 sec (total 36 cycles) and a final extension at 72°C for 10 min. The amplicons were separated on 4% agarose SFR gels (Amresco, Solon, Ohio, USA) with 1x TBE buffer at 4°C and visualized with ethidium bromide.

Sixteen SSR primers producing amplicons in all three genera, except for two primers that did not produce amplicon in *Erianthus*) were selected with chromosomal positions in unique linkage groups of *S. bicolor* and *M. sinensis*. Sequence information of the SSRs (Swamminathan *et al*. 2012) is listed in Table 2.3. A total of 258 fragments were amplified among the 42 accessions from the 16 SSR primers. To genotype accessions, products were amplified in 10 µl PCR reactions containing 1 µl of genomic DNA (5–10 ng), 0.02 µl of M13 tailed forward primer, 0.1 µl of each reverse and fluorescent M13 primers (100 µM stock), 3.78 µl of ddH2O and 5 µl of 2X GoTaq Colorless Master Mix (Promega, Madison, Wisconsin, USA). TwoM13 primers tagged with FAM and VIC at the 5′ end were used. Touchdown PCR was same as described above. Size separation of the amplicons was carried out by the Keck Center for Functional Genomics at UIUC, on an ABI 3730xl with the LIZ500 size standard. Marker scoring was done using the Genemarker software (Softgenetics, LLC State College, Pennsylvania, USA).

Cluster analysis using SSR markers

A binary matrix generated from coding 1 or 0 depending on presence or absence of each relevant SSR allele, respectively, was used for calculating the distance matrix using Jaccard's similarity coefficient (Jaccard, 1908). Jaccard's similarity coefficient was used since the binary information was asymmetric (the shared absence of an allele does not represent genetic similarity). The dendrogram was generated with Unweighted pair-group method using arithmetic

average (UPGMA) method using software package NTSYS-pc version 2.21m (Rohlf, 2002). The dendrogram was rooted using *S. bicolor* as the outgroup. Principal coordinate analysis (PCoA, = Multidimensional scaling, MDS) based on Jaccard's similarity matrices were used to visualize genetic similarities and relationship among different accessions using NTSYS-pc version 2.21m (Rohlf, 2002).

2.4. Results

Morphological traits of *Miscanthus* accessions

The morphological traits of most *Miscanthus* accessions conformed to their expected taxonomy. All *M. sacchariflorus* and *M. sacchariflorus var. lutarioriparius* accessions displayed elongated (monopodial) rhizomes, glabrous leaves, awnless spikelets, buds at the nodes and rhizomatous growth habit (Figure 2.1, Supplementary table 2.1). In contrast, all horticultural *M. sinensis* accessions have small (sympodial) rhizomes, pilose leaves, awned spikelets, tufted growth habit with no buds at the nodes (Figure 2.1, Supplementary table 2.1). Some of *M. sinensis* and *M. floridulus* accessions from USDA showed glabrous abaxial leaves and/or awnless spikelets. *Miscanthus* 'purpurascens' has morphological traits intermediate between *M. sinensis* and *M. sacchariflorus*, having glabrous abaxial leaves (Figure 2.1b), awns in spikelets (Figure 2.1e) and elongated rhizomes (Figure 2.1j) but no bud at the nodes (Supplementary table 2.1). *M. x giganteus* accessions could not be distinguished from *M. sacchariflorus* (Figure 2.1g; Supplementray table 2.1) but their growth habits are intermediate between *M. sinensis* and *M. sacchariflorus* (Figure 2.1o). Ten accessions were found to be misclassified of the 101 horticultural accessions (Supplementary Table 2.1).

Chromosome counts of *Miscanthus* species

All *Miscanthus* accessions investigated had a basic chromosome number of x=19 where most accessions was diploids (2n=2x=38), except for a triploid accession 'Goliath' (2n=57) (Figure 2.2b) and a tetraploid accession 'KB giganteus' (2n=76) (Table 2.1 and Figure 2.2d). The latter accession was listed by the nursery source as *M. x giganteus* but a chromosome number of 76 and morphology suggest that it is a tetraploid *M. sacchariflorus*. The chromosome number of 'NG 51-088' was 30 (Figure 2.1f) and those of *Erianthus* accessions were 40 and 60 for 'MPTH98-283' (Figure 2.1g) and 'MPTH97-326' (Figure 2.1h), respectively. The chromosome numbers of USDA accessions were obtained from the literature (Table 2.1). Price and Daniels (1968) observed fragmented chromsomes in some of 'Fiji' accessions. For example, 'Fiji 59' has 96 chromosomes and 2 fragments and, thus, is listed as having 96+2f chromosomes (Table 2.1).

Nuclear DNA contents and ploidy levels of *Miscanthus* and related genera

Total nDNA contents and monoploid genome sizes (Cx-value) of some of horticultural *Miscanthus* and related genera from USDA-NPGS are shown in Table 2.1. Total nDNA content of *Miscanthus* accessions ranged from 4.26 pg for *M. sacchariflorus var. lutarioriparius* 'PF30022' (Figure 2.3b) to 8.65 pg for 'KB giganteus' (Figure 2.3i), which was a tetraploid accession (Figure 2.2d). Diploid *M. sinensis* accessions ranged between 5.00 and 5.42 pg and all the of horticultural *M. sinensis* accessions fell within this range (data not shown). The highest total nDNA content in diploids were observed in *M. tinctorius* which ranged between 5.49 and 5.66 pg. 'Goliath' was classified as triploid *M. sinensis* since its morphology was the same as other horticultural *M. sinensis* accessions and its total nDNA coentent (7.57 ± 0.18 pg; Figure

2.3f) was similar to the expected value for triploid *M. sinensis* (mean Cx-value of *M. sinensis* 2.6 pg × 3 = 7.8 pg) and displayed 57 chromsomes. Similarly, 'KB giganteus' and 'Gotemba gold' (Figure 2.3g) were classified as tetraploid *M. sacchariflorus* based on the expected value for tetraploid *M. sacchariflorus* (2.224 × 4 = 8.90 pg).

Total nDNA content of *Saccharum* and *Erianthus* accessions are also listed in Table 2.1. *S. officinarum* 'LA purple' had 7.75 pg of total nDNA content (Figure 2.4d) and that of *Saccharum* hybrid 'Fiji' accessions ranged between 9.98 and 10.57 pg (Table 2.1). *Miscanthus* accession 'Raiatea' was similar to 'Fiji' accessions in terms of total nDNA content, showing 10.03 pg of nDNA (Table 2.1; Figure 2.4e). The 'NG51-088' accession showed similar genome size with *Erianthus* (see below) and 3.66 pg of nDNA (Figure 2.4a). *Erianthus* accessions had 4.75 – 4.90 pg of nDNA (Table 2.1; Figure 2.4b), except for 'MPTH97-260' having 7.27 pg of nDNA (Figure 2.4c). Based on the basic chromosome number of x = 10 for *Erianthus* (D'Hont *et al*., 1995), 'NG51-88' and 'MPTH97-260' were classified as triploid and hexaploid and other MPTH accessions as tetraploids (Table 2.1).

Classification of *Miscanthus* accessions and related genera using morphology and monoploid genome size

Accessions were grouped based on their estimated Cx-values and morphology. Cx-values (pg) were calculated by dividing nDNA content by their estimated ploidy level using the basic chromosome number of 10 for *Saccharum* and *Erianthus* (D'Hont *et al*., 1995) and 19 for *Miscanthus* accessions (Adati & Shiotani, 1962). For determining ploidy levels of accessions where chromosome counts were not conducted, chromosome numbers were estimated by total nDNA contents. For example, ploidy level of *M x giganteus* 'Gilded tower' was projected as

triploid since its morphology and nDNA content were similar to those of *M. x giganteus* 'Illinois', which contains 57 choromosomes. The basic chromosome numbers of 'Fiji' were estimated as 10 since their chromosomes ranged between 95 and 100 if fragmented chromsomes were included.

All accessions could be classified into six groups across taxa based on combined analysis of their morphology and genome sizes (Table 2.1). The first group was the genus *Saccharum* and thieir hybrids, ranging from 0.97 to 1.06 pg of Cx-value. The second group was *Erianthus*, displaying Cx-values from 1.19 to 1.23 pg. The 'NG51-088' accession was included in the *Erianthus* group based on the similarity of monoploid genome size with other *Erianthus* accessions (Table 2.1). *M. sacchariflorus* and *M. sacchariflorus var. lutarioriparius* represented the third group and the former had a slightly greater genome size (2.20 – 2.25 Cx/pg) than the latter (2.10 – 2.16 Cx/pg) (Table 2.1). This group also contained 'KB giganteus' (2.16 Cx/pg) and 'Gotemba gold' (2.10 Cx/pg) with genome sizes close to *M. sacchariflorus var. lutarioriparius* (2.13 Cx/pg). All accessions in this group shared similar morphological characteristics and thus 'KB giganteus' and 'Gotemba gold' were classified as *M. sacchariflorus*. *M. x giganteus* accessions with monoploid genome sizes ranging between 2.29 and 2.33 Cx/pg represented the fourth group. The fifth group, *Miscanthus* 'Purpurascens' accessions showed Cx-values ranging from 2.44 to 2.46 pg with morphological traits intermediate between *M. sacchariflorus* and *M. sinensis* groups. The last and largest group consisted of accessions in the section *Miscanthus* with genome sizes ranging from 2.45 Cx/pg in 'M75-062' to 2.83 Cx/pg in *M. tinctorius* 'Gifu 2010-12-15'. Although there was large variation in genome sizes in this group, they shared similar morphological characteristics (Supplementray table 2.1) and, thus, were grouped together.

nDNA contents of interspecific hybrids of *Miscanthus* species

Total nDNA content of the artificial hybrids were close to the mid-parent mean of their parent genotypes (table 2.2; Figure 2.3c, d, e, h, I and j). Their morphological traits were also intermediate between two parental accessions (data not shown). This suggests that *Miscanthus* 'Purpurascens' and *M. x giganteus* accessions may be hybrids generated from crosses between section *Miscanthus* and diploid *M. sacchariflorus* and between section *Miscanthus* by tetraploid *M. sacchariflorus*, respectively.

Amplification of SSR markers

Sixteen SSR primers were selected (Table 2.3) and half of them were amplified in *S. bicolor* (Table 2.4), which is predictable since only one cultivar of sorghum (*S. bicolor* cv. Pioneer 8695) was used as a standard. Fourteen of the 16 primers were polymorphic between and within genera and produced amplicons in accessions of *Saccharum*, *Erianthus* and *Miscanthus*. The remaining two primers displayed polymorphism between *Saccharum* and *Miscanthus* but did not generate amplicons in *Erianthus* (Table 2.4).

Cluster analysis of *Micanthus* and related genera using SSR markers

Thirty-eight accessions including *Erianthus*, *Saccharum*, *Miscanthus* and *S. bicolor* accessions (Table 2.1) were used in the cluster analysis. The UPGMA dendrogram revealed three major clusters, which separated all *Erianthus*, *Saccharum* and *Miscanthus* accessions (Figure 2.5). *Erianthus* accessions separated out as a distinct cluster from *Miscanhus* and *Saccharum*. As expected from their genome size, 'NG51-088' and *Miscanthus* 'Raiatea' were

included in *Erianthus* and *Saccharum* clusters, respectively (Figure 2.5). The *Miscanthus* accessions were largely divided into two groups, similar to the sections *Miscanthus* and *Triarrhena*. In the section *Miscanthus* cluster, three subgroups were observed including most of the USDA-NPGS *Miscanthus* accessions (except for *Miscanthus* 'Raiatea'), whose species status was not identified by the USDA (Figure 2.5). All subgroups contained *M. sinensis* as designated by USDA-NPGS based on their collector records. (Figure 2.5). The two *M. floridulus* accessions belonged in two subgroups with each subgroup containing *M. sinensis* from USDA-NPGS. Accessions in the *Triarrhena* group were subgrouped according to their species names and ploidy levels. Two *M. x giganteus* accessions were separated as 'Gilded tower' were grouped with tetraploid *M. sacchariflorus* and 'Illinois' with *M. sinensis* and *M. floridulus* (Figure 2.5). Two accessions of *M. tinctorius* fell into two different subgroups but each subgroup also included *M. sinensis* from USDA-NPGS.

The same 38 accessions (Table 2.1) as in cluster analysis were used in the PCoA using Jaccard's similarity coefficient (Figure 2.6a). The PCoA plot suggested there are four distinct groups in the genus *Miscanthus*, which are clearly separated from *Saccharum* and *Erianthus*. *M. tinctorius*, *M. transmorrisonensis*, *M. sinensis*, *M. floridulus* and other *Miscanthus* species from USDA-NPGS all grouped together. In contrast, diploid and tetraploid *M. sacchariflorus* were grouped differently (Figure 2.6a). In PCoA plot (Figure 2.6b) using a total of 30 *Miscanthus* accessions of three diploid and one triploid hybrid progeny (Table 2.2) and all 26 *Miscanthus* accessions except for 'Raiatea', section *Miscanthus* were further divided into three subgroups. Each group also contained *M. sinensis* accessions from USDA-NPGS, probably representing variation within the species. In the PCoA plot (Figure 2.6b), *M. x giganteus* 'Illinois' was not

included in any group while the other *M. x giganteus* accession, 'Gilded tower' were still grouped together with tetraploid *M. sacchariflorus*.

Putative hybrid detection in *Miscanthus* accessions using known hybrids

Hybrids from a controlled pollinated between diploid *M. sacchariflorus* 'Golf course' and diploid *M. sinensis* 'Grosse Fontaine' were between two parental accessions on PCoA plots (Figure 2.6b) A triploid hybrid between tetraploid *M. sacchariflorus* 'KB giganteus' and diploid *M. sinensis var. condensatus* 'Cabaret' was also positioned between the two parents although the progeny genotype are closer to the tetraploid *M. sacchariflorus* parent (Figure 2.6b; red dots connected by red lines). Mean Jaccard's similarity coefficients between the three diploid hybrid progeny and the diploid *M. sinensis* and *M. sacchariflorus* parents were 0.49 and 0.42, respectively (Table 2.5). Comparably, Jaccard's similarity coefficients of the triploid hybrid progeny from the cross of diploid *M. sinensis* with tetraploid *M. sacchariflorus* was also intermediate between the parents and were 0.47 and 0.50, respectively (Table 2.5). 'Purpurascens' accessions were located between section *Miscanthus* and *Triarrhena* (Figure 2.6b) suggesting that these accessions may be putative hybrids generated from a cross between an accession from the section *Miscanthus* groups and diploid *M. sacchariflorus*. *M. x giganteus* accessions could be putative hybrids between section *Miscanthus* and tetraploid *M. sacchariflorus* as the latter is clearly distinct from diploid *M. sacchariflorus* (Figure 2.6a and b). These results also conform to the genome size data where hybrids between two parental accessions show mid-parent values for both diploid 'Purpurascens' and triploid *M. x giganteus*.

2.5. Discussion

Genome size estimation

The basic chromosome number of ancestral Saccharinae species is believed to be 10 as this is the number of chromosomes found in the majority of current species (Burner, 1991). The basic chromosome number of *E. arundinaceus* is also x=10 (D'Hont *et al.*, 1995). This is supported by our results where the three *Erianthus* accessions were observed to have 30, 40 and 60 chromosomes, (Table 2.1; Figure 2.2) and by the 2C-DNA value (Table 2.1) and cluster analysis (Figure 2.5). All *Miscanthus* accessions investigated had the basic chromosome number of x=19. The estimated genome size of species of *Miscanthus* from our analysis (2058 - 2773 Mbp, Table 2.1) is comparable to a previous report by Rayburn et al. (2009; 2150–2650 Mbp) and Nishiwaki *et al.* (2011). The estimated genome size of *S. officinarum* 'LA purple' was 950.9 Mbp (Table 2.1) and this is very similar to the 957.2 Mbp reportd by Zhang *et al.* (2012).

Classification of accessions using morphological, flow cytometry and molecular markers

There is no consensus as yet on the taxonomic definition and the number of species or subspecies of *Miscanthus* (Clifton-Brown *et al.*, 2008; Sun *et al.*, 2010). Taxonomic studies conducted with *Miscanthus* have been based on morphology or molecular maker analysis. Flow cytometry and chromosome counts can provide additional and important information on genome size and ploidy levels of the genus *Miscanthus* but only limited information was previously available on nDNA content (Raybrun *et al.*, 2009) and chromosome numbers (Adati & Shiotani, 1962; Burner, 1991; Linde-Laursen, 1993). This study provides the first

comprehensive analysis combining morphology, molecular marker analysis and genome size to improve our understanding of the taxonomy and evolution of the genus *Miscanthus*.

Miscanthus accessions were first introduced in to the United States from Japan in the late 1800s (Quinn *et. al*., 2010) with additional ornamental *Miscanthus* accessions introduced thereafter. Errors in the proper taxonomic identification of many accessions exist due to the lack of taxonomic knowledge and molecular marker information or from simple mislabeling of plants. Using our combined analyses, some of misclassified accessions have been corrected, including *M. sacchariflorus* 'KB giganteus', *Saccharum* spp. 'Raiatea' (Table 2.1 and Figure 2.5) and other horticultural accessions (data not shown). Genome size and clustering analysis suggest that 'NG51-88' is more likely *Erianthus* (Table 2.1; Figure 2.5) although Arro *et al*. (2006) reported that 'NG55-088' is *S. robustum* and groups with both *Erianthus* and *S. robustum*. Given that the mean monoploid genome size of *S. robustum* is similar to that of *S. officinarum* (984.9 Mbp) (Zhang *et al*., 2012), which is smaller than 'NG55-088' (1195.6 Mbp, Table 2.1), we feel 'NG51-88' can be assigned to *Erianthus*.

Based on morphology and genome size information of *Miscanthus* accessions and of diploid and triploid artificial hybrids, we hypothesize that *Miscanthus* 'Purpurascens' accessions are hybrids between sections *Miscanthus* and diploid *M. sacchariflorus* and *M. x giganteus* between section *Miscanthus* and tetraploid *M. sacchariflorus*. Also, 'NG51-088' and *Miscanthus* 'Raiatea' belong to *Erianthus* and *Saccharum* hybrids, respectively. To support this with molecular maker analysis, SSR markers were developed for classifying accessions. For ornamental *Miscanthus* accessions, species status has been also assigned by private nurseries, not by taxonomic authorities. Therefore, we included USDA-NPGS accessions as standards for species determination in the combined analysis.

Nearly all *Miscanthus* accessions fell into 2 major clusters which is generally in agreement with the two taxonomic sections, *Miscanthus* and *Triarrhena* (Figure 2.5) with 6 distinctive subgroups (Table 2.1 and Figure 2.6b). Three subgroups in section *Miscanthus* probably represent intraspecific variation since all three subgroups include *M. sinensis* accessions from USDA-NPGS (Figure 2.5). Low genetic similarities (< 0.5, Figure 2.3; Table 2.5) of a number of the accessions within the section *Miscanthus* group is most likely due to a bias from using SSR primers initially designed for detecting polymorphisms between *M. sinensis* accessions for genetic map construction (Swaminathan *et al*., 2012). That two USDA-NPGS *M. floridulus* accessions were grouped with *M. sinensis* accessions (Figure 2.5) make the distinct species status of these accessions questionable, which is consistant with a previous report (Hodkinson *et al*. 2002b). *M. tinctorius* belongs to section *Kariyasua* (Lee, 1993) but the two *M. tinctorius* accessions in this study were also grouped with *M. sinensis* accessions (Table 2.1 and Figure 2.5 and 2.6a and b). Therefore, this study did not observe the distinct taxonomic status of section *Kariyasua*. Previous research supports three sections in the genus *Miscanthus* since *M. oligostachyus* was included in the *Miscanthus* group but was separated from other *Miscanthus* species (Hodkinson *et al*., 2002b).

The section *Triarrhena* group was divided into 3 subgroups, diploid *M. sacchariflorus*, tetraploid *M. sacchariflorus* and *M.sacchariflorus var. lutarioriparius*, and *M.* 'Purpurascens'. Morphology and genome size data could not distinguish the first two subgroups (Table 2.1; Figure 2.1) but cluster analysis clearly divided these accessions into two groups (Figure 2.5 and 2.6). Sun *et al*. (2010) also reported that the only difference between diploid *M. sacchariflorus* and *M. sacchariflorus var. lutarioriparius* is plant height. Thus, the second subgroup containing

tetraploid *M. sacchariflorus* and *M.sacchariflorus var. lutarioriparius* can be classified as a variety within *M. sacchariflorus*.

M. x giganteus and 'Purpuracens' accessions were grouped with *M. sacchariflorus* accessions in UPGMA dendrogram (Figure 2.5). However, morphology (Figure 2.1), genome size (Table 2.1) and the PCoA plot (Figure 2.6) clearly distinguished these accessions from *M. sacchariflorus*, suggesting their hybrid origins. *M. x giganteus* has been hypothesized as hybrid between tetraploid *M. sacchariflorus* and diploid *M. sinensis* (Lindle-Laursen, 1993, Hodkinson *et al*., 2002c; Rayburn *et al*., 2009). The combined analyses using morphology, genome size, genetic similarity and comparison with known triploid hybrid progeny supports this observation (Table 2.1 and 2 and Figure 2.6b). The taxonomic status of 'Purpurascens' accessions have been unclear and assigned to *M. oligostachyus* (Brand, 1999), *M. sinensis var.* purpurascens or *Miscanthus* 'Purpurscens' (Deuter, 2000) and *M. sacchariflorus* (Hodkinson *et al*., 2002a). Here, we suggest a hybrid origin of *Miscanthus* 'Purpurascens' between diploid *M. sacchariflorus* and section *Miscanthus* based on the combined analyses (Table 2.1 and 2 and Figure 2.6b).

Evolution of genome size in *Miscanthus* and related genera

Understanding the evolution of plant genome size involves 1) genome size distribution within and among plant taxa, 2) changes in genome size in a phylogenetic context, 3) the mechanisms changing genome size and 4) phenotypic consequences influencing taxonomic and geographical distribution of genome size variation among species (Bennett & Leitch, 2005).

Monoploid genome sizes of *S. bicolor* and *Saccharum* accessions were 852.6 and 940.8-1038.8 Mbp, respectively (Table 2.1), and these results were generally consistent with previous reports for *Eusorghum* (745–818 Mbp) (Price *et al*, 2005) and *Saccharum* speices (732.7-1046

Mbp) (Zhang *et al*., 2012). Monoploid genome sizes of *Erianthus* ranged from 1166.2 to 1205.4 Mbp (Table 2.1), which is the first report of genome size estimation for accessions in the genus *Erianthus*. *Eusorghum*, *S. officinarum* and *Erianthus* typically have the basic chromosome numbers of x = 10. In contrast, the genus *Miscanthus* has a consistant base chromosome number of x = 19 with genome sizes ranging from 2058 to 2773 Mbp (Table 2.1). The UPGMA dendrogram suggests that *Saccharum* and *Miscanthus* are more closely related than *Erianthus* (Figure 2.5).

Chromosome number, genome size, and geographical information on the genus *Miscanthus* in this study support the hypothesis of *Miscanthus* originating via whole genome duplication (Paterson *et al*. 2010). Recent genetic mapping studies revealed whole genome duplication and a chromosome fusion event in neopolyploid *Miscanthus* during the process of its divergence from other genera (Ma *et al*. 2012; Swaminathan *et al*., 2012). The possible genome size of ancestral accessions of the genus *Miscanthus* may be between 2.45 and 2.58 pg which was determined from four wild accessions collected in New Guinea, Taiwan and Thailand (Figure 2.5) where other closely related genera coexist. The genome size of *Miscanthus* compared to *Eusorghum* and *Saccharum* species suggest additional increases in genome size occurred after whole genome duplication, which is probably associated with the content of repetitive DNA in the *Miscanthus* genome comprising 95% of the genome of *M. x giganteus* (Swaminathan *et al*., 2010). After divergence from other related genera, the *Miscanthus* genome may have gone through genome size alteration in the process of diversification resulting in the variation observed across accessions (Table 2.1, Figure 2.5). Additional genome doubling in *M. sacchariflorus* may have occurred in Korea (Figure 2.7) where sympatric populations of diploid and tetraploid *M. sacchariflorus* are commonly observed (Dr. Yoon Ho Moon, Personal

communication). This is supported by reports that all *M. sacchariflorus* accessions sampled in Japan were tetraploids, while the ploidy level of *M. sacchariflorus* in China is almost exclusively diploid (Nishiwaki *et al*., 2011).

Phenotypic consequences of genome size variation

One form of phenotypic variation that is associated with geographic distribution of species is days to flowering. All diploid *M. sacchariflorus* accessions generally flower one or two months earlier than most *M. sinensis* accessions (section *Miscanthus*) both in the greenhouse (Chae *et al*., 2012) and in the field (T. Voigt, personal communication). Given that the diploid *M. sacchariflorus* accessions 'Robustus' was collected from Siberia or Northern China (Deuter, 2000) and that our diploid *M. sacchariflorus* accessions showed a high degree of genetic similiary (Table 2.5 Figure 2.5 and 2.6a and b), this region may be the origin of *M. sacchariflorus* (Figure 2.7). The smaller genome size of *M. sacchariflorus* could be related to early flowering which may facilitate completion of the life cycle of plants under a shorter growing season. A positive relationship between nDNA content and minimum generation time across plant species (in review of Bennett & Leitch, 2005) supports the origin of *M. sacchariflorus* in Siberia. In contrast, one of the tetraploid *M. sacchariflorus* accessions, 'KB giganteus' did not complete its life cycle and did not flower in field plots at UIUC. Access to only two tetraploid *M. sacchariflorus* accessions and a limited number of diploid *M. sacchariflorus* accessions in the U.S. precludes the investigation into phenotypic difference between different ploidy levels among *M. sacchariflorus* accessions. Korea may be the best place to study flowering time differences between diploid and tetraploid *M. sacchariflorus*, where natural diploid and tetraploid *M. sacchariflorus* populations co-exist.

In conclustion, combined taxonomic analysis based on morphology, flow cytometry, chromosome counts and SSR markers suggests that the genus *Miscanthus* consists of three species (*M. sinensis*, *M. sacchariflorus* and *M. x giganteus*) and one *M. sacchariflorus* variety (tetraploid *M. sacchariflorus* and *M.sacchariflorus var. lutarioriparius*) and a putative hybrid, *Miscanthus* 'Purpurascens' in our accessions. The combined analysis can be useful for *Miscanthus* germplasm characterization for biofuel crop improvement and to understand the evolution of the genus *Miscanthus*.

Table 2.1. Origin, chromosome number, 2C-DNA value of two *Saccharum* accessions, seven putative hybrids between *Saccharum* and *Miscanthus* and five *Erianthus arundinaceus* 37 *Miscanthus* genotypes

Group	Accessions	Sources or USDA PI No.	Origin	Chr. No.	Internal standard	2C-DNA content (pg)	Ploidy level	Cx-value (pg)	Monoploid genome size (Mbp)
1	F_1 between CP 52-68 (*S. spp.*) x Tainan (*S. spontaneum*) 'L97-1002'*	Louisiana State University		-	s[z]	9.57 ± 0.16	10x	0.96	940.8
1	*S. officinarum* 'LA Purple'*			80	s	7.75 ± 0.04	8x	0.97	950.6
1	*Saccharum* hybrid 'Fiji 53'	PI 271853	Fiji	93+2f[z]	s	9.98 ± 0.13	10x	1.00	980.0
1	*Miscanthus spp.* 'Raiatea'*	Q 37075	Hawail		s	10.03 ± 0.05	10x	1.00	980.0
1	*Saccharum* hybrid 'Fiji 17'*	PI 212268	Fiji	100[z]	s	10.05 ± 0.07	10x	1.01	989.8
1	*Saccharum* hybrid 'Fiji 54'	PI 268060	Fiji	96+1f[z]	s	10.14 ± 0.24	10x	1.01	989.8
1	*Saccharum* hybrid 'Fiji 59'	PI 268061	Fiji	96+2f[z]	s	10.23 ± 0.16	10x	1.02	999.6
1	*Saccharum* hybrid 'Fiji 55'*	PI 271854	Fiji	98[z]	s	10.38 ± 0.17	10x	1.04	1019.2
1	*Saccharum hybrid* 'Fiji 57'	PI 276960	Fiji	98+1f[z]	s	10.57 ± 0.17	10x	1.06	1038.8
2	*E. arundinaceus* 'MPTH98-283'*	USDA	Thailand	40	s	4.75 ± 0.19	4x	1.19	1166.2
2	*E. arundinaceus* 'MPTH98-326'*	USDA	Thailand	-	s	4.77 ± 0.06	4x	1.19	1166.2
2	*E. arundinaceus* 'MPTH97-221'*	USDA	Thailand	-	s	4.80 ± 0.03	4x	1.20	1176.0
2	*E. arundinaceus* 'MPTH97-260'*	USDA	Thailand	60	s	7.27 ± 0.09	6x	1.21	1185.8
2	*S. robustum* 'NG51-088'*	PI 210240	New Guinea	30	s (z)	3.66 ± 0.11	3x	1.22	1195.6
2	*E. arundinaceus* 'MPTH97-194'*	USDA	Thailand	-	s	4.90 ± 0.09	4x	1.23	1205.4
3	*M. sacchariflorus* 'Gotemba gold'*	Glasshouse Works	Nursery	-	s	8.40 ± 0.10	4x	2.10	2058.0
3	*M. sacchariflorus* var. *lutarioriparius* *	'PF30022'		-	s	4.26 ± 0.04	2x	2.13	2087.4
3	*M. sacchariflorus* 'KB giganteus'*	Bluemel	Nursery	76	s (z)	8.65 ± 0.21	4x	2.16	2116.8
3	*M. sacchariflorus* 'Golf course'*			38	s	4.40 ± 0.08	2x	2.20	2156.0
3	*M. sacchariflorus* 'Robustus'*	Earthly pursuits	China	38	s	4.43 ± 0.02	2x	2.22	2175.6
3	*M. sacchariflorus*	Bluemel	Nursery	-	s	4.45 ± 0.11	2x	2.23	2185.4
3	*M. sacchariflorus**	Hortico	Nursery	-	s	4.47 ± 0.07	2x	2.24	2195.2
3	*M. sacchariflorus*	Earthly pursuits	Nursery	-	s	4.49 ± 0.05	2x	2.25	2205.0
5	*M. x giganteus* 'Illinois'*	Chicago botanic garden	Kew Garden	57	s (z)	6.86 ± 0.13	3x	2.29	2244.2
5	*M. x giganteus*	Hortico	Nursery	-	s	6.93 ± 0.21	3x	2.31	2263.8
5	*M. x giganteus* 'Freedom'	MSU-Baldwin		-	s	6.94 ± 0.25	3x	2.31	2263.8
5	*M. x giganteus*	Walla Walla	Nursery	-	s	6.96 ± 0.06	3x	2.32	2273.6
5	*M. x giganteus*	Long's Garden	Nursery	-	s	6.99 ± 0.07	3x	2.33	2283.4

Table 2.1. Continued

Group	Accessions	Sources or USDA PI No.	Origin	Chr. No.	Internal standard	2C-DNA content (pg)	Ploidy level	Cx-value (pg)	Monoploid genome size (Mbp)
5	*M. x giganteus* 'Glided Tower'*	Plant Delights	Nursery	-	s	6.99 ± 0.15	3x	2.33	2283.4
6	*M. oligostachyus* 'Purpurascens'*	SoyFACE (UIUC)	Nursery	38	s	4.88 ± 0.12	2x	2.44	2391.2
6	*M. oligostachyus* 'Purpurascens'	Walla Walla	Nursery	-	s	4.90 ± 0.17	2x	2.45	2401.0
6	*M. sinensis var.* 'purpurascens'*	Bluemel	Nursery	-	s	4.92 ± 0.06	2x	2.46	2410.8
7	*Miscanthus spp.* 'M75-062'*	PI 423566	Taiwan	-	s	4.90 ± 0.09	2x	2.45	2401.0
7	*Miscanthus spp.* 'US64-0016-03'*	PI 302423	Zambia	-	s	4.94 ± 0.15	2x	2.47	2420.6
7	*M. sinensis* 'US47-0011'*	CANE 9233	Taiwan	38[y]	s	5.00 ± 0.10	2x	2.50	2450.0
7	*M. sinensis* 'US 64-0004-02'*	PI 294602	Japan	38 [y]	s	5.03 ± 0.11	2x	2.52	2469.6
7	*M. sinensis* 'Goliath'*	SoyFACE (UIUC)	Nursery	57	s	7.57 ± 0.18	3x	2.52	2469.6
7	*M. floridulus* 'US 56-0022-03'*	PI 230189	Thailand	38[z]	s	5.06 ± 0.04	2x	2.53	2479.4
7	*M. floridulus**	PI 295762	Japan	-	s	5.06 ± 0.09	2x	2.53	2479.4
7	*Miscanthus spp.* 'NG77-022'*	PI 417947	New Guinea	38[y]	s	5.15 ± 0.03	2x	2.58	2528.4
7	*M. sinensis* 'Grosse fontaine'*	SoyFACE (UIUC)	Nursery	38	s	5.17 ± 0.06	2x	2.59	2538.2
	M. sinensis 'Undine'*	SoyFACE (UIUC)	Nursery	38	s	5.19 ± 0.08	2x	2.60	2543.1
7	*M. sinensis* 'Uruyu'		Japan		s	5.21 ± 0.02	2x	2.61	2557.8
7	*M. sinensis var. condensatus* 'Cosmopolitan'		Nursery	38	s	5.22 ± 0.11	2x	2.61	2557.8
7	*M. transmorrisonensis* 'Arnin's Weeping'	Joy Creek Nursery	Nursery	-	s	5.24 ± 0.13	2x	2.62	2567.6
7	*M. sinensis var. condensatus* 'Condensatus'*		Nursery	-	s	5.27 ± 0.11	2x	2.64	2587.2
7	*M. transmorrisonensis**	Walla Walla	Nursery	-	s	5.37 ± 0.08	2x	2.69	2636.2
7	*M. sinensis* 'US64-0007-01'*	PI 294605	Japan	-	s	5.37 ± 0.10	2x	2.69	2626.4
7	*M. sinensis var. condensatus* 'Cabaret'*		Nursery	-	s	5.42 ± 0.09	2x	2.71	2655.8
7	*M. tinctorius* 'Gifu 2010-12-06'*	Seed	Japan	-	s	5.49 ± 0.12	2x	2.75	2695.0
7	*M. tinctorius* 'Gifu 2010-12-15'*	Seed	Japan	-	s	5.66 ± 0.22	2x	2.83	2773.4

[z] Price and Daniels (1968) Cytology of South Pacific Sugarcane and related grasses. The Journal of heredity **59**, 141-145.
[y] Burner (1991) Cytogenetic analyses of sugarcane relatives (Andropogoneae : Saccharinae) Euphytica **54**, 125-133 .
* Accessions used in cluster analysis.
[z] S indicate sorghum and s(z) means nuclear DNA content of accessions were confirmed both sorghum and maize.

Table 2.2. Nuclear DNA content and genome size of four parental accessions and diploid and triploid hybrid progeny.

Parents and progeny hybrid accessions	nDNA content (pg)	Cx-value (pg)
M. sacchariflorus 'Golf course' (2x) × *M. sinensis* 'Grosse Fontaine' (2x)		
GCxGF 2x-07*	4.78 ± 0.06	2.39
GCxGF 2x-49	4.96 ± 0.03	2.48
GCxGF 2x-28*	4.97 ± 0.19	2.49
GFxGC 2x-09	4.89 ± 0.03	2.45
GFxGC 2x-02*	4.96 ± 0.08	2.48
Parent 1: *M. sacchariflorus* 'Golf course'	4.46 ± 0.14	
Parent 2: *M. sinensis* 'Grosse Fontaine'	5.25 ± 0.03	
Expected values for progenies 1x from P3 (2.23) + 1x from P4 (2.625)	4.86	2.43
M. sacchariflorus 'KB giganteus' (4x) × *M. sinensis var condensatus* 'Cabaret' (2x)		
10UI-032-01	6.82 ± 0.10	2.27
10UI-032-02	7.02 ± 0.01	2.34
10UI-032-03*	6.91 ± 0.17	2.30
Parent 3: *M. sacchariflorus* 'KB giganteus' (4x)	8.59 ± 0.17	2.15
Parent 4: *M. sinensis var condensatus* 'Cabaret' (2x)	5.39 ± 0.08	2.70
Expected values for progenies: 2x from P1 (4.295) + 1x from P2 (2.695)	6.92	2.31

* Accessions used in cluster analysis.
Flow cytometry analyses for a set of all hybrid lines and parental lines were conducted at the same day and each set was replicated three times on different days.

Table 2.3. Chromosomal position in sorghum and *Miscanthus* and sequence information of 16 SSR primers.

Primer	Sor. chr.	Mis. LGs	SSR unit	Forward primers	Reverse primers
ESSR02	1	1	GGC	CGGACTGTTCGAAAGCAACC	GGCCAAAAGCTTACGATCCC
GSSR69	1	2	GTG	GAGGATCAGGAACTGGGAGTGAC	CTCCGATCTAATCCATTCCATCAC
ESSR12	2	3	CGC	AACGAGGACAACGCCAACTC	TCGATCACGAGAGGAGAGCC
GSSR15	2	4	CACG	TACTAGCTAAACGGACGTGCAGTG	GAAGATTCTCTGTGCTGACTGGGT
ESSR29	3	5, 6	GA	GCCAGCCATCAGCAATAGGT	GCGAGCTCCAATTTGCAATC
ESSR25	4	7, 8	ACCCT	GCTCCTCCCACTCGAGGAAT	TTGAGACGACGAGGGTTGGT
GSSR45	5	9	TTTGT	ATGTTTACTTCACCCTTTTTGGGC	CACGAAGTTTTTCGCGTTTCTTAC
GSSR12	5	10	TGC	TGCAAAAACACTCCAATGACTTGT	GTTCATGACACTTTGACTCGATGG
GSSR19	6	11	AG	TCAAGTTGGAGTTTCATCAGCATC	AGGTACGCAGGGACCACATATAAA
ESSR11	6	11	TCC	CCAGCTTCACCGAGGCATAC	GTAGGAGGACGGGTGCTGGT
ESSR08	6	12	CAAGCA	CAGCCATTCAAGGACAAGCC	CACGCCAATCACGGTTATCA
GSSR41	7	13	TG	ATTTTGAAATAAGAAAGACGGCCA	ACACACAAACACACTCATCATTCG
GSSR75	8	14	GCACCG	TGGACATTTACATCCACTTTGCAG	AGAGAAACAGGAGAGGCACTAGCA
GSSR14	8	15	TCC	GACATCCACATGTCGTACTTGGAC	ATCAGAACTTCAAGAGCTTCCGC
GSSR05	9	16, 17	CCG	ACCCAATAAGCACGGAACCCTA	AAGCGAGCGTACCTGTAGAGGAG
ESSR04	10	18, 19	CGG	CAAGGTCGTCTCCATCCTGG	CACTCCCAGGGCTTCAGGTT

Sor. Chr.: Sorghum chromsomes, *Mis*. LGs: *Miscanthus* lingkage groups (Swaminathan *et al*., 2012)

Table 2.4. Intergeneric and interspecific amplification of selected *Miscanthus* EST and genomic SSRs.

Marker name	*S.bi*	*Sa.spp.*	*E.ar*	*M.lu*	4x *M.sa*	2x *M.sa*	*Mxg*	*M. fl*	*M.* 'Pur'	*M si*	*M. tr*	*M.ti*
ESSR02	+	+	-	+	+	+	+	+	+	+	+	+
GSSR69	+	+	+	+	+	+	+	+	+	+	+	+
ESSR12	+	+	+	+	+	+	+	+	+	+	+	+
GSSR15	-	+	-	-	-	-	+	+	-	+	+	+
ESSR29	+	+	+	+	+	+	+	+	+	+	+	+
ESSR25	-	+	+	+	+	+	+	+	+	+	+	+
GSSR45	-	+	+	+	+	+	+	+	+	+	+	+
GSSR12*	-	+	+	+	+	+	+	+	+	+	-	+
GSSR19	-	+	+	+	+	+	+	+	+	+	+	+
ESSR11	-	+	+	+	+	+	+	+	+	+	+	+
ESSR08	+	+	+	+	+	+	+	+	+	+	+	+
GSSR41	-	+	+	-	+	+	+	+	+	+	+	+
GSSR75	+	+	+	+	+	+	+	+	+	+	+	+
GSSR14	-	+	+	+	+	+	+	+	+	+	+	+
GSSR05	+	+	+	+	+	+	+	+	+	+	+	+
ESSR04	+	+	+	+	+	+	+	+	+	+	+	+

+ Amplification; -No amplification

S.bi: *Sorghum bicolor*, *Sa.spp*: *Saccharum* species, *E. ar*: *Erianthus arundinaceus*, *M.lu*: *Miscanthus lutarioriparius*, 4x *M.sa*: Tetraploid *M. sacchariflorus*, 2x *M.sa*: diploid *M. sacchariflorus*, *Mxg*: *M. x giganteus*, *M.fl*: *M. floridulus*, *M.* 'pur': *Miscanthus* 'Purpuracens', *M.si*: *M. sinensis*, *M.tr*: *M. transmorrisonensis*, *M.ti*: *M. tinctorius*

* GSSR12 was excluded since genotyping results were not consistent with screening result.

Table 2.5. Jaccard's similarity coefficient matrices showing genetic similarity of artificial hybrids and their parents and putative hybrid *Miscanthus* 'Purpuracens' and *M. x giganteus* 'Illinois' and 'Gilded tower'.

No	Accessions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	M.fl_US56-22-03	1.00																	
2	M.sa_Golfcourse	0.17	1.00																
3	M.sa_Robustus	0.19	0.92	1.00															
4	M.sa_Gotembagold	0.24	0.46	0.48	1.00														
5	M.sa_KBgiganteus	0.28	0.34	0.33	0.52	1.00													
6	M.si_US64-04-02	0.51	0.17	0.18	0.22	0.23	1.00												
7	M.si_Goliath	0.36	0.23	0.25	0.19	0.28	0.36	1.00											
8	M.sivc_Cabaret	0.51	0.18	0.17	0.19	0.27	0.40	0.41	1.00										
9	M.ol_Purpurascens	0.27	0.43	0.43	0.29	0.29	0.27	0.44	0.24	1.00									
10	M.si_Purpurascens	0.27	0.43	0.46	0.32	0.29	0.27	0.44	0.21	0.94	1.00								
11	M.tr	0.33	0.29	0.30	0.27	0.21	0.36	0.33	0.30	0.31	0.33	1.00							
12	Mxg_Illinois	0.42	0.31	0.33	0.37	0.43	0.40	0.43	0.40	0.36	0.36	0.32	1.00						
13	Mxg_Gildedtower	0.28	0.29	0.29	0.50	0.82	0.25	0.31	0.27	0.31	0.31	0.25	0.51	1.00					
14	M.si_Grossefontaine	0.38	0.15	0.17	0.13	0.24	0.33	0.51	0.44	0.23	0.23	0.32	0.37	0.26	1.00				
15	10UI-32-13(3x)	0.38	0.31	0.33	0.48	0.50	0.31	0.35	0.47	0.31	0.33	0.30	0.51	0.58	0.26	1.00			
16	GCxGF2x-7	0.30	0.41	0.43	0.31	0.31	0.27	0.40	0.33	0.29	0.32	0.33	0.41	0.33	0.51	0.33	1.00		
17	GCxGF2x-28	0.30	0.45	0.47	0.32	0.30	0.25	0.43	0.31	0.35	0.35	0.33	0.40	0.29	0.44	0.35	0.54	1.00	
18	GFxGC2x-2	0.25	0.40	0.42	0.29	0.33	0.25	0.46	0.33	0.33	0.36	0.34	0.40	0.32	0.53	0.32	0.78	0.62	1.00

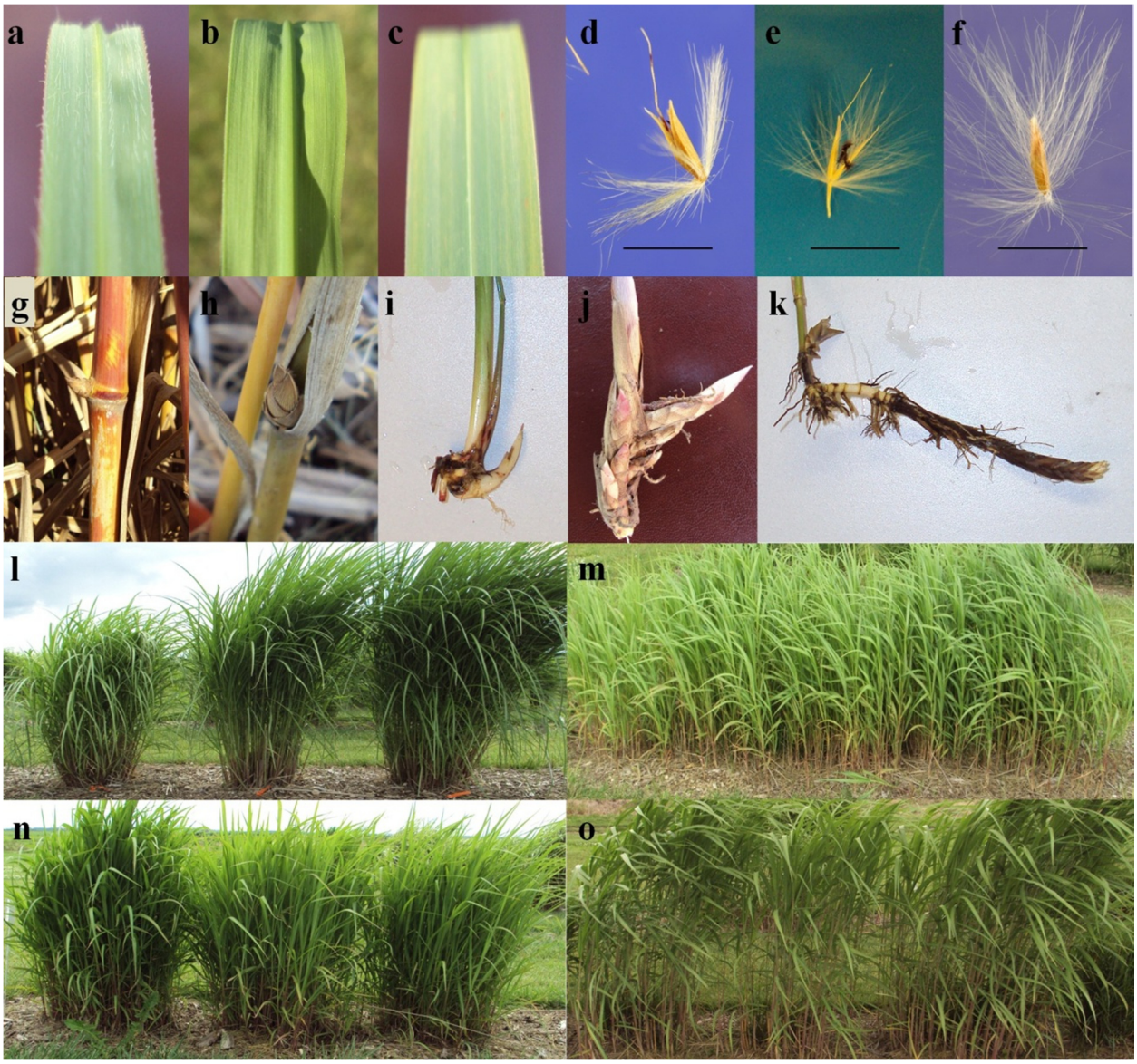

Figure 2.1. Key traits for morphology-based taxonomy of *Miscanthus* genus. a) a pilose abaxial leaf of *M. sinensis*; glabrous abaxial leaaves of b) *M. sacchariflorus* and c) 'Purpurascens'; awns in spikelets of d) *M. sinensis* e) 'Purpurascens'; f) awnless spikelets of *M. sacchariflorus*; buds at the node of g) *M. xgiganteus* 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 habits of *M. sacchariflorus*; intermediate growth habits of n) 'Purpurascens'and o) *M. x giganteus*. *Miscanthus* accessions were panted at 0.9 m spacing and have grown for 5 years. Bars represent 5mm. Pictures were taken from May to October 2010 in the Energy Bioscience Institute Farm at UIUC. Images of d and f were provided as a courtesy of Justin M. Gifford.

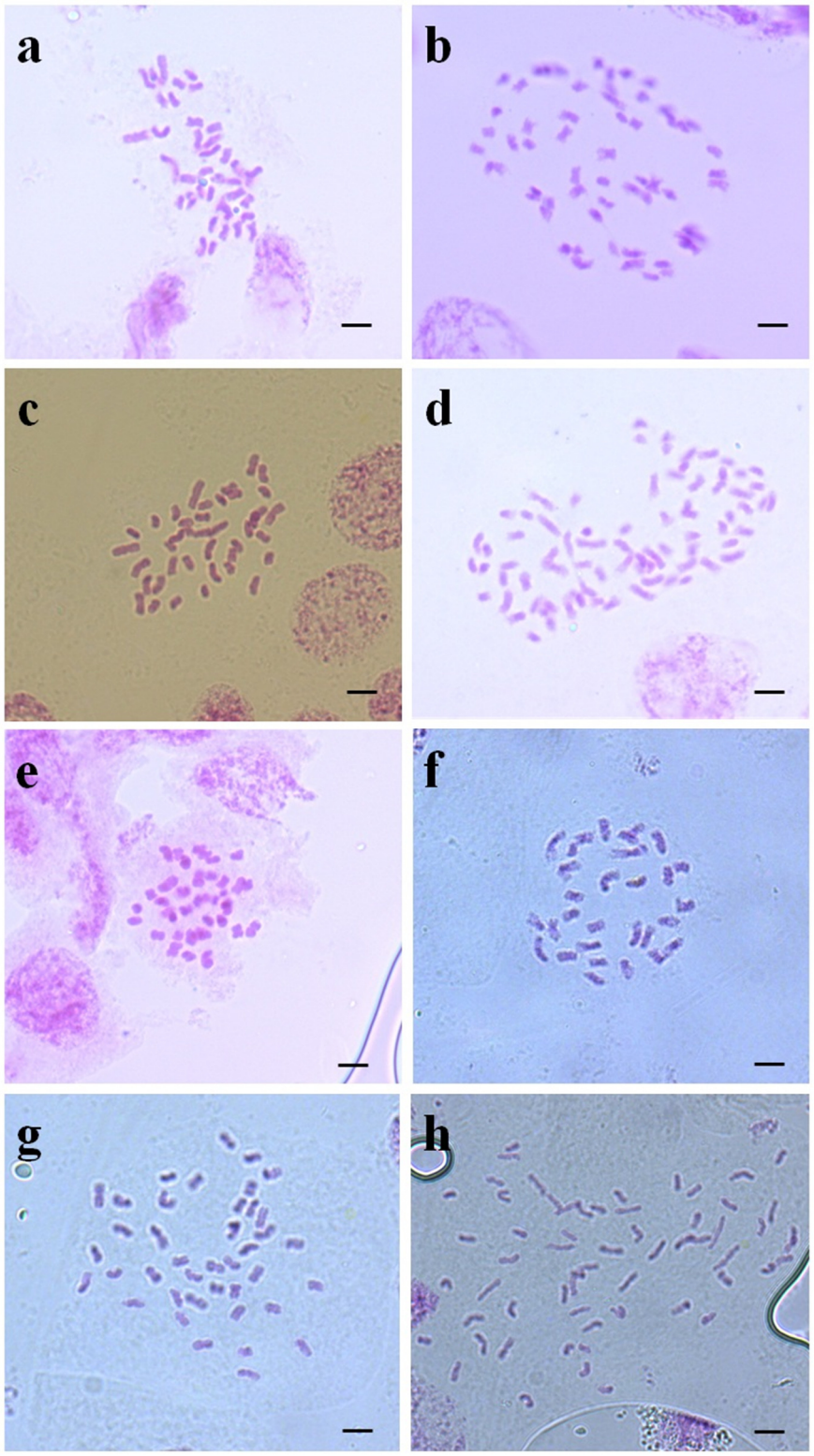

Figure 2.2. Chromosomes of five *Miscanthus* and three *Erianthus* accessions. Scale bars = 10 µm. (a) *M. sinensis* 'Undine' (2n=2x=38); (b) *M. sinensis* 'Goliath' (2n=3x=57); (c) *M. sinensis var. condensatus* 'Cosmopolitan' (2n=2x=38) (d) M. sacchariflorus 'KB giganteus' (2n=4x=76); (e) *Miscanthus* 'Purpurascens' (2n=2x=38); (f) 'NG 51-088' (2n=3x=30); (g) *E. arundinaceus* 'MPTH98-283' (2n=4x=40); (h) *E. arundinaceus* 'MPTH97-260' (2n=6x=60). Images of b, c, d and e were provided courtesy of Dr. Sae Jin Hong.

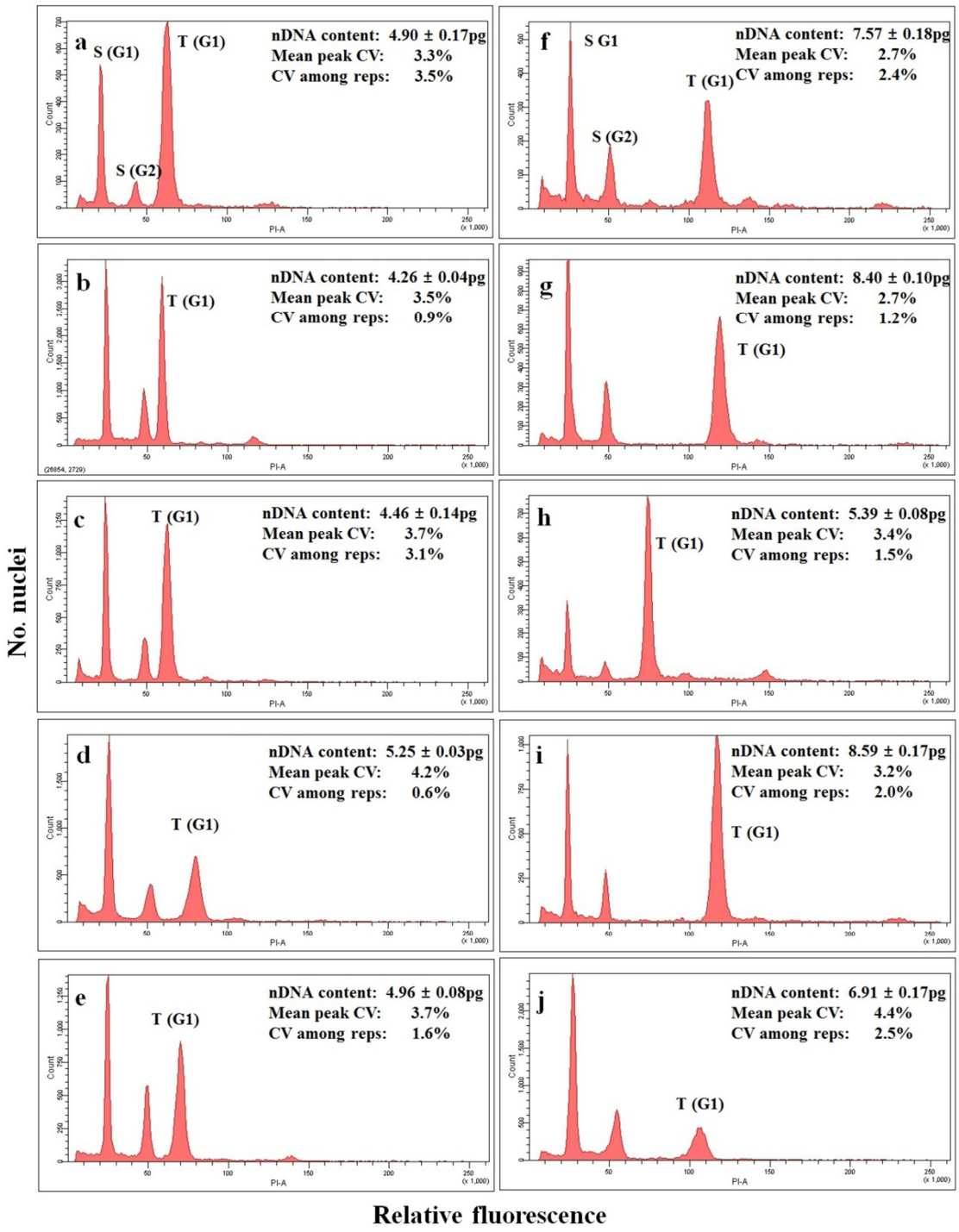

Figure 2.3. Histogram of nuclei extracted from leaf tissue of *Miscanthus* accessions and known interspecific hybrid progeny and its parents of *Miscanthus* species. Nuclei were stained with propidium iodide and sorghum was used as an internal standard. (a) *M.* 'Purpurascens'; (b) *M. sacchariflorus* var. *lutarioriparius*; (c) 2x *M. sacchariflorus* 'Golf course'; (d) 2x *M. sinensis* 'Grosse fontaine'; (e) 2x hybrid between (c) and (d); (f) *M. sinensis* 'Goliath'; (g) *M. sacchariflorus* 'Gotemba gold' (h) *M. sinensis var condensatus* 'Cabaret'; (i) 4x *M. sacchariflorus* 'KB giganteus'; (j); 3x hybrid '10UI-032-03' between (h) and (i). S (G1) and S (G2) indicate sorghum G0/G1peak and G2 peak, respectively. T (G1) indicates a target G0/G1peak. Coefficient of variation (CV) of a peak is a normalized standard deviation defined as CV = 100 × Standard Deviation / Mean of peak. CV among reps is calculated as 100 × standard deviation / Mean nDNA content.

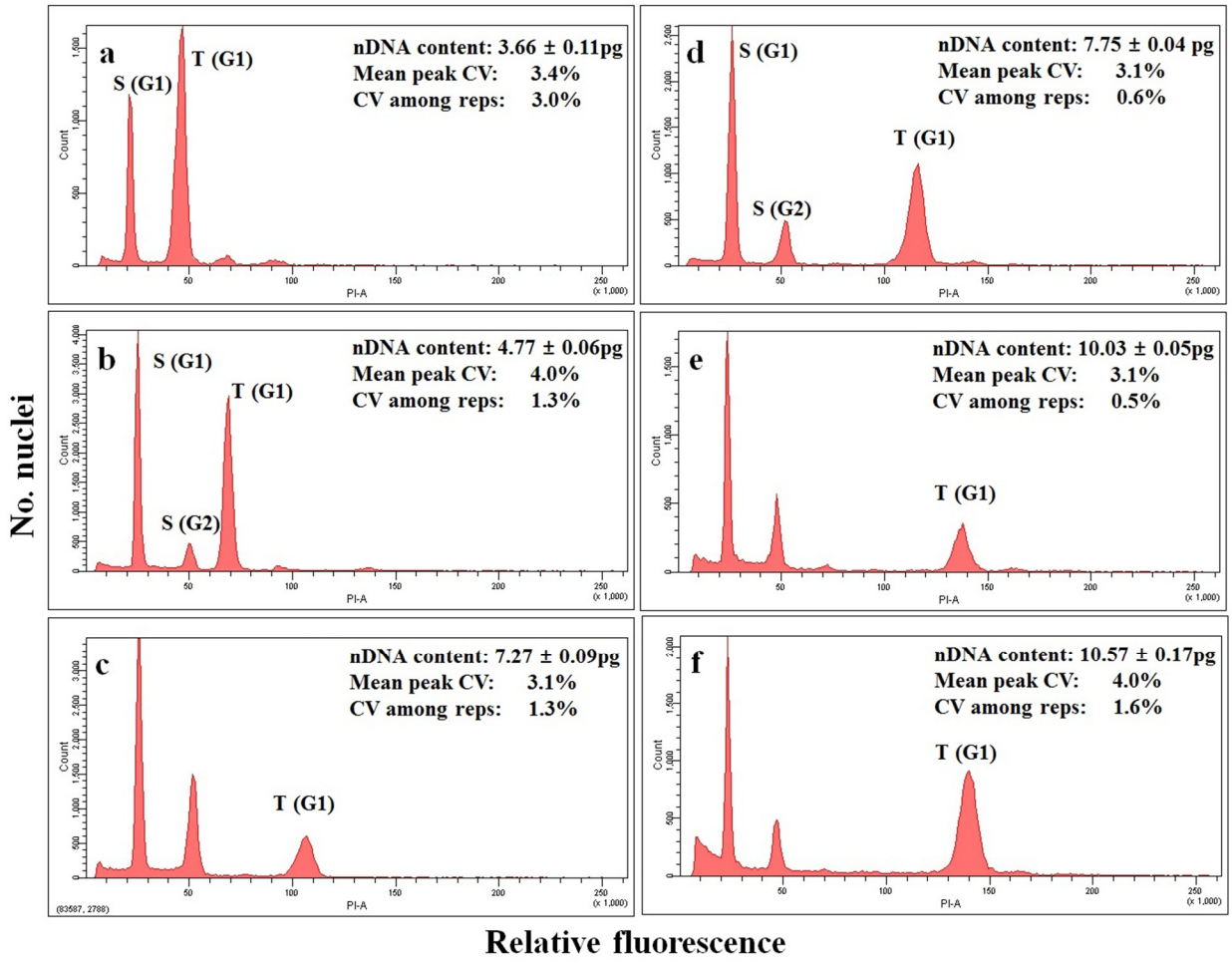

Figure 2.4. Histogram of nuclei extracted from leaf tissue of *Erianthus* and *Saccharum* accessions. Nuclei were stained with propidium iodide and sorghum was used as an internal standard. (a) *S. robustum* (*E. arundinaceus*) "NG51-088' (b) *E. arundinaceus* 'MPTH98-326'; (c) *E. arundinaceus* 'MPTH97-260'; (d) *S. officinarum* 'LA Purple'; (e) *Miscanthus spp.* 'Raiatea'; (f) *Saccharum* hybrid 'Fiji 55'. S (G1) and S (G2) indicate sorghum G0/G1peak and G2 peak, respectively. T (G1) indicates a target G0/G1peak. Coefficient of variation (CV) of a peak is a normalized standard deviation defined as CV = 100 × Standard Deviation / Mean of peak. CV among reps is calculated as 100 × standard deviation / mean nDNA content.

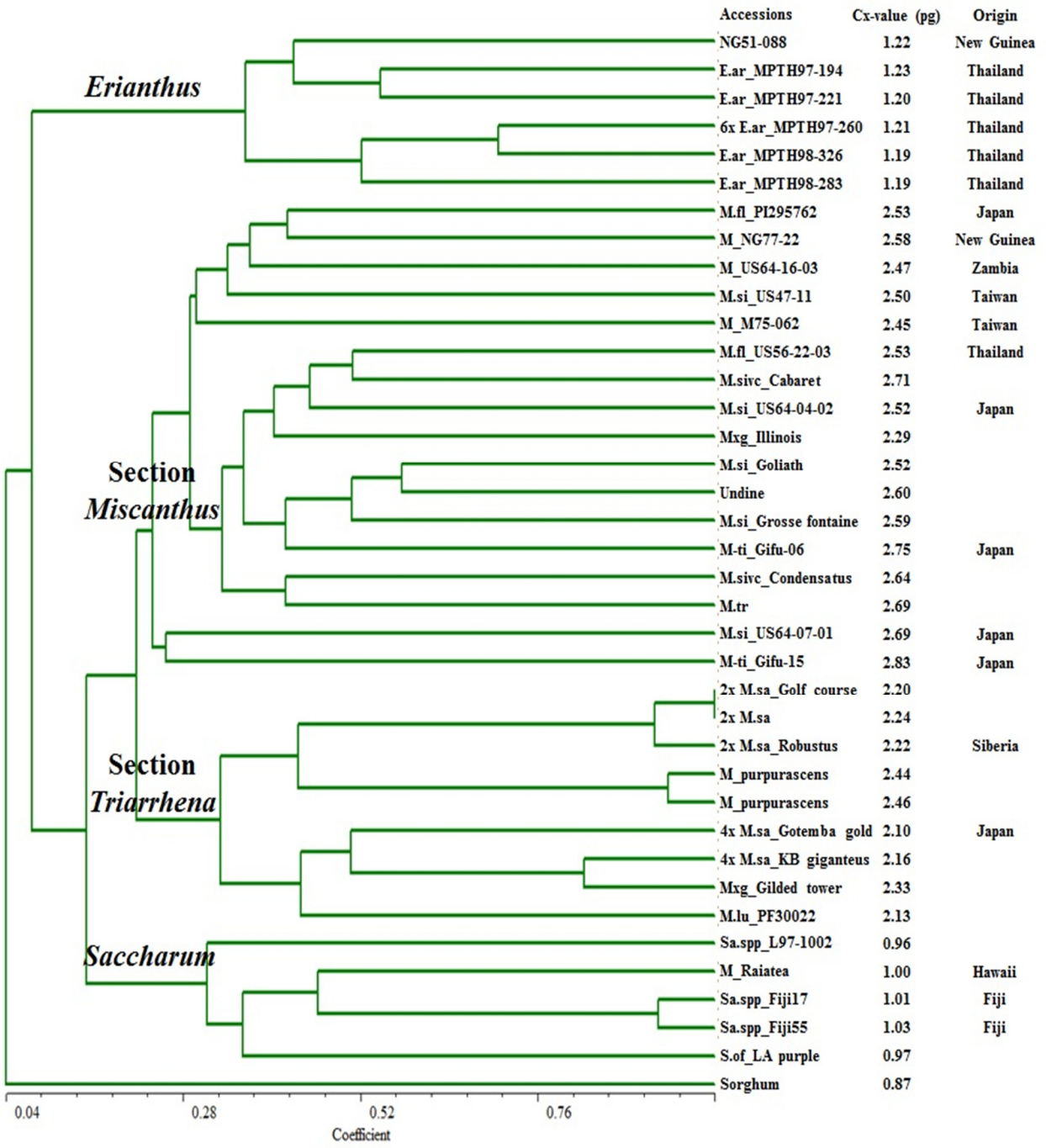

Figure 2.5. UPGMA dendrogram showing variation across taxa based on 16 SSR primers, genome size and origin among accession of *Saccharum*, *Erianthu*s, and *Miscanthus*. Sorghum is used as the reference outgroup. Species codes are *Sa.spp*: *Saccharum* species, E. ar: *Erianthus arundinaceus*, M_: *Miscanthus* species, M.lu: *Miscanthus sacchariflorus var. lutarioriparius*, 4x M.sa: Tetraploid *M. sacchariflorus*, 2x M.sa: diploid *M. sacchariflorus*, Mxg: *M. x giganteus*, M.fl: *M. floridulus*, M.si: *M. sinensis*, M.tr: *M. transmorrisonensis*, M.ti: *M. tinctorius.*

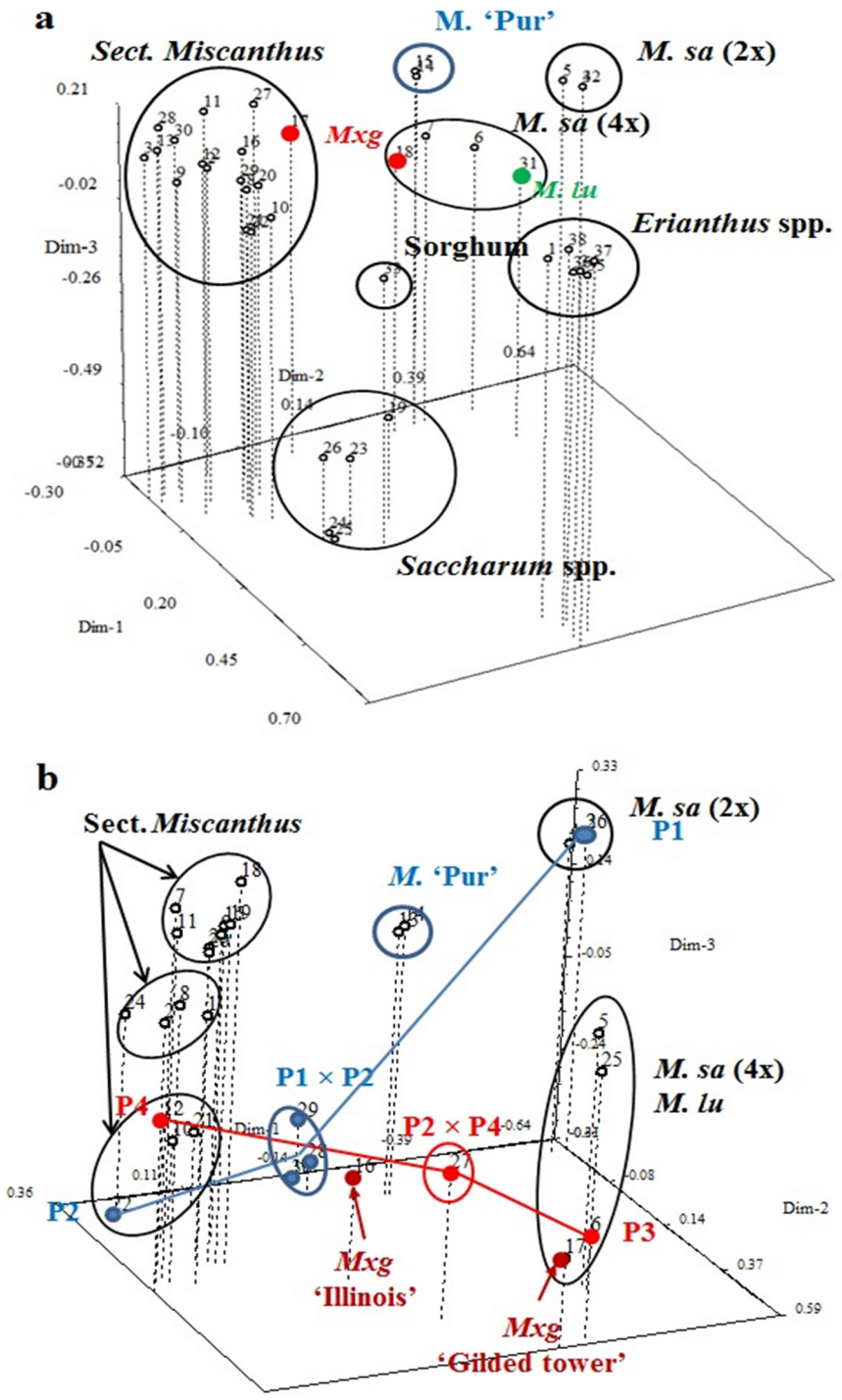

Figure 2.6. Principal coordinate analysis plot including all *Miscanthus* and related genera (a) and *Miscanthus* accessions and artificial hybrids (b) showing genetic similarity among *Saccharum*, *Erianthus*, and *Miscanthus*. Blue dots represent two parents and three 2x hybrids and red dots two parents and one 3x hybrid. Blue lines connect two parents of 2x hybrids and red lines connect two parents of 3x hybrids. Species codes are M.lu: *Miscanthus sacchariflorus var. lutarioriparius*, M.sa (4x): Tetraploid *M. sacchariflorus*, M.sa (4x): diploid *M. sacchariflorus*, Mxg: *M. x giganteus*, P1: *M. sacchariflorus* 'Golf course' (2x), P2: *M. sinensis* 'Grosse Fontaine' (2x), P1×P2: diploid hybrids from P1 × P2, P3: *M. sacchariflorus* 'KB giganteus' (4x), P4: *M. sinensis var condensatus* 'Cabaret' (2x), P3 × P4: triploid hybrids from P3 × P4.

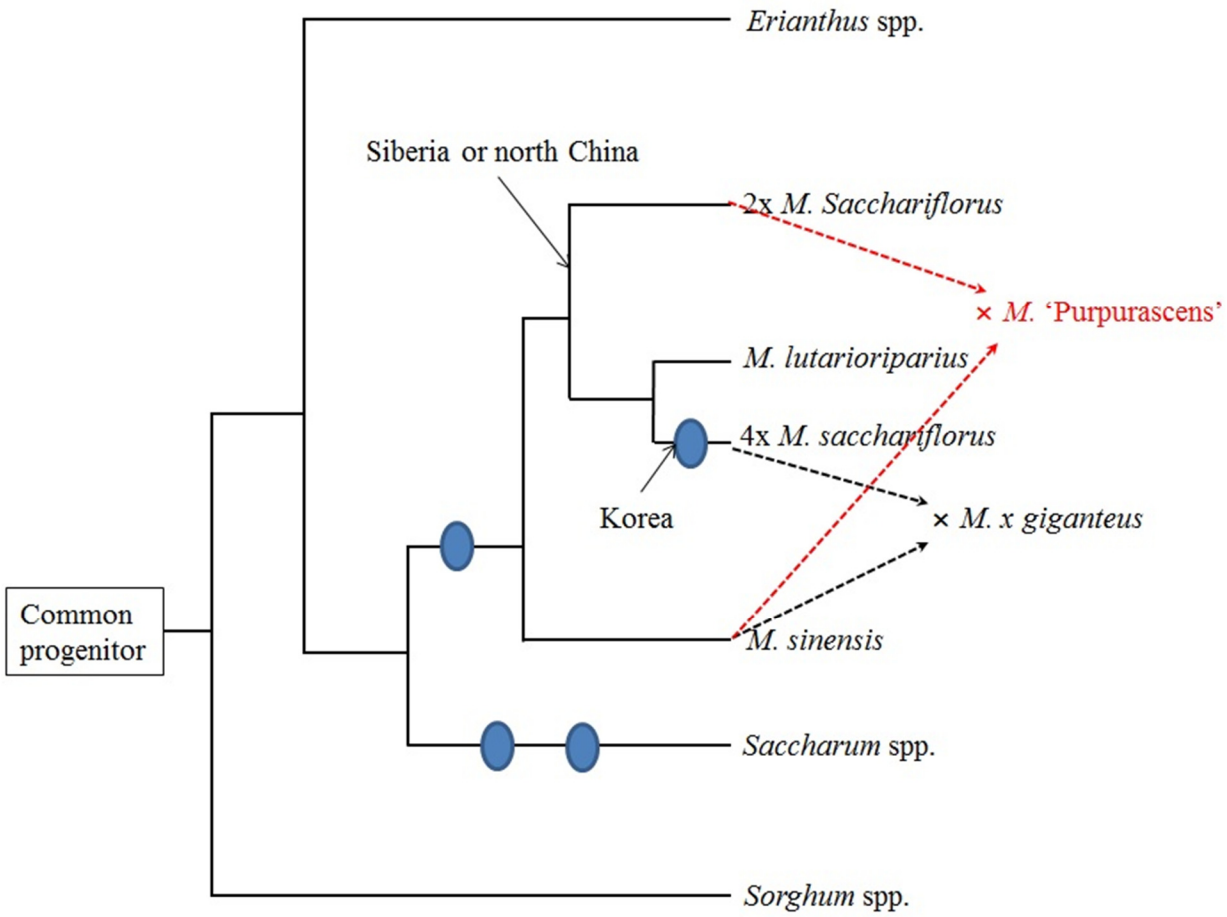

Figure 2.7. Hypothetical dendrogram for the evolution of *Miscanthus* species. Ovals indicate whole genome duplications. After whole genome duplication in common progenitor of *Miscanthus* (Kankshita *et al*., 2012), second genome duplication may have occured in Korea (See text). *M. sacchariflorus* may have diverged from other *Miscanthus* in Siberia or north China. *M. x giganteus* and *Miscanthus* 'Purpurascens' are putative hybrids between diploid *M. sinensis* and tetraploid *M. sacchariflorus* and between diploid *M. sinensis* and *M. sacchariflorus*, respectively (Table 2.1; Figure 2.1 and 2.4).

Supplementary Table 2.1. Morphological traits of 101 *Miscanthus* germplasm in UIUC.

No.	Acessions	Location	Awn in Spikelet	Buds at the nodes of culms	Abaxial leaf	Growth habit	Rhizome
1	*M. sacchariflorus* 'KB sacchariflorís'	EBI	-	+	G	R	S
2	*M. sacchariflorus* 'Earthly pursuit'	GH	-	+	G	R	S
3	*M. sacchariflorus* 'Golf course'	GH	-	+	G	R	S
4	*M. sacchariflorus* 'NIL'	GH	-	+	G	R	S
5	*M. sacchariflorus* 'Robustus'	EBI	-	+	G	R	S
6	*M. sacchariflorus* 'Silver banner grass'	EBI	-	+	G	R	S
7	*M. sacchariflorus* 'Hercules'*	EBI	-	+	G	R	S
8	*M. sacchariflorus* 'KB giganteus'*	EBI	-	+	G	R	S
9	*M.sacchariflorus var. lutarioriparius* 'PF30022'	GH	-	+	G	R	S
10	*M. sinensis* 'Adagio'	EBI	+	-	G	T	M
11	*M. sinensis* 'Allegro'	EBI	+	-	P	T	M
12	*M. sinensis* 'Altweibersommer'	EBI	+	-	P	T	M
13	*M. sinensis* 'Andante'	EBI	+	-	P	T	M
14	*M. sinensis* 'Andante'	Soyface	+	-	P	T	M
15	*M. sinensis* 'Arabesque'	EBI	+	-	P	T	M
16	*M. sinensis* 'Autumn light'	EBI	+	-	P	T	M
17	*M. sinensis* 'Berlin'	EBI	+	-	P	T	M
18	*M. sinensis* 'Blondo'	EBI	+	-	P	T	M
19	*M. sinensis* 'Bluetenwunder'	EBI	+	-	P	T	M
20	*M. sinensis* 'Bluetenwunder'	Soyface	+	-	P	T	M
21	*M. sinensis* 'Dixie land'	EBI	+	-	P	T	M
22	*M. sinensis* 'Ferner osten'	EBI	+	-	P	T	M
23	*M. sinensis* 'Flamingo'	EBI	+	-	P	T	M
24	*M. sinensis* 'Giraffe'	EBI	+	-	P	T	M
25	*M. sinensis* 'Gold & silber'	EBI	+	-	P	T	M
26	*M. sinensis* 'Gold & silber'	Soyface	+	-	P	T	M
27	*M. sinensis* 'Gold bar'	EBI	+	-	P	T	M
28	*M. sinensis* 'Goldfeder'	EBI	+	-	P	T	M
29	*M. sinensis* 'Goldfeder'	Soyface	+	-	P	T	M
30	*M. sinensis* 'Goliath'	EBI	+	-	P	T	M
31	*M. sinensis* 'Goliath'	Soyface	+	-	P	T	M
32	*M. sinensis* 'Gracillimus'	EBI	+	-	P	T	M
33	*M. sinensis* 'Gracillimus nana'	Soyface	+	-	P	T	M
34	*M. sinensis* 'Graziella'	EBI	+	-	P	T	M
35	*M. sinensis* 'Grosse fontaine'	EBI	+	-	P	T	M
36	*M. sinensis* 'Grosse fontaine'	Soyface	+	-	P	T	M
37	*M. sinensis* 'Haiku'	EBI	+	-	P	T	M

* Species were corrected based on morphology.

+ or -: presence or absecence of a given trait; P: pilose and G: glabrous abaxial leaves; T: tufted and R: Rhizomatous growth habits; M: monopodial and S: symopdial rhizomes; I: intermediate in a given trait.

Supplementary Table 2.1. Continued

No.	Acessions	Location	Awn in Spikelet	Buds at the nodes of culms	Abaxial leaf	Growth habit	Rhizome
38	*M. sinensis* 'Heiga reich'	EBI	+	-	P	T	M
39	*M. sinensis* 'Juli'	EBI	+	-	P	T	M
40	*M. sinensis* 'Kaskade'	EBI	+	-	P	T	M
41	*M. sinensis* 'Kaskade'	Soyface	+	-	P	T	M
42	*M. sinensis* 'Kirk alexander'	EBI	+	-	P	T	M
43	*M. sinensis* 'Kleine silberspinne'	EBI	+	-	P	T	M
44	*M. sinensis* 'Little kitten'	EBI	+	-	G	T	M
45	*M. sinensis* 'Little zebra'	EBI	+	-	P	T	M
46	*M. sinensis* 'Malepartus'	EBI	+	-	P	T	M
47	*M. sinensis* 'Minuett'	EBI	+	-	P	T	M
48	*M. sinensis* 'Morning light'	EBI	+	-	P	T	M
49	*M. sinensis* 'Mysterious maiden'	EBI	+	-	P	T	M
50	*M. sinensis* 'Nippon'	EBI	+	-	G	T	M
51	*M. sinensis* 'Nippon'	Soyface	+	-	P	T	M
52	*M. sinensis*	Soyface	+	-	P	T	M
53	*M. sinensis* 'November sunset'	EBI	+	-	P	T	M
54	*M. sinensis* 'Positano'	EBI	+	-	P	T	M
55	*M. sinensis* 'Positano'	Soyface	+	-	P	T	M
56	*M. sinensis* 'Puenktchen'	EBI	+	-	P	T	M
57	*M. sinensis* 'Red tango'	EBI	+	-	P	T	M
58	*M. sinensis* 'Rigoletto'	EBI	+	-	P	T	M
59	*M. sinensis* 'Rigoletto'	Soyface	+	-	P	T	M
60	*M. sinensis* 'Roland'	Soyface	+	-	P	T	M
61	*M. sinensis* 'Roter pfeil'	EBI	+	-	P	T	M
62	*M. sinensis* 'Rotsilber'	EBI	+	-	P	T	M
63	*M. sinensis* 'Rotsilber'	Soyface	+	-	P	T	M
64	*M. sinensis* 'Sarabande'	EBI	+	-	P	T	M
65	*M. sinensis* 'Silberfeder'	EBI	+	-	P	T	M
66	*M. sinensis* 'Silberfeder'	Soyface	+	-	P	T	M
67	*M. sinensis* 'Silberfeil'	Soyface	+	-	P	T	M
68	*M. sinensis* 'Silberspinne'	EBI	+	-	P	T	M
69	*M. sinensis* 'Silberspinne'	Soyface	+	-	P	T	M
70	*M. sinensis* 'Silbertum'	Soyface	+	-	P	T	M
71	*M. sinensis* 'Sirene'	EBI	+	-	P	T	M
72	*M. sinensis* 'Stardust'	EBI	+	-	P	T	M
73	*M. sinensis* 'Strictus'	EBI	+	-	P	T	M
74	*M. sinensis* 'Strictus'	Soyface	+	-	P	T	M
75	*M. sinensis* 'Super stripe'	EBI	+	-	P	T	M
76	*M. sinensis* 'Undine'	EBI	+	-	P	T	M
77	*M. sinensis* 'Undine'	Soyface	+	-	P	T	M
78	*M. sinensis* 'Variegatus'	EBI	+	-	P	T	M

Supplementary Table 2.1. Continued

No.	Acessions	Location	Awn in Spikelet	Buds at the nodes of culms	Abaxial leaf	Growth habit	Rhizome
79	*M. sinensis* 'White kaskade'	Soyface	+	-	P	T	M
80	*M. sinensis* 'Zebrinus'	EBI	+	-	P	T	M
81	*M. sinensis* 'Zwergzebra'	EBI	+	-	P	T	M
82	*M. sinensis* 'November sunset'	Soyface	+	-	P	T	M
83	*M. sinensis* 'Varigatus'	Soyface	+	-	P	T	M
84	*M. sinensis* 'Zebrinus'	Soyface	+	-	P	T	M
85	*M. sinensis* 'Zwergzebra'	Soyface	+	-	P	T	M
86	*M. sinensiss* 'Amur silver grass'*	GH	+	-	P	T	M
87	*M. sinensis* 'Silver banner grass'*	GH	+	-	P	T	M
88	*Miscanthus* 'Purpurascens'*	EBI	+	-	G	T	I
89	*Miscanthus* 'Purpurascens'*	Soyface	+	-	G	T	I
90	*M. sinensis. var. condensatus* 'Cabaret'	EBI	+	-	G	T	M
91	*M. sinensis var. condensatus* 'Condensatus'	Soyface	+	-	P	T	M
92	*M. sinensis var. condensatus* 'Cosmopolitan'	EBI	+	-	P	T	M
93	*M. transmorrisonensis* 'Evergreen maiden grass'	EBI	+	-	P	T	M
94	*M. x giganteus* (*M. floridulus*)*	Soyface	-	+	G	I	S
95	*M. x giganteus* 'Arabesque'*	Soyface	-	+	G	I	S
96	*M. x giganteus* 'Autumn light'*	Soyface	-	+	G	I	S
97	*M. x giganteus* 'Malepartus'*	Soyface	-	+	G	I	S
98	*M. x. giganteus* 'Cleveland giganteus'	EBI	-	+	G	I	S
99	*M. x. giganteus* 'G-Max'	EBI	-	+	G	I	S
100	*M. x. giganteus* 'Illinois giganteus'	EBI	-	+	G	I	S
101	*M. x.* giganteus 'Ohio giganteus'	EBI	-	+	G	I	S

CHAPTER 3

A framework genetic map for *Miscanthus sinensis* from SSR and RNAseq-based markers shows recent tetraploidy[1]

3.1. Abstract

Miscanthus (subtribe Saccharinae, tribe Andropogoneae, family Poaceae) is a genus of temperate perennial C_4 grasses whose high biomass production makes it, along with its close relatives sugarcane and sorghum, attractive as a biofuel feedstock. The basic chromosome number of *Miscanthus* (x = 19) is different from that of other Saccharinae and approximately twice that of the related *Sorghum bicolor* (x = 10), suggesting large-scale duplications may have occurred in recent ancestors of *Miscanthus*. Owing to the complexity of the *Miscanthus* genome and the complications of self-incompatibility, a complete genetic map with a high density of markers has not yet been developed.

Deep transcriptome sequencing (RNAseq) from two *M. sinensis* accessions was used to define 1536 single nucleotide variants (SNVs) for a GoldenGate™ genotyping array, and found that simple sequence repeat (SSR) markers defined in sugarcane are often informative in *M. sinensis*. A total of 658 SNP and 210 SSR markers were validated via segregation in a full sibling F_1 mapping population. Using 221 progeny from this mapping population, a genetic map for *M. sinensis* was constructed that resolves into 19 linkage groups, the haploid chromosome number expected from cytological evidence. Comparative genomic analysis documents a genome-wide duplication in *Miscanthus* relative to *S. bicolor*, with subsequent insertional fusion of a pair of chromosomes. The utility of the map is confirmed by the identification of two

[1]Reprinted, with permission, from K. Swaminathan, W. Chae, T. Mitros *et al.* 2012, "A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy" BMC Genomics 13:142.

paralogous C_4-pyruvate, phosphate dikinase (C_4-PPDK) loci in *Miscanthus*, at positions syntenic to the single orthologous gene in Sorghum.

The genus *Miscanthus* experienced an ancestral tetraploidy and chromosome fusion prior to its diversification, but after its divergence from the closely related sugarcane clade. The recent timing of this tetraploidy complicates discovery and mapping of genetic markers for *Miscanthus* species, since alleles and fixed differences between paralogs are comparable. These difficulties can be overcome by careful analysis of segregation patterns in a mapping population and genotyping of doubled haploids. The genetic map for *Miscanthus* will be useful in biological discovery and breeding efforts to improve this emerging biofuel crop, and also provide a valuable resource for understanding genomic responses to tetraploidy and chromosome fusion.

3.2. Introduction

The grass subtribe Saccharinae (sugarcanes, sorghums, *Miscanthus*, and related C_4 species) includes a remarkable array of recently and independently derived polyploids that arose from a common diploid progenitor. For example, sugarcanes carry even multiples of a haploid complement of x = 10 or x = 8 chromosomes, and exhibit polysomic inheritance that presumably arose via auto-polyploidy (D'Hont *et al.*, 1998; Sreenivasan *et al.*, 1987; Brandes E, 1956) over the past several million years. This scenario is consistent with the similar monoploid DNA content of sugarcane (approximately 750 million base pairs (Mbp) for *Saccharum spontaneum* and 930 Mbp for *S. officinarum* (D'Hont & Glaszmann, 2001) and 730 Mbp for *Sorghum bicolor* (Paterson *et al.*, 2009). The ten chromosome pairs of diploid *Sorghum bicolor* likely represents the ancestral Saccharinae condition. Polyploidy in Saccharum arose at least twice, and

[1]This chapter was the output of a team-based project. My responsibilities were to generate, transplant and maintain the mapping population, prepare DNA samples from this population, to develop SSR markers, genotype the mapping population with SSR markers and map C4-PPDK alleles in the genome. I was also involved in genotyping data analyses and genetic map construction.

chromosome number in sugarcane is so flexible as to allow a range of natural and artificial auto- and allo-polyploids up to dodecaploid.

In contrast, the genus *Miscanthus* has a base chromosome number of x = 19, with nominally diploid (2 N = 2x = 38) and tetraploid (2 N = 4x = 76) species, plus the highly productive triploid interspecific hybrid, *Miscanthus x giganteus*. Among a number of possibilities for the distinctive chromosome number, the most likely is the whole genome duplication (tetraploidization) of an ancestor possessing N = 10 pairs of chromosomes (Paterson *et al*., 2010) although this has not been demonstrated. Direct comparisons of the DNA content of *Miscanthus* to sorghum and sugarcane is not obviously informative, as the N = 19 monoploid DNA content of *Miscanthus* spans 2150–2650 Mbp (Rayburn *et al*., 2009), more than three times longer than the monoploid content of eusorghum (745–818 Mbp) (Price *et al*, 2005). The possible origin of the nearly doubled chromosome number and tripled haploid size via polyploidy is further obscured by the high repetitive content of the *Miscanthus* genome, recently shown by sample sequencing to be ~95% in *M. x giganteus* (Swaminathan *et al*., 2010).

Chromosome numbers can be unreliable indicators of even relatively recent polyploidy. For example, 2n = 20 maize is a paleopolyploid comprising two sub-genomes that diverged ~12 Mya (Swigonová *et al*., 2004). Comparative mapping and sequence analysis reveals that the progenitors of these sub-genomes also had 2n = 20, a fact obscured karyotypically by subsequent chromosome fusions in the maize lineage. Conversely, while diploid *S.bicolor* has 10 pairs of chromosomes, other diploid Sorghum species with comparable DNA content have only five pairs, presumably a consequence of chromosomal fusions (Price *et al*., 2005). Similarly, diploid *Brachypodium distachyon* has 2n = 10 chromosomes, but other *Brachypodium* species with comparable DNA content have 2n = 20 (Draper *et al*., 2001). In any event, even in a whole-

genome duplication scenario, the odd base chromosome number of *Miscanthus* would require additional chromosome-scale events such as loss or fusion. The description of *M. sinensis* as "diploid" with 2n = 2x = 38 chromosomes is based on chromosome counting, and the observations that chromosome pairing during meiosis regularly produces bivalents (Burner, 1991; Lafferty and Lelley, 1994).

Despite *Miscanthus*' unusual chromosome and DNA complement relative to other Saccharinae, relatively few genetic resources have been developed for elucidating the relationship of the *Miscanthus* genome to those of its close relatives. This is in part due to the fact that the most widely grown *Miscanthus* biomass crop is the vegetatively propagated triploid *M. x giganteus* (2n = 3x = 57), which produces no viable seed (Lindle-Laursen, 1993), and therefore no segregating progeny. *M. x giganteus* is among the most productive in known grasses (Heaton *et al*., 2008) and evidence to date indicates it derives from a cross between a diploid *M. sinensis* and a tetraploid *M. sacchariflorus* (Hodkinson *et al*., 2002b). Another complicating factor is self-incompatibility, which makes the production of homozygous genotypes difficult and forces the independent mapping of meiotic products from each parent in F_1 progeny.

M. sinensis, the likely diploid parent of *M. x giganteus*, is widely grown as an ornamental grass with rich genetic diversity, and is itself highly productive. Although a preliminary genetic linkage map for *M. sinensis* using RAPD markers and an "offspring cross" mapping strategy has been published (Atienza *et al*., 2002), this map resolves 28 linkage groups (LGs), many more than the expected 19 LGs. The marker density of the map is not sufficient for fine-scale mapping and the reproducibility of RAPD markers is difficult. These problems can be mitigated by utilization of simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers, which are plentiful in the *Miscanthus* genome and are also reproducible across

laboratories. Additionally, SSR markers can be used for the search of homoeologous chromosomes in the mapping of polyploid plants (Ming *et al*., 1998; Okada *et al*., 2010).

The present study reports the discovery of genetic variation in *M. sinensis* using both SNP markers discovered by deep transcriptome sequencing and SSRs that were previously shown to be variable in sugarcane. A dense map of all 19 linkage groups was obtained in *M. sinensis* with 846 segregating markers. Comparison with the *S. bicolor* genome reveals whole-genome duplication in *Miscanthus*, with a single chromosome fusion accounting for the odd base chromosome number of the genus. The two sub-genomes of *Miscanthus* are quite similar, resulting in variant frequencies among paralogs that are only modestly higher than those observed between alleles. Despite this recent duplication, whether by allo- or auto-tetraploidy, our map is consistent with disomic inheritance in *Miscanthus*, in contrast to the polysomic inheritance found in the closely related polyploid sugarcane. The genetic map of *Miscanthus* provides a valuable resource that can be used to apply both functional genomics to this perennial C_4 grass, and marker-assisted breeding to biomass crop improvement.

3.3. Materials and Methods

Grosse Fontaine x Undine reciprocal mapping population

A full sib (F_1) population was produced by reciprocally crossing two ornamental *M. sinensis* accessions, ‘Grosse Fontaine’ (GF) and ‘Undine’ (UN) propagated in the greenhouse, from rhizomes of single plants established at the SoyFACE plot (Figure 3.1A and 3.1B; plant locations SF20 and SF5 respectively) located at the Crop Sciences Research and Education center, on the University of Illinois at Urbana-Champaign (UIUC). The temperature of the greenhouse was maintained between 22.2–29.4°C and supplemental light (threshold of 600

W/m2) was provided from 6am to 8pm. These accessions are variable for a number of phenotypes including plant height, stem diameter and flowering time which weretaken from mature plants growing in the greenhouse (Figure 3.1C). Parent plants were isolated in a greenhouse room and reciprocally cross-pollinated to produce seeds from both parents. Seeds were collected separately for each direction of the cross and germinated in seed trays in the Plant Science Laboratory greenhouse at UIUC. A full -sib population from both reciprocal crosses was grown in the greenhouse. The rhizome from each individual was split and planted in a randomized block design with 3 clonal replicates per plant at the Energy Biosciences Institute (EBI) farm at UIUC in May 2010 (Figure 3.1D).

Genomic DNA of the mapping population and the two parental genotypes was extracted from young leaves using the Puregene protocol (Qiagen, Valencia, California, USA) and used for SSR and SNP marker development and genotyping. After removing individuals that showed non-parental alleles, likely due to pollen contamination, 221 F_1 individuals defined our mapping population, including 113 with GF as maternal parent and 108 with UN as the maternal parent. All plants were genotyped for mapping using SNP and SSR markers as described below.

Transcriptome sequencing and assembly (Dr. Kanshita Swaminathan and Dr. Therese Mitros)

Total RNA was extracted from young leaves from GF and UN (the two parents of the mapping population) using a CTAB RNA extraction method (Chang and Puryear, 1993). Paired-end RNA-seq libraries were made using the Illumina RNAseq kit (cat # RS-930-1001) as per the manufacturer's instructions. The libraries were sequenced at the Keck Center for Functional Genomics at the University of Illinois on an Illumina GA II platform. A total of 144 million 80

bp RNAseq reads were generated from 6 lanes of sequencing, with 5 of the lanes producing successful paired-end reads (found at NCBI short read archive, accession number SRA051293). De novo assembly of the raw RNAseq reads for each parent was performed using ABySS (Birol *et al*., 2009) with k-mer lengths k = 25, 30, 35, 40, 45 and 50 bp. All assemblies were run on fifteen nodes of a cluster (Dual-quad cores (2.83 GHz Xeons), 16 GB RAM). The assemblies were made non-redundant by removing contigs that were identical or completely contained within a larger contig. The resulting contigs from Undine and Grosse Fontaine were then merged using Phrap (Green *et al*., unpublished observations) version 1.080721, -revise_greedy, -minmatch = 20 and -penalty = –9). This combined assembly was used as the reference sequence for the discovery of single nucleotide variants (SNVs). Identification of single nucleotide variations from RNAseq data RNA-seq reads were aligned back to the combined Undine and Grosse Fontaine transcriptome assembly using Bowtie (Langmead *et al* 2009; Langmead 2010) and bwa (Li and Durbin, 2009; Li and Durbin, 2010). Bowtie was run with the -k option set to 1 and with the -best option. Bwa was run with -q 15. The sam output was converted to bam and sorted using view and sort functions from the samtools suite (Li *et al*., 2009).

Duplicate reads were removed using the samtools rmdup function as these could be an artifact of the PCR step during the construction of the RNAseq libraries. The bam file was then converted to pileup format using samtool's pileup function and SNVs were identified computationally using VarScan (Koboldt *et al*., 2009). For the GoldenGate probe set, only SNVs flanked by at least 50 bp of invariant sequence that had a minimum of ten reads corroborating each allele were chosen. There was no tolerance for indels.

To obtain probes appropriate for genotyping with genomic DNA, we screened these 101 bp sequences (the SNVs chosen for the GoldenGate assay plus the 50 invariant bases on

both flanks) using BLAT (Kent, 2002) against the fully assembled genomes of four grasses (sorghum, maize, rice, and *Brachypodium distachyon*) to eliminate sequences that contained splice junctions. Illumina further filtered probes for robustness with respect to the GoldenGate assay.

Single nucleotide variant (SNV) genotyping using the GoldenGate™ and Genome Studio (Dr. Kanshita Swaminathan and Dr. Therese Mitros)

Genomic DNA from the F_1 mapping population and both parents, as well as two doubled haploid M. sinensis (IGR-2011-001 and IGR-2011-002) and their parents (IGR-2011-003 and IGR-2011-004, respectively), were assayed at the Keck Center for Functional Genomics at the University of Illinois using the 1536 SNV GoldenGate array described above, following the manufacturer's protocols. Genotypes were called using Genome Studio (Illumina), which characterizes each genotype according to the signal intensities measured for the alternate nucleotides that define a SNV. Here and below, denote these alternate nucleotides "A" and "B."

For each SNV, Genome Studio clusters signal intensities to define homozygous and heterozygous genotype calls. For a segregating (diploid) SNP, one or two homozygous clusters and one heterozygous cluster are expected, depending on whether or not the SNP is variable in both parents (Figure 3.2A, C and E show clusters observed in biallelic SNPs). In contrast, a SNV that represents a fixed single nucleotide difference between paralogs (i.e., a non-segregating variant at two unlinked loci) is revealed when both parents, and the entire F_1 population, form a single "heterozygous" cluster (data not shown). SNVs that are fixed at one locus but segregating at another form two or three clusters as for a conventional diploid SNP, but with skewed signal intensities (Figure 3.2B, D and F). Each SNV cluster was reviewed manually in Genome Studio.

SNVs forming more than three clusters were discarded. Two doubled haploid lines and their parents were also genotyped, and used to confirm clustering (Figure 3.2).

SSR marker development

Primers for sugarcane SSRs derived from expressed sequence tags (ESTs) and intergenic sequences were previously designed and characterized by James *et al*. (2011). These primers were tested in *M. sinensis* to screen for markers that are polymorphic within one or both parental genotypes, GF and UN. Products were amplified in 10 µl PCR reactions containing 1 µl of genomic DNA (5–10 ng) from GF or UN, 0.1 µl of forward and reverse primers (100 µM stock each), 3.8 µl of ddH2O and 5 µl of 2X GoTaq Green Master Mix (Promega, Madison, Wisconsin, USA). PCR conditions for the screening were as follows: 3 min of denaturation at 94°C, 36 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 45 sec followed by a final extension at 72°C for 10 min. The amplicons were separated on 4% agarose SFR gels (Amresco, Solon, Ohio, USA) with 1X TBE buffer at 4°Cand visualized with ethidium bromide. Polymorphic markers resulting from this screen were used for subsequent genotyping of the *Miscanthus* mapping population (Figure 3.3).

To genotype the mapping population, products were amplified in 10 µl PCR reactions containing 1 µl of genomic DNA (5–10 ng), 0.02 µl of M13 tailed forward primer, 0.1 µl of each reverse and fluorescent M13 primers (100 µM stock), 3.78 µl of ddH2O and 5 µl of 2x GoTaq Colorless Master Mix (Promega, Madison, Wisconsin, USA). Four M13 primers tagged with FAM, VIC, NED and PET at the 5′ end were used in this analysis to fluorescently label the SSR amplicons. All primers were ordered from Integrated DNA Technologies (idtdna.com). Touchdown PCR was used to amplify the SSRs: denaturation at 94°C for 3 min followed by 2

cycles of 94°C for 30 sec, 65°C for 30 sec, and 72°C for 45 sec. The annealing temperature was decreased every 2 cycles by 2°C until 57°C. The amplification was finished with 26 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 45 sec (total 36 cycles) and a final extension at 72°C for 10 min. Electrophoresis of the amplicons was carried out by the Keck Center of functional genomics at the University of Illinois, on an ABI 3730xl with the LIZ600 size markers. Marker scoring was done using the Genemarker software (Softgenetics, LLC State College, Pennsylvania, USA).

Linkage analysis and map construction

The 221 F_1 offspring were genotyped using fragment analyses of 210 amplicons from 107 SSR primers and GoldenGate analysis of 1536 SNVs (Table 3.1). A total of 868 markers showed clear polymorphisms or clustering and were used in the linkage analysis. Genotypes were converted into JoinMap "CP" (cross pollinator) codes based on the genotypes of the parents and the type of marker. The linkage map was constructed using the JoinMap 4.1 software (Van Ooijen, 2011). Thirty-five of the markers displayed a segregation ratio greater than 2:1 when the expected ratio was 1:1, and greater than 10:1 when the expected ratio was 3:1 (Table 3.1). These 35 markers were not considered during the construction of the initial framework map. None of the individuals included had missing data for more than 13 markers. A minimum independence logarithm of odds (LOD) score of 11 and a maximum recombination frequency of 0.4 was used to define linkage groups in all map calculations mentioned here on.

Synteny with *S. bicolor* genome (Dr. Kanshita Swaminathan and Dr. Therese Mitros)

Mapped *Miscanthus* markers were aligned to the *S. bicolor* genome using blastn 2.2.25+ (Camacho *et al*., 2009) with wordsize 10 and BLAT (default parameters) (Kent, 2002). From

these two alignments the SNP markers were assigned to a position in sorghum if they had the largest number of identical residues and shared at least 80% of the residues in the probe. The positions of markers in centiMorgans on the 19 *Miscanthus* linkage groups were plotted versus these aligned positions to the sorghum genome coordinates.

Comparison to sorghum genetic map (Dr. Kanshita Swaminathan and Dr. Therese Mitros)

The consensus map for *S. bicolor* developed in Mace *et al*. (2009) was adopted. Sequence-tagged markers were extracted from supplemental materials of this paper and Genbank, and aligned to the chromosome sequences of sorghum (Paterson *et al*., 2009). Sorghum map positions for our *Miscanthus* markers were then inferred by linear interpolation using flanking markers from the sorghum map, assuming locally constant recombination rates.

Sequence of the 3rd intron of *Miscanthus* C_4-PPDK (Liang Xie)

PPDK sequences in Genbank (accession numbers AY262272.1, AY262273.1), and the GF and UN RNAseq sequences and assemblies were aligned to the genomic PPDK locus on sorghum chromosome 9. Primers PPDK-int3F and PPDK-int3R (5′-AACCTGGCGGAGATGT CGA-3′ and 5′-AGGTAGACTTCCTTGTACTGA-3′, respectively) were designed to amplify the third intron of C_4-PPDK from both Undine and Grosse Fontaine. The primers amplified two fragments, between 1500 and 2000 bp, from each parent. Each amplicon was cloned separately into pGEM-t easy (Promega) and a total of 45 clones (10 to 15 clones from each band) were Sanger sequenced using three oligonucleotide primers, (SP6, T7 and 5′-GAGACAGCGATTGGACTAAGC-3′). The sequences were aligned using the Sequencher sequence analysis software (Gene Codes Corporation, Ann Arbor, MI USA).

Phylogenetic analysis of intron sequences (Dr. Therese Mitros)

Intron and flanking exon sequences from primer to primer were aligned with Muscle (Edgar, 2004) and trimmed to remove ambiguous sites. Orthologous introns from S. bicolor, S. officinarum, and Z. mays were identified from sequences in Genbank. For the purposes of phylogenetic analysis, identical sequences were removed. Gblocks (Castresana, 2000) was used to identify blocks of well-aligned sequence with a minimum of 6 sequences for a conserved position, 8 for a flanking position, 8 as the maximum number of contiguous non-conserved positions, half allowed gap positions, and a minimum block size of 5. The final alignment had 1,337 positions. MrBayes (Huelsenbeck & Ronquist, 2001) was used to produce a consensus phylogenetic tree (50,000 generations, with sampling frequency, 100), using an inverted gamma distribution for rate variation. Midpoint rooting was used.

Mapping *Miscanthus* C_4-PPDK loci

The G/A polymorphism at position 397 in the sequence alignment was used as a CAPS marker [marker identifier EBI-847] as this polymorphism results in the presence of an NheI restriction enzyme site (5′-GCTAGC-3′). NheI (NEB # R0131S) was used to digest amplicons obtained from PPDK-int3F and PPDKint3R in the parents and population. The population was scored for the presence of one or two bands as marker EBI 847. A second SSLP marker (EBI-848) was designed around two indels between positions 1354 and 1388. Oligos PPDK-UD3F and PPDK-UD3R (5′-AAAGGTGAACATAGTTTCG-3′ and 5′-CATAGTTCG(T/A)AGCGTGAG-3′ respectively), were designed around these indels and used to amplify the locus from the population and the parents. The plants either amplified a single

fragment (132 bp) or amplified two fragments (132 bp and 118 bp). The 118 bp amplicon segregated in the population and was scored as EBI 848.

3.4. Results and Discussion

RNAseq and the genomic sequence of related species can be used to define SNVs

To develop a collection of putative SNVs for *Miscanthus*, we sequenced transcriptomes of M. sinensis 'Grosse Fontaine' and 'Undine' leaves and leaf rolls using deep RNAseq. Across both accessions, we generated over 21 Gbp in predominantly paired 80 bp Illumina GA II reads. From these RNAseq data we assembled a unified set of 29,933 contigs longer than 100 bp. The median contig length was 522 bp, with half of the total contig length accounted for by 6,433 contigs longer than 1,071 bp (the contig N50). We identified SNVs by realigning the RNAseq reads against the assembled transcriptome contigs and requiring strong support for two alternate variants embedded in otherwise nearly identical flanking sequence, to enable straightforward high-throughput genotyping. Other variation observable in the dataset was not considered further.

Since our aim was to define variants that could be genotyped by a GoldenGate assay with genomic rather that transcriptomic samples, we excluded from consideration probe sequences that spanned a putative exon-exon boundary. To do this in the absence of a *Miscanthus* genomic reference, we took advantage of the extensive conservation of exon-exon boundaries in grasses (Paterson *et al.*, 2009) to identify and reject likely exon-junction-spanning probe sequences by comparison with the genomes of sorghum, maize, and rice. To facilitate syntenic comparisons between *Miscanthus* and related species, we also chose for genotyping those SNVs that (1) could be readily assigned to homologs in sorghum by sequence similarity and (2) had homologs that were distributed across all sorghum chromosomes.

Results of GoldenGate genotyping

Out of 1,536 putative markers on the *Miscanthus* GoldenGate array, 1,243 showed one or more clusters in GoldenGate signal space (Figure 3.2), indicating consistent genotyping across individuals. The remaining 293 putative markers showed dispersed or very low signal in Genome Studio and were considered failed assays, and not investigated further. Of the 1,243 successful oligonucleotide assays, 93 assays showed signal for only one probe, and appear to be homozygous across both parents and their progeny or represent cases where the second oligo probe failed. After excluding these failed or invariant assays, 1,150 markers were left, of which 658 formed 2 or 3 clusters in signal space. The remaining markers appear as either a single centrally located cluster, more than three clusters, or dispersed signal, and were not considered further.

Intepretation of GoldenGate SNP genotypes

By considering the patterns of genotypes across our F_1 mapping population, many of the SNV's discovered by RNAseq analysis are indeed segregating biallelic markers (i.e., single nucleotide polymorphisms, or SNPs). Others, however, represent fixed differences between closely related paralogous loci. Furthermore, many segregating biallelic markers have their GoldenGate signal affected by a closely related paralog that has the same sequence as the marker allele. Signal from such paralogous alleles causes the cluster positions in Genome Studio to be skewed in a characteristic manner that is readily recognized. A plot of normalized theta (ratio of signal intensities assayed for A and B SNP alleles) against normalized R (signal intensity) per marker for each individual can be used to visualize genotypes in a segregating population (Figure 3.2A–

F). The values of normalized theta are close to 0 in samples where the genotype is AA, close to 0.5 if it is AB and close to 1 if it is BB.

In situations where more than one locus is being sampled, and where the sequence of a second (paralogous) locus matches one of the two allelic states of the SNV in the segregating locus, the clusters are skewed towards the allele sharing the common nucleotide (Figure 3.2B, C and D). In Figure 3.2B, locus 1 is heterozygous for A and B SNVs in both parents and hence produces AA, AB, or BB progeny, whereas the second paralogous locus is fixed for the B SNV in both parents and progeny. This results in all three clusters being skewed to the right due to the higher dosage of SNV B. Figure 3.2D shows a scenario where the GF parent is AB and the Undine parent BB at locus 1, whereas the second locus is fixed for SNV A in both parents and progeny, which shifts clusters to the left due to higher dosage of SNV A. A similar situation is shown in Figure 3.2F where UN rather than GF is segregating at locus 1.

For mapping of segregating loci, panels A and B indicate markers that are heterozygous in both GF and UN parents, panels C and D show markers heterozygous in only the GF parent, and panels E and F markers heterozygous in only the Undine parent. Markers shown in Figure 3.2 E and F share the feature where the genotype of the two different sampled doubled haploid lines carry either the A or B SNV, but no progeny share the B/B genotype because their parents have either an A/A (GF) or A/B (UN) genotype. Notably, 26% of the two-cluster SNV's showed skewed signal intensities in the GoldenGate assay, indicating that the two alternative sequences are not present in equal dosages. This observation is consistent with the sequence variants being detected from more than one locus, and suggests that many of the variant sequence pairs A and B appear as heterozygous alleles at one locus (A/B) but are fixed at a second locus (i.e., A/A or B/B), resulting in a ~3:1 ratio of signal intensities on the GoldenGate assay. If both parents show

allelic variation at one locus but are fixed for the same allele at a second paralogous locus, then segregating progeny may show 2:2, 3:1, and 4:0 dosages, consistent with observations (Figure 3.2A. EBI 832, EBI 693 and EBI 635).

A second class of SNV (33%) formed only a single cluster of genotypes (data not shown). For these SNVs, both parents and all progeny had the same genotype. This is consistent with the pattern expected from fixed differences between paralogous loci (e.g., A/A at one locus and B/B at another) that do not segregate in progeny. These SNV's are not useful as genetic markers, since both parents and all progeny fall into a single "heterozygous" cluster and there is no genetic segregation of alleles. The proportion of both single cluster and skewed two cluster SNVs (59%) should not be used as a direct estimate of the degree of paralogy due to the potential biases introduced by our SNV discovery and selection. These paralogous loci, however, do suggest extensive paralogy in the *Miscanthus* genome, which is corroborated by the genetic map as shown below.

Only a small minority (5 out of 1536) of the SNVs that we identified by RNAseq analysis formed more than three clusters in signal space, and could not be simply interpreted either as segregating alleles or fixed paralogous variants. The rarity of such SNV's in this analysis suggests that a similar RNAseq-based protocol could be useful in SNP discovery from other *Miscanthus* populations and species lacking genomic reference sequences. For 658 out of 1,150 genotyped *Miscanthus* SNVs, the GoldenGate intensities in our F_1 mapping population could be grouped into two (467) or three (191) clusters of genotypes in signal space, indicating variants that are found in both homozygous and heterozygous states in the population. The two-cluster class of SNV's can be interpreted as segregating SNPs that are heterozygous in one parent and homozygous in the other, with progeny of both types.

Similarly, the three-cluster classes of SNVs are interpreted as SNPs that are heterozygous in both parents, allowing for homozygous offspring of two types as well as heterozygotes. The interpretation of these SNV as segregating SNPs in our cross is supported by the integration of these markers into a consistent linkage map with limited segregation distortion (below).

Corroboration of allelic and fixed differences using doubled haploid lines

To test the hypothesis that many SNV's represent fixed differences between paralogous loci, two *M. sinensis* double haploid lines and their parents were also genotyped. Since the doubled haploids were developed by another culture from outbred diploid parents (Glowacka, unpublished observations), there are two expectations.

First, for the SNV's that are inferred to be biallelic SNPs in F_1 cross, some of them will correspond to heterozygous loci in other *M. sinensis* accessions, including the outbred parents of the doubled haploid lines. If these SNV's are bona fide allelic variants, however, then the doubled haploids should be homozygous for all such variants. Figure 3.2G shows the segregation of alleles in the GoldenGate assay. In situations where two or three clusters are observed in the GoldenGate, consistent with a biallelic SNP, the double haploids are either A/A or B/B homozygotes while the mapping population has all three allelic states, as expected.

Second, for SNV's that are inferred to be fixed differences between paralogs, both variant states should be observed in the doubled haploids as well as their parents. This is observed as a single AB cluster on the GoldenGate array (Figure 3.2G).

Genotyping summary

Taken together, analyses of the F_1 mapping population and the two doubled haploid lines show that segregating allelic variants at a single locus can be distinguished from fixed differences between paralogs, even in the face of extensive gene duplication. These data suggest that many *Miscanthus* genes have a closely related paralog that cannot be easily differentiated in the short read transcript data, but which assort independently. Using segregation patterns from a high density of genetic markers a linkage map can be constructed.

SSR primers from sugarcane identify allelic and paralogous polymorphism in *Miscanthus*

Since *Saccharum* (sugarcane) is a close relative of *Miscanthus*, primer pairs that amplify simple sequence repeats in *Saccharum* would also be likely to amplify polymorphic SSRs in *Miscanthus* (Cordeiro *et al*., 2001). Sixty-eight percent of the 2,640 SSRs primer pairs mined from sugarcane ESTs produced amplicons when tested with *Miscanthus*. Only 51% of the 2,628 SSR primer pairs derived from *Saccharum* genomic sequences produced amplicons with *Miscanthus*. Of these, 188 EST- and 237 genome-derived primers generated polymorphic amplicons between the two parental genotypes. Primers that produced non-specific amplicons were excluded. We genotyped the F_1 mapping population using 107 primers pairs (29 and 78 primers from EST and intergenic sequences, respectively) out of 425 polymorphic primers. One hundred and seven primers produced 20 marker configurations (Figure 3.3). Among them, 69 primers follow disomic marker configurations but 38 primers (35.5%) do not fit disomic configurations, producing more than 3 amplicons in one or both parents. One hundred and seven primers produced a total of 301 amplicons and among them, 210 were polymorphic between two

parental genotypes and segregated in progeny populations. One hundred ninety three amplicons out of 210 were actually mapped.

An integrated linkage map for *M. sinensis*

Using the 868 segregating markers defined above, we constructed an integrated linkage map for *M. sinensis* using JoinMap 4.1. A newly implemented multipoint maximum likelihood model was adopted to construct a map from an F_1 cross of two outbred parents, using the Haldane mapping function (Van Ooijen, 2011). In contrast to a pseudo-testcross approach, which utilizes markers that are heterozygous in one parent but homozygous in the other, the new method can also incorporate markers that are heterozygous in both parents. While pseudo-testcross based analysis results in separate maps for each parent, the combined approach allows direct integration into a single map of crossovers that occur in either or both parents by using the markers that are heterozygous in both parents as anchors.

Only 48 out of 868 markers show segregation distortion ($p <= 0.005$ using the chi-squared goodness of fit test). Of these, 35 were highly distorted markers, and not included in the initial framework map. These highly distorted markers include 21 with a segregation ratio greater than 2:1 when they should have been 1:1, and 14 with a segregation ratio greater than 10:1 when they should have been 3:1 (Table 1). Of the remaining 833 segregating markers (641 SNPs and 192 SSRs), 829 were incorporated into 19 major linkage groups using a minimum logarithm of odds (LOD) score of 11 and maximum recombination frequency of 0.4. The JoinMap 4.1 maximum likelihood method was used to calculate the map order of the framework map. Four SNP markers that were placed more than 40 cM away from the nearest marker on the linkage group were excluded and marker order for those linkage groups was recalculated. An

attempt was then made to replace the 35 highly distorted markers on the ML map, keeping the marker order of the framework map constant and using the same map calculation parameters as before. Seventeen of the 35 highly distorted markers were incorporated. This map with 846 markers is shown in Figure 3.4.

Independent regression maps for each parent were also constructed to corroborate the robustness of marker order. The total length of the 19 linkage groups on the ML map is 1782 cM, with an average intermarker spacing of 2.7 cM (excluding markers with identical map positions). Thus we expect that the missing map length from the telomeric ends of the linkage groups (Knapik *et al*., 1996; Chakravarti *et al*., 1991) accounts for roughly 2 × 19 × 2.7 cM = 102 cM, for a total estimated map length of 1884 cM. In the Grosse Fontaine map, 94% of the markers lie within 10 cM of each other, while in the Undine map only 90% meet this criterion. In the integrated map, 97% of the mapped markers lie within 10 cM of another marker, attesting to the dense coverage of the map.

Disomic inheritance and limited segregation distortion

Transmission of each linkage group is consistent with pure disomic inheritance in *M. sinensis* (i.e., complete preferential pairing of homologs), with no evidence for tetrasomic inheritance (i.e., pairing and recombination between homoeologs). Furthermore, very few markers show segregation distortion (48 out of 868; $p <= 0.005$ using the chi-squared goodness of fit test), and those that do are concentrated on Ms2, Ms3, Ms4, Ms12, and Ms13. Overall there is more segregation distortion in Undine. Twenty of the 24 distorted UN markers lie on Ms4. Potential causes of segregation distortion include the following three possibilities: (1) Failure to complement deleterious recessive alleles heterozygous in both GF and UN parents that

reduce viability of F_1 progeny; (2) Interactions between genomes, e.g., meiotic drive in F_1 gametophytes, gametophytic competition or pollen-pistil interactions like self-incompatibility; (3) Proximity to areas of suppressed recombination like centromeres and nucleolus organizer regions. The design of our cross makes it difficult to differentiate among these possible explanations.

Whole genome duplication with extensive conserved synteny to sorghum

Since our *Miscanthus* markers were derived from (1) transcribed regions with reduced sequence variation (SNPs) and (2) sequences from conserved ESTs and intergenic regions (SSRs), many of them could be unambiguously assigned to orthologous (i.e., evolutionarily homologous) positions on the Sorghum bicolor genome sequence by straightforward sequence alignment. Out of 653 SNP loci on the integrated *Miscanthus* map, 618 could be placed on the sorghum genome. Similarly, out of 193 SSRs on the map, 126 could be placed on the sorghum genome.

A simple dot plot (Figure 3.5A) strikingly reveals complete whole-genome duplication in *M. sinensis* relative to sorghum, with most chromosomes showing near perfect colinearity at the scale of our genetic map. After recognizing this extensive synteny, we oriented and renumbered the *Miscanthus* linkage groups to emphasize this correspondence between *Miscanthus* and sorghum. Every sorghum chromosome exhibits nearly complete marker synteny with a pair of *Miscanthus* linkage groups. Eight sorghum chromosomes are completely duplicated, showing a 1:2 correspondence to *Miscanthus* linkage groups. We infer the whole genome nature of the duplication by the density of colinear markers in sorghum euchromatin, where our gene-biased markers are found. The only evident rearrangement in these

chromosomes is a small inversion near the top of Sb4 relative to Ms8 and Ms9. Since Ms8 and Ms9 share the same ordering in this region, this inversion either occurred in the sorghum lineage, or in the stem lineage of *Miscanthus* prior to the tetraploidization event, or is an error in the sorghum genetic map or sequence assembly.

The remaining two sorghum chromosomes, Sb4 and Sb7, are also duplicated over their entire euchromatic spans, but show a more complex pattern of synteny with *Miscanthus*. Ms8 is an intact copy of Sb4, and Ms13 is an intact copy of Sb7. The second copies of these two sorghum chromosomes, however, are fused into the single linkage group Ms7. Ms7 then appears as a copy of Sb7 inserted into the centromeric region of Sb4 (Figure 3.5B). This single fusion explains the odd base chromosome number of *Miscanthus*. By following the relative orientations of sorghum chromosome arms in *Miscanthus*, we see that this fusion has the characteristic form of a type of insertion previously observed in other grasses (Luo *et al*., 2009). Since all *Miscanthus* species have the same base chromosome number, this fusion presumably occurred in the lineage leading to the last common *Miscanthus* ancestor.

Mapping C_4-PPDK loci in *Miscanthus*

C_4 photosynthesis in the Panicoideae (including maize, Saccharinae, millet, switchgrass, and *Miscanthus*) is facilitated by a C_4-specific form of the pyruvate, phosphate dikinase enzyme (C_4-PPDK). Physiological and molecular evidence suggest that altered expression of C_4-PPDK may contribute to cold tolerant C_4 photosynthesis in *M. x giganteus* (Naidu *et al*, 2003; Wang *et al*., 2009).

The closely related *S. bicolor* has a single C_4-PPDK gene located on chromosome 9 (Paterson *et al*., 2009). Sequencing of cloned cDNAs from triploid *M. x giganteus* identified five

distinct transcripts, including one apparent pseudogene (Naidu *et al*, 2003), which suggests even greater genetic complexity than three homoeologous C_4-PPDK alleles. Based on our observation of whole genome duplication, we reasoned that *M. sinensis* might have an unlinked pair of paralogous C_4-PPDK genes. Based on synteny considerations, these C_4-PPDKs would lie on *Miscanthus* LG's, 16 and 17, both of which are syntenic to Sorghum 9.

To look for C_4-PPDK paralogs in *M. sinensis*, and to identify the genetic map position or positions of the gene(s), we designed primers to amplify the third intron of the gene based on the single previously known *Miscanthus* C_4-PPDK. Two amplicons were observed in both parents, and both were cloned and sequenced. Cladistic methods identified two distinct paralogs of C_4-PPDK (Figure 3.6A), which we named C_4-PPDK1 and C_4-PPDK2.

By aligning partial sequences of C_4-PPDK in *M. sinensis* with the homologous sequence in *S. bicolor*, *S. officinarum*, and *Z. mays*, we measured the sequence divergence and phylogenetic relationship between the two *Miscanthus* homoeologs and homologous sequences in related outgroups (Figure 3.6A). The divergences between Ms C_4-PPDK1 and sorghum and sugarcane C_4-PPDK are comparable, suggesting that the origin of *Miscanthus* could be contemporaneous with the split between sorghum and sugarcane. Ms C_4-PPDK2 branches outside of the Ms C_4-PPDK1/sorghum/sugarcane clade, which could indicate that the other parent involved in *Miscanthus* tetraploidy was more divergent. These inferences, however, are weak due to the limited sequence length used in the analysis.

To map the two evident paralogs of C_4-PPDK, we designed markers for each gene based on observed intronic sequence variation. Marker EBI 847 is a Cleaved Amplified Polymorphic Sequence (CAPS) marker designed to detect the SNV at position 397 and marker EBI 848 is a sequence length polymorphism (SLP) marker that detects two indels between 1354 bp and 1388

bp. Both markers show a 1:1 segregation ratio. EBI 847 maps to *Miscanthus* linkage group Ms16 at 36.8 cM on the integrated map (41.2 cM on the GF maximum likelihood map) while EBI 848 is placed on linkage group Ms17 at 19.2 cM on the integrated map (20.1 cM on the UN maximum likelihood map). *Miscanthus* linkage groups 16 (C_4-PPDK1) and 17 (C_4-PPDK2) are the homoeologs of *S. bicolor* chromosome 9, which contain sorghum C_4-PPDK (Figure 3.6D). This demonstrates both the utility of our genetic map and sorghum synteny for mapping genes in *Miscanthus*. This is the first documentation of the presence of two paralogous (indeed, homoeologous) C_4-PPDKs in *Miscanthus*. The presence of two paralogs provides an opportunity for regulatory divergence and could contribute to the ability of *Miscanthus* to perform cold tolerate photosynthesis.

3.5. Conclusions

All grasses are paleopolyploid by virtue of an ancient whole genome duplication that occurred ~70 million years ago (mya) in a common ancestor of extant Poaceae (Thiel *et al*, 2009; Salse *et al*, 2008; Bolot *et al*., 2009). Many lineages within the grasses have also experienced more recent polyploidization events superimposed on this early event. This study has shown that *M. sinensis* is a recent polyploid. Through comparative analysis of our *M. sinensis* genetic map with the *S. bicolor* genome, we account for the base chromosome number x = 19 of the genus *Miscanthus* by a doubling of the ancestral Sacccharinae number x = 10, and a subsequent chromosome fusion. Some taxonomists have included in the *Miscanthus* genus several African accessions that have a base chromosome number of x = 15 (Amalraj & Balasundaram 2006; Hodkinson *et al*. 1997; 2002b) and Himalayan accessions where 2n = 40 (Amalraj and

Balasundaram 2006). These may represent ancestral configurations (e.g., 2n =40), additional karyotypic changes (x = 15), or misclassifications.

Since most common *Miscanthus* species (*M. sinensis*, *M. sacchariflorus*, *M. lutarioriparia*, and *M. floridulus*) share the base chromosome number 19, both the genome duplication event and the chromosome fusion likely occurred within the last several million years, at or near the base of the Saccharinae. Although we cannot rule out recurrent polyploidizations in the lineages of multiple *Miscanthus* species, a single origin is most parsimonious. Our *M. sinensis* map is consistent with disomic inheritance, without pairing of homoeologous chromosomes despite their limited sequence divergence. The situation in *Miscanthus* is similar to that found in hexaploid wheat, where closely related species hybridized in allopolyploid fashion, retaining their original chromosomal pairing patterns in a larger genome. Tetraploidization provides the opportunity for a lineage to explore the regulatory and functional diversification of duplicated genes (Otto, 2007; Adams & Wendel, 2005; Adams *et al*., 2004; Feldman & Levy, 2009).

Remarkably, when measured in map units, the *M. sinensis* and *S. bicolor* genetic maps are linearly related, indicating that the inserted repetitive sequence in the *Miscanthus* genome is not recombinogenic. The total length of the 19 linkage groups of our *M. sinensis* map (~1890 cM) is comparable to the map length of the 10 linkage groups in the S. bicolor genome (~1605 cM, Mace *et al*., 2009). Naively, the doubling of chromosome number would be expected to substantially increase the total map length, based on the rule of thumb that each chromosome arm experiences approximately one crossover per meiosis. This suggests that the *Miscanthus* duplication is recent enough that whatever cellular mechanism is responsible for regulating crossover frequency has not had time to adjust to the new karyotype.

It is tempting to speculate that the ensuing chromosome fusion was a critical evolutionary event that established the genus through reproductive isolation of the nascent *Miscanthus* population. Subsequent radiation could then have produced numerous *Miscanthus* species, some of which (e.g., tetraploid *M. sacchariflorus*) underwent additional polyploidization. The nature of these later events is unknown. The ancestral chromosome fusion itself can be understood as arising from the insertion of one chromosome into the centromeric region of another (Figure 3.7). Insertional fusions have been inferred in other grasses (Cordeiro *et al*., 2001), and an insertion with the same orientational properties that we observed has been described for *Aegilops tauschii* chromosome 4D.

The recent and extensive nature of the *Miscanthus* genome duplication, coupled with our use of RNA-seq to discover single nucleotide variant markers, required a careful analysis of segregation patterns in our F_1 mapping population to extract bona fide allelic polymorphisms from a background of comparable sequence variation that arises from fixed differences between paralogous (and nominally homoeologous) loci. Given the large genome size of *Miscanthus*, deep RNA-seq was an efficient and cost-effective way to identify many single nucleotide variants. Our integration of the resulting single nucleotide polymorphism markers with simple sequence repeat markers confirms the validity of this approach. A new maximum likelihood method for full sib mapping (Van Ooijen, 2011) allows the integration of parental maps. These methods may be useful for rapidly developing markers and maps for other species with complex ploidy.

Since our *M. sinensis* genetic map has good coverage of all 19 linkage groups and shows limited segregation distortion that is clustered in three regions, it will be useful for further exploration of the *Miscanthus* genome. As a first step in this direction, the genetic map and

knowledge that *Miscanthus* is recently duplicated relative to sorghum was used to discover and map two homoeologous copies of the C_4 pyruvate, phosphate dikinase enzyme (C_4-PPDK), which appears at the expected syntenic position relative to sorghum C_4-PPDK. Whether or not the two C_4-PPDK genes have distinct roles is unknown.

The ability to separate homoelogous loci suggests that this map could be valuable for both identifying quantitative trait loci in *Miscanthus*, and for marker-assisted breeding improvement of this emerging bioenergy crop.

Table 3.1. Summary of SSR and SNP Marker, as published in Swaminathan *et al*. (2012).

Type of Marker	Number of primer pairs/ SNPs	Number of amplicons	Polymorphic amplicons	Markers polymorphic in Undine		Markers polymorphic in Grosse Fontaine		Markers polymorphic in both parents		
				Segregation ratio						
				1:1	> 2:1	1:1	> 2:1	3:1	>10:1	1:2:1
EST SSRs	29	91	73	33		21	2	15	2	
Genomic SSRs	78	210	137	48	2	50		25	12	
SNPs	1536	n/a	n/a	238	16	212	1			191
Total		n/a	n/a	318	17	280	0	40	0	191

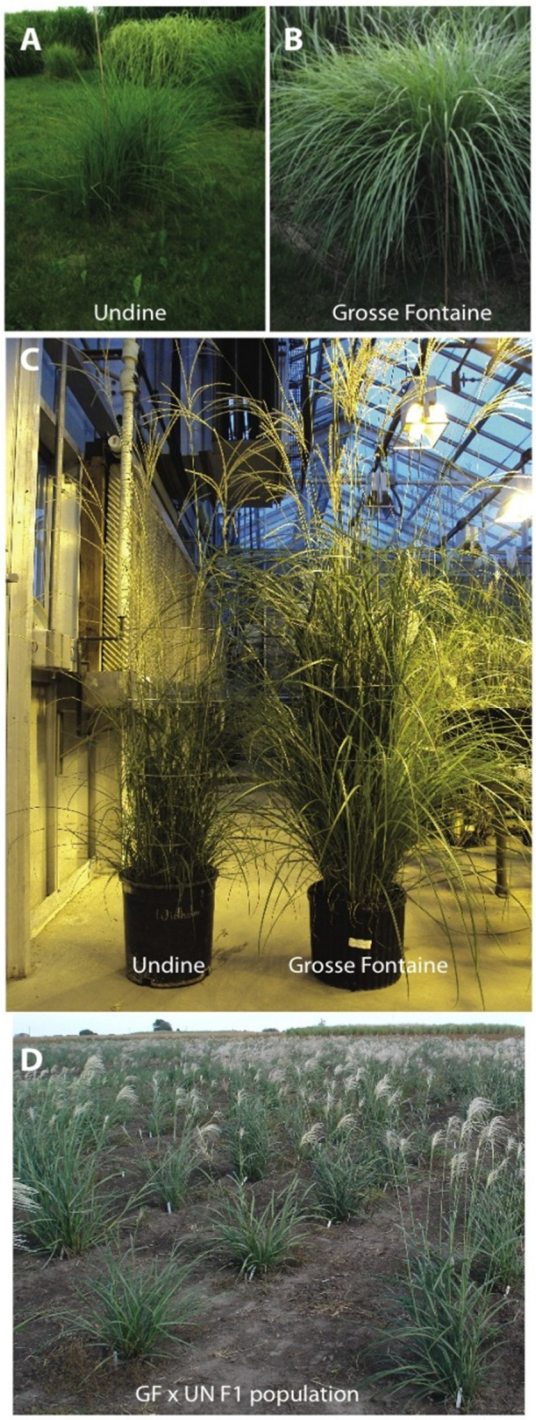

Figure 3.1. Parents of the *M. sinensis* mapping population, as published in Swaminathan *et al*. (2012). A full sib (F_1) population was produced by reciprocally crossing two ornamental *M. sinensis* accessions, "Grosse Fontaine" (GF) and "Undine" (UN). The parents of the mapping population were propagated from rhizomes taken from two individual plants, UN (panel A, plant location SF20) and GF (panel B, plant location SF5), established by Emily Heaton in the spring of 2001 at the SoyFACE plot located at the Crop Sciences Research and Education center, on the University of Illinois campus. The photographs of GF and UN were taken on July 11 2011. Panel B shows the individual GF and UN plants used to make the F_1 mapping population, flowering in the greenhouse (photograph taken on Dec 14th 2009). Panel D shows the full sib F_1 mapping population in August 2010 at the University of Illinois Energy Farm. This population was established in the spring of 2010.

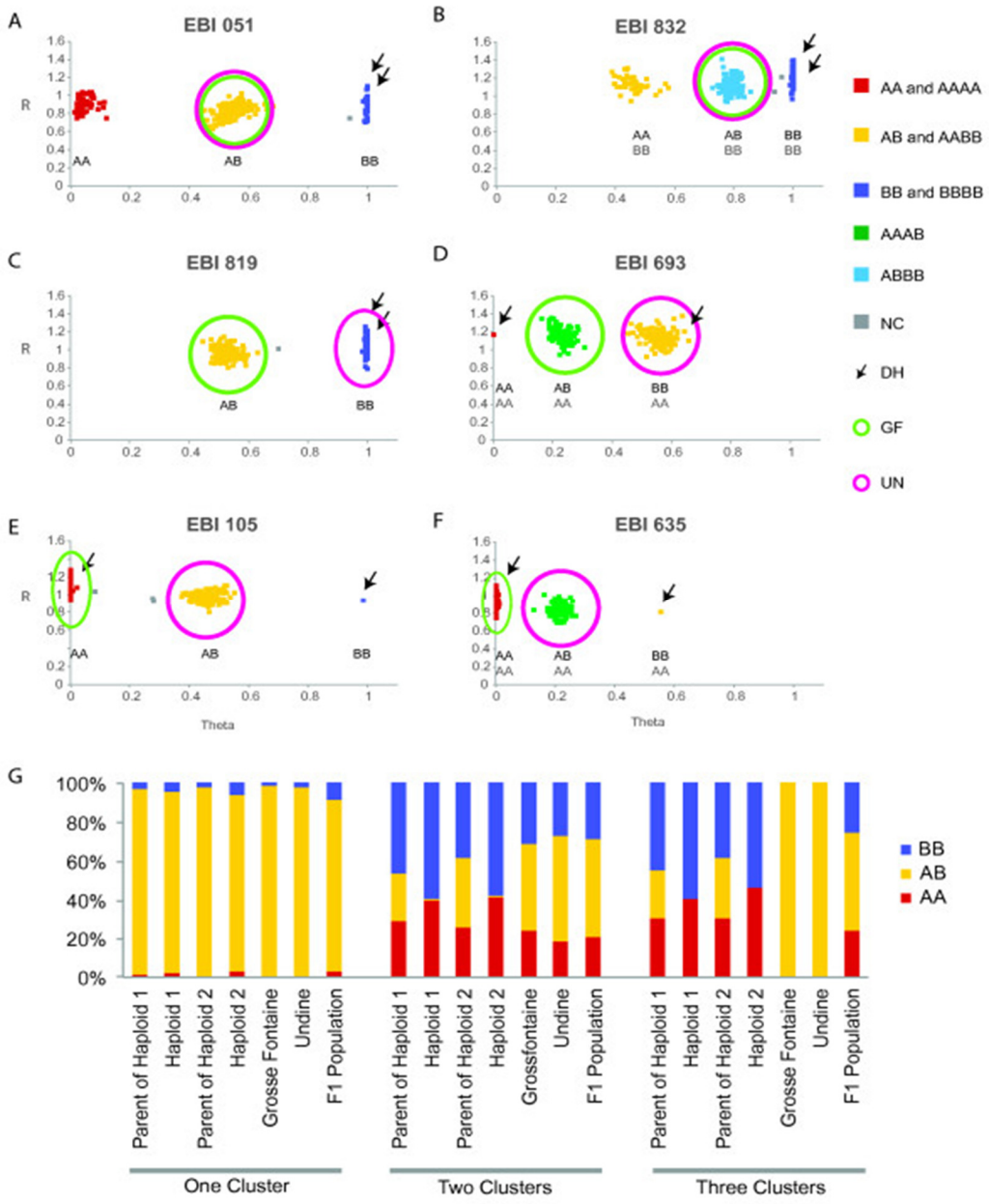

Figure 3.2. Genotype calling using the *Miscanthus* GoldenGate™ array, as published in Swaminathan *et al*. (2012). The graphs in panels A–F plot normalized theta (ratio of signal intensities assayed for A and B SNP alleles) against normalized R (signal intensity) for each individual represented as a colored square. Panels A, C, and E illustrate markers that cluster as predicted for a biallelic SNP, which segregate as AA (red), AB (yellow), or BB (blue). Panels B, D, and F illustrate markers that cluster as predicted for a SNP distinguishing alleles for one of two duplicated and unlinked loci, where theta is skewed by the relative dosage of A and B SNPs. In all panels, clusters are defined as sharing alleles with either the Grosse Fontaine (green circles) or Undine (pink circles) parents, individuals that fall outside the cluster are marked as "no calls" (NC, grey), and the doubled haploid genotype is indicated by the black arrow. Panel G reports the relative fraction of genotyped segregating SNPs within each clustering type among the Grosse Fontaine and Undine parents, the population of their F_1 progeny, as well as the two doubled haploids and their respective parents. Single cluster markers (fixed differences between paralogs) behave similarly in diploids and doubled haploids. In contrast, while diploid accessions show extensive heterozygosity at segregating loci (two- and three-cluster markers), doubled haploids show no heterozygosity.

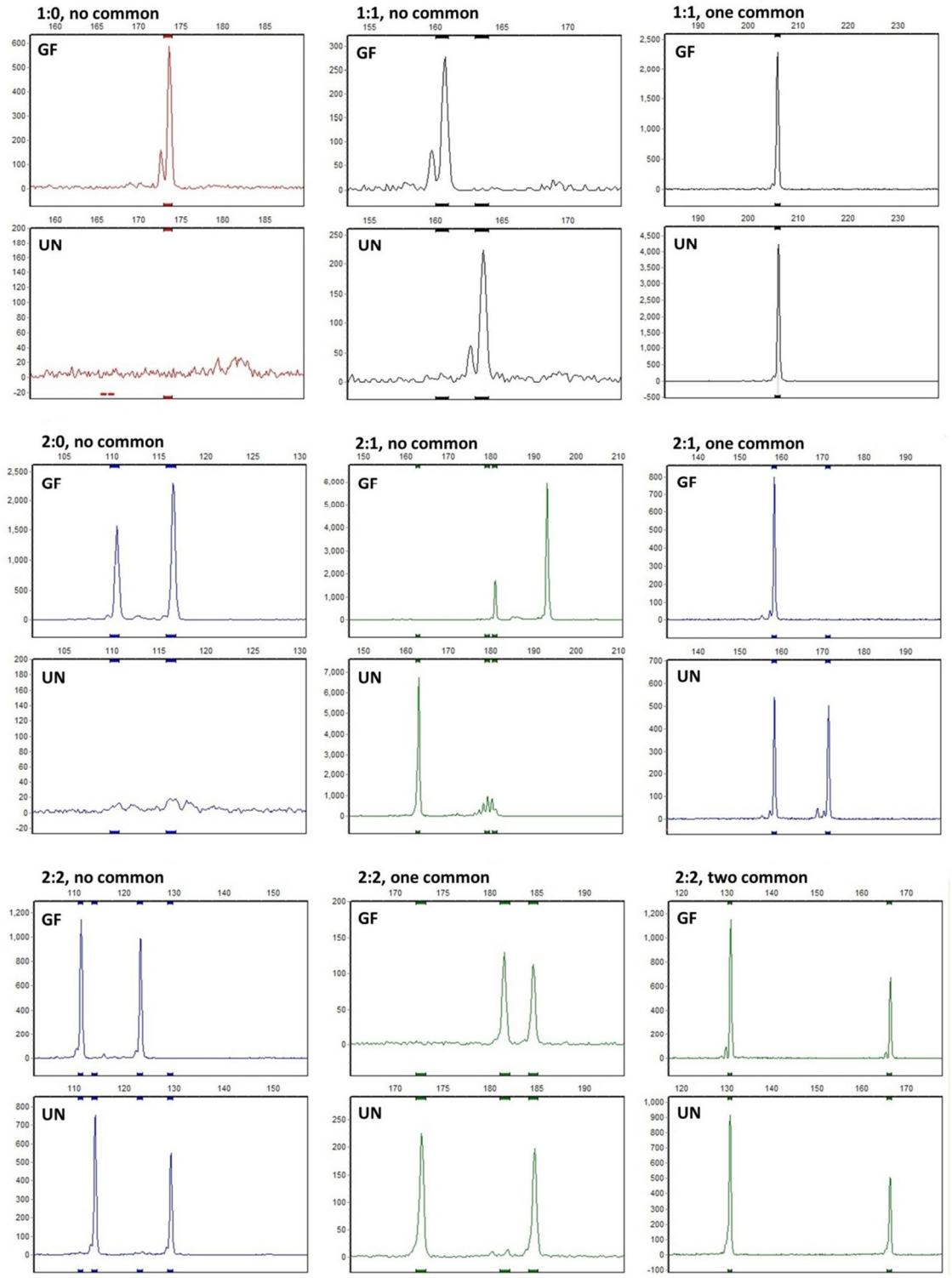

Figure 3.3. Different amplicon profiles seen in the fragment analysis of SSR markers, as published in Swaminathan *et al.* (2012). The length of the amplicon, in bp, is shown on the horizontal axis and the fluorescence intensity on the vertical axis. Several profiles show "stutter peaks" that are associated with a main peak. These are not counted as distinct marker states.

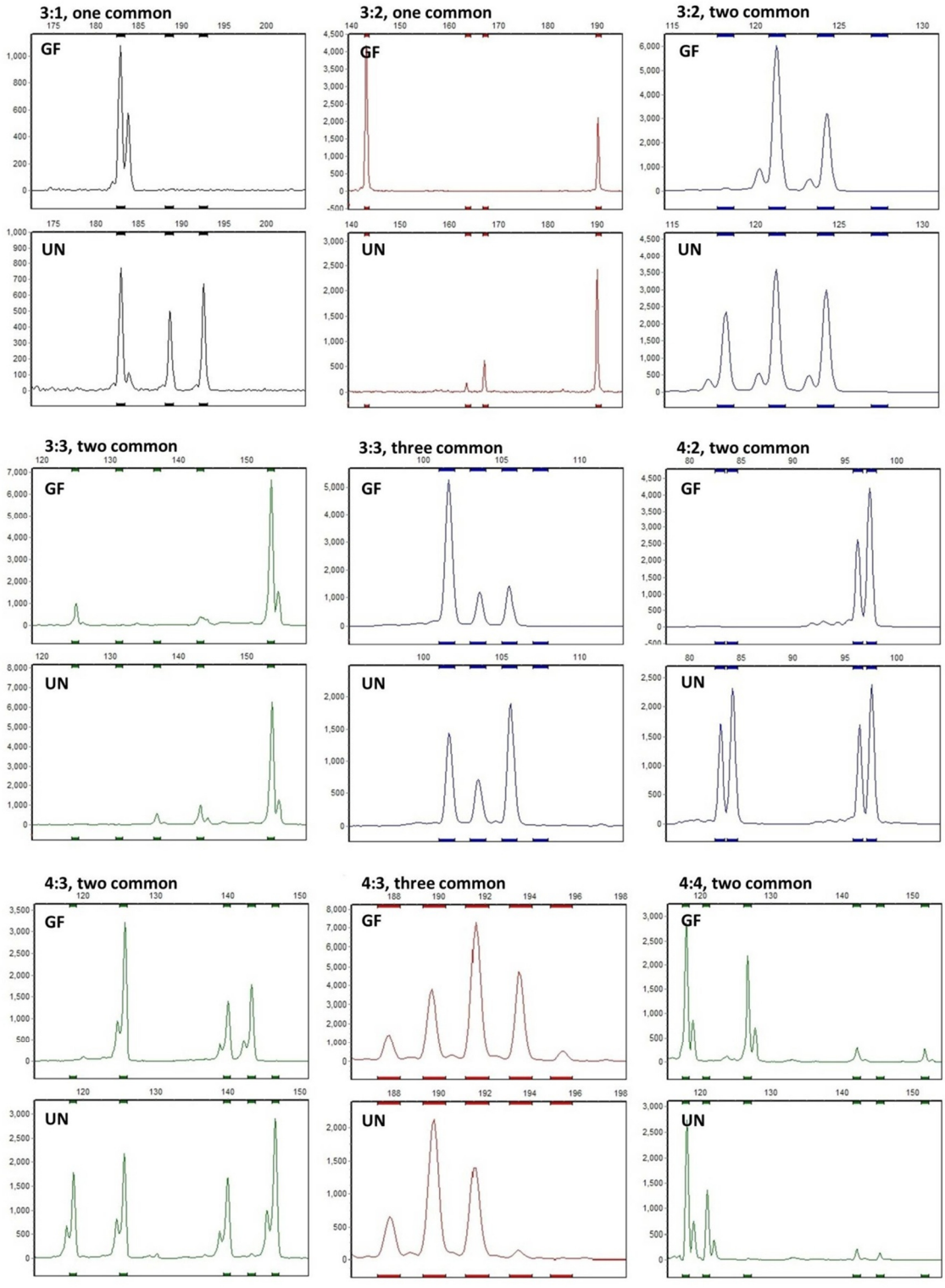

Figure 3.3. Continued

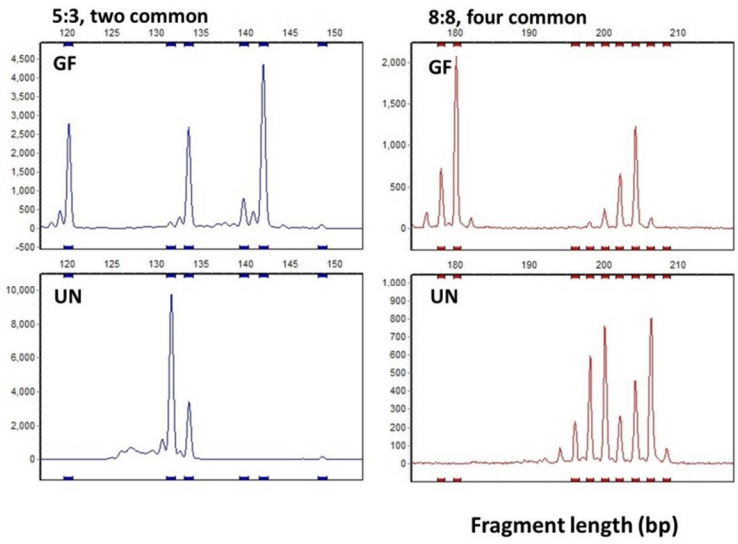

Figure 3.3. Continued

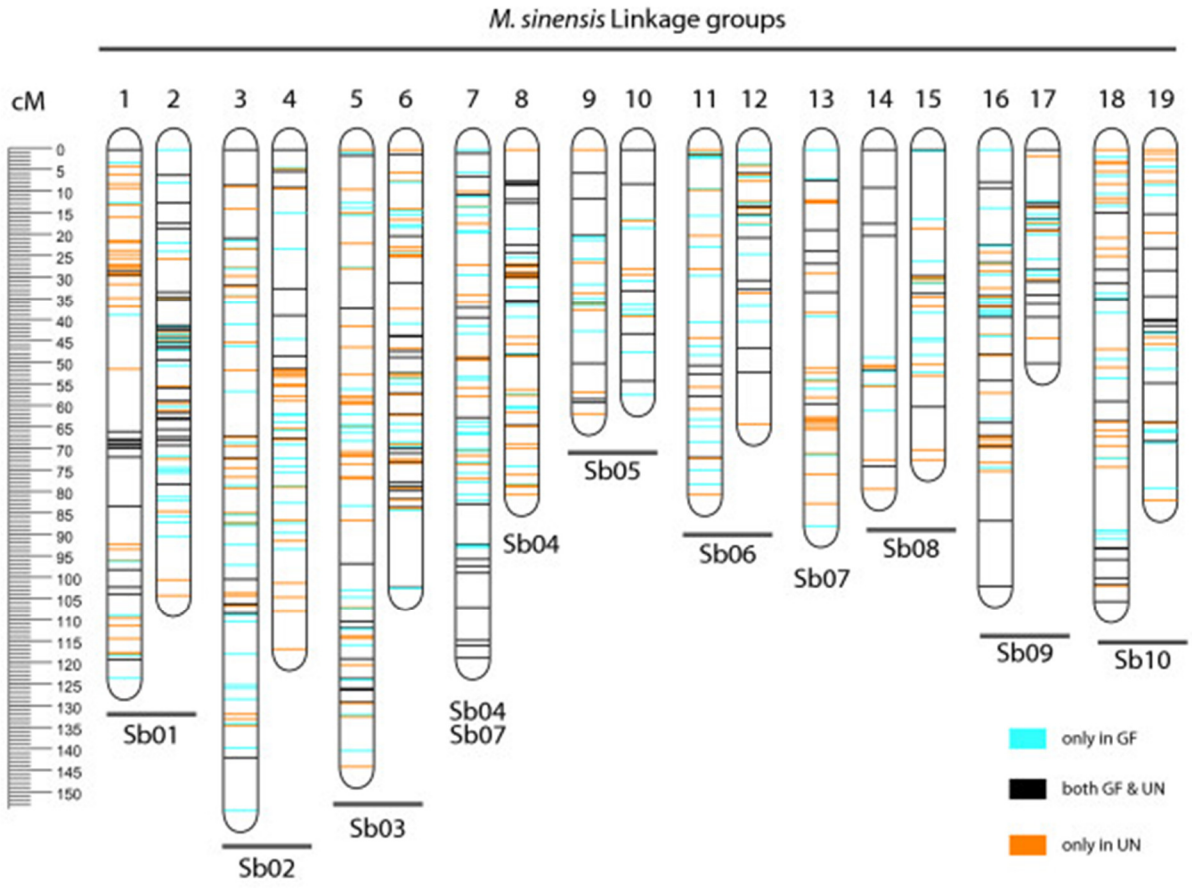

Figure 3.4. Integrated genetic map for *M. sinensis*, as published in Swaminathan *et al.* (2012). Markers on each of the 19 linkage groups are shown as horizontal lines at their estimated map position on the integrated map, with scale in centiMorgans on left. Marker type is shown by color (blue, heterozygous in GF only; orange, heterozygous in UN; black, heterozygous in both GF and UN). Markers that are heterozygous in both parents allow individual parental maps to be combined. Marker types and positions are specified in Additional file 9: Table S5. Linkage groups are numbered based on extensive synteny with sorghum (see Figure 4)

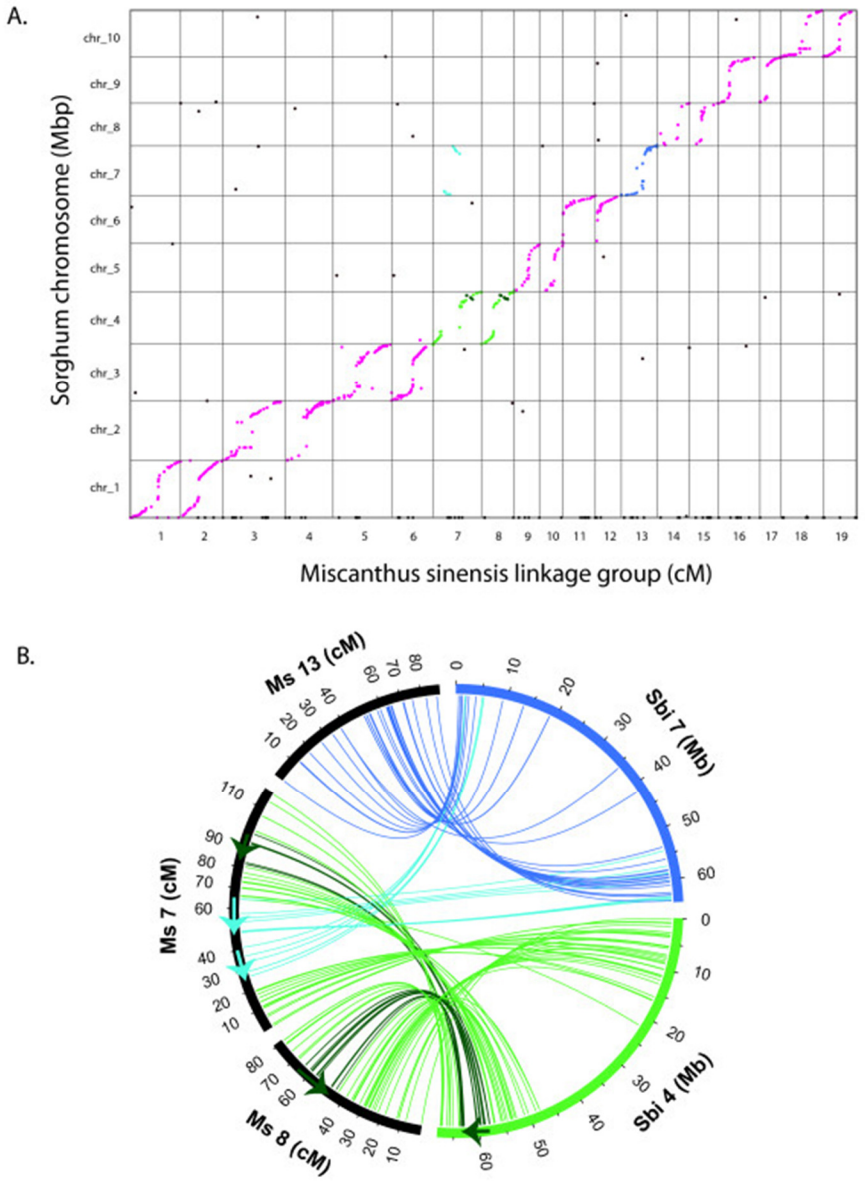

Figure 3.5. Tetraploidy of *Miscanthus* relative to sorghum, with extensive colinearity and a single chromosome fusion, as published in Swaminathan *et al*. (2012). (A) Horizontal axis shows genetic map position of markers on the 19 *Miscanthus* linkage groups, in centiMorgans; vertical axis shows physical map position of markers aligned to the 10 sorghum chromosomes in megabases. Each dot corresponds to a single marker. Markers that could not be uniquely mapped to sorghum are shown along the horizontal axis as black dots. Duplication and colinearity of nearly all chromosomes is evident (markers in magenta). A copy of sorghum chromosome 7 (markers in sky blue) has been inserted into a copy of sorghum chromosome 4 (markers in green) to produce *Miscanthus* linkage group 7. Markers on *Miscanthus* linkage group 13, which are also syntenic with sorghum chromosome 7, are shown in a darker blue. (B) Circos plot showing centromeric insertion of sorghum chromosome 7 into sorghum chromosome 4 to form *Miscanthus* linkage group 7 (approximate boundaries indicated by arrows). Each line represents an orthologous relationship between a mapped *Miscanthus* marker and its unique counterpart on the *Sorghum bicolor* genome. Both *Miscanthus* linkage groups 7 and 8 have a region corresponding to sorghum chromosome 4, which is inverted with respect to the other markers (dark green arrow and lines). As also shown, *Miscanthus* linkage group 8 is an intact copy of sorghum chromosome 4, and *Miscanthus* linkage group 13 is an intact copy of sorghum chromosome 7.

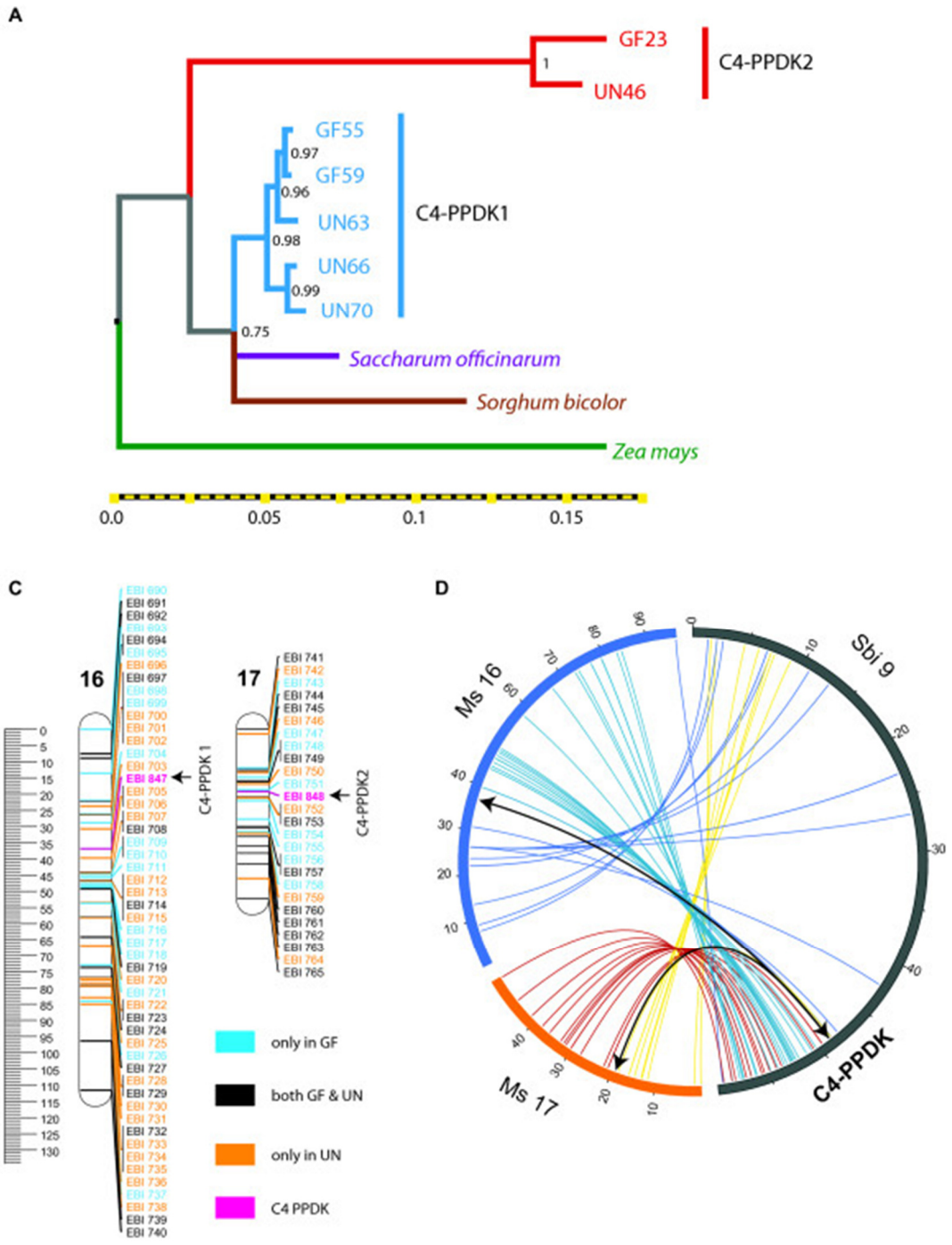

Figure 3.6. Mapping PPDK, as published in Swaminathan *et al*. (2012). (A) Phylogenetic tree of representative C_4-PPDK sequences showing the clear separation of the two paralogous sequences. (B) Placement of the C_4-PPDK markers, EBI 847 and EBI 848 (pink) on the linkage map. (C) Synteny of corresponding linkage groups (Ms16 and Ms17) to Sorghum chromosome 9. The position of PPDK is illustrated by the black arrows.

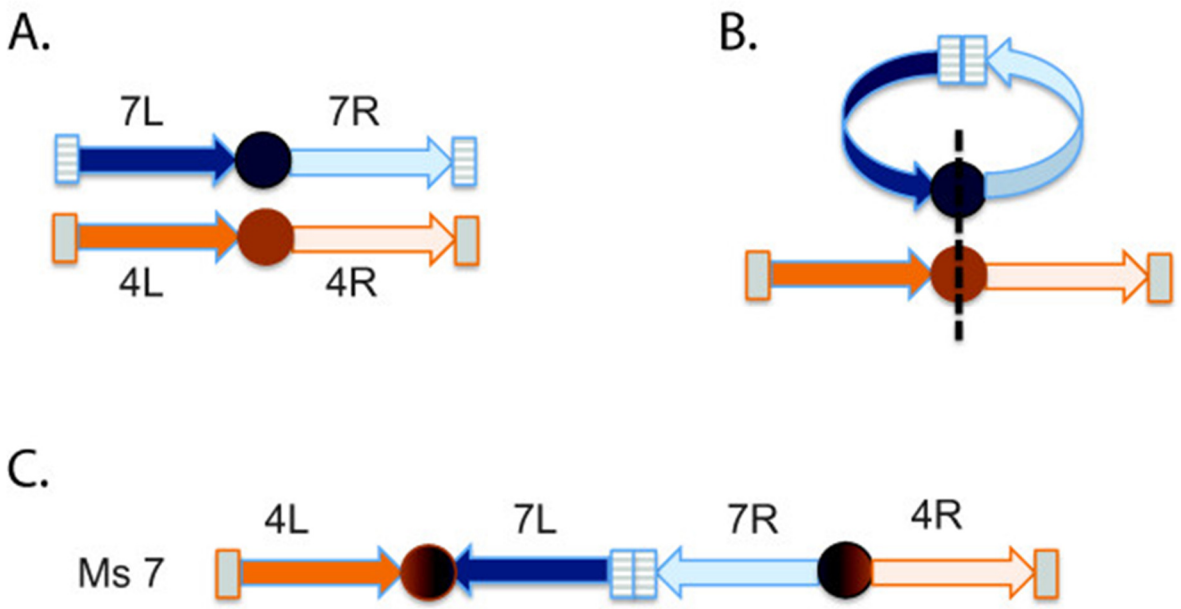

Figure 3.7. Mechanism for insertional dysploid reduction of ancestral chromosomes (Cordeiro *et al*, 2001) , as published in Swaminathan *et al*. (2012). The arms of the progenitors of Sb7 and Sb4 are indicated in blue and orange, respectively. Centromeres are shown as solid circles, and telomeres as grey rectangles. (A) original configuration. (B) intermediate state in which (1) chromosome 7 circularizes, and (2) 7 and 4 recombine at breaks that occur within centromeres. The order of occurrence of circularization and break/recombination are unspecified. (C) The resulting order and orientation matches that found in Ms7 (Figure 3.5B). Note that the original telomeres of 7 lie within the fused chromosome, and are presumably lost. Only one of the two centromeres (shown as mix of orange and blue) survives. An analogous event is proposed to have produced chromosome 4 of *Aegilops tauschii* (Cordeiro *et al*, 2001).

CHAPTER 4

Synthetic polyploid production of *Miscanthus sacchariflorus*, *M. sinensis* and *M.* x *giganteus*[2]

4.1. Abstract

Plants from the genus *Miscanthus* are potential crop candidates for renewable sources of lignocellulosic biomass for energy production. A potential strategy for *Miscanthus* crop improvement involves interspecific manipulation of ploidy levels to generate superior germplasm and to circumvent reproductive barriers for the introduction of new genetic variation into core germplasm. Synthetic autotetraploid lines of *Miscanthus sacchariflorus* and *M. sinensis*, and autoallohexaploid *M. x giganteus* were produced in tissue culture from oryzalin treatments to seed- and immature inflorescence-derived callus lines. This is the first report of the genome doubling of diploid *M. sacchariflorus*. Genome doubling of diploid *M. sinensis*, *M. sacchariflorus,* and triploid *M. x giganteus* to generate tetraploid and hexaploid lines was confirmed by stomata size, nuclear DNA content, and chromosome counts. A putative pentaploid line was also identified among the *M. x giganteus* synthetic polyploid lines by nuclear DNA content and chromosome counts. Comparisons of phenotypic performance of synthetic polyploid lines with their diploid and triploid progenitors in the greenhouse found species-specific differences in plant tiller number, height, and flowering time among the doubled lines. Stem diameter tended to increase after polyploidization but there were no significant improvement in biomass traits. Under field conditions, *M. x giganteus* synthetic polyploid lines showed greater phenotypic variation, in terms of plant height, stem diameter and tiller number, than their progenitor lines. Production of synthetic autopolyploid lines displaying significant phenotypic

[2]This chapter has been accepted for publication in Global Change Biology: Bioenergy pending minor revision.

variation suggests that ploidy manipulation can introduce useful genetic diversity in the limited *Miscanthus* germplasm currently available in the United States. The role of polyploidization in the evolution and breeding of the genus *Miscanthus* is discussed.

4.2. Introduction

Polyploidy is a common genomic feature in all eukaryotes, particularly prominent in plants. It is broadly classified into auto- and allopolyploidy, which are formed by doubling the diploid genome within a species and from hybrids between species, respectively. Polyploidization, together with hybridization, plays an important role in plant evolution and speciation (Soltis & Soltis, 2009). Different from the gradual evolutionary process whereby new species evolve from isolated populations, new species of plants can emerge abruptly via polyploidization. Differences in chromosome number provide instant reproductive isolation, limiting gene flow between neopolyploids and their progenitors (Tate *et al*., 2005; Otto, 2007). Furthermore, physiological and morphological changes due to polyploidization may alter reproductive biology of neopolyploids (Thompson & Lumaret, 1992).

Polyploidization can alter cytogenetic, genetic and epigenetic characteristics of organisms resulting in phenotypic variation among neopolyploid plants, which would be a target for natural selection. Genome duplication increases cell volume by increasing genome size and has been associated with delayed development of polyploid plants. Larger genomes require more time for cell replication and tend to display slower cell growth rates (Bennett & Leitch, 2005). A slower cell cycle can result in delayed and/or prolonged flowering (Ramsey and Schemske, 2002) by postponing the termination of apical growth which can lead to a longer period of vegetative growth (Salas Fernandez *et al*., 2009).

The genome of newly formed polyploid plants usually undergoes extensive genetic and epigenetic changes. Genomes of neopolyploids are usually unstable and experience rapid repatterning (Wendel, 2000) often beginning in the first generation after polyploidization (Levy & Feldman, 2004). The extensive and rapid genomic rearrangements are likely due to sequence rearrangements, homoeologous recombination, and sequence elimination (Adams & Wendel, 2005; Otto, 2007). The heritable epigenetic changes, such as DNA methylation, histone modification and RNA interference can also alter gene expression (Wolffe & Matzke, 1999; Liu & Wendel, 2003; Doyle *et al*., 2008). Additionally, the activation or suppression of transposable elements is considered an important component of the evolution of polyploid genomes (Matzke & Matzke, 1998). These genetic and epigenetic changes can result in phenotypic variation.

Miscanthus species are perennial C_4 grasses and have a base chromosome number of x=19, with nominally diploid (2n=2x=38) and tetraploid (2n=4x=76) accessions. One of these species, *Miscanthus x giganteus* (2n=3x=57) is a candidate for dedicated bioenergy crop production due to its high biomass productivity and capacity to capture greenhouse gases by sequestering carbon in underground rhizomes (McLaughlin & Walsh, 1998). While *M. x giganteus* is considered an excellent bioenergy crop due to its high biomass productivity and cold tolerance (Heaton *et al*., 2010), it has limited genetic diversity with very few different genotypes available in Europe and the United States (Greef *et al*., 1997). Introducing new variability in *M. x giganteus* is constrained by the sterility associated with its triploid genome. The fertile, diploid parental species of *M. x giganteus*, *M. sinensis* and *M. sacchariflorus* are also potential candidates as bioenergy crops in that they can produce substantial biomass in certain environments (Clifton-Brown *et al*., 2001) and show extensive genetic variability in their native ranges in eastern Asia (Jakob *et al*., 2009; Sacks *et al*., 2012).

In the United States, *M. sinensis* was first introduced from Japan in the late 1800s (Quinn *et al*., 2010) and various *Miscanthus* accessions have been introduced thereafter, mainly as ornamentals. Efforts are underway to add diverse *Miscanthus* genotypes to the current germplasm pool in the United States with accessions from Korea, Japan and China but export and import restrictions have been a major barrier and slowed germplasm acquisition (Jakob *et al*., 2009). Currently at University of Illinois Urbana-Champaign (UIUC), more than 101 accessions of *Miscanthus* have been collected (mostly ornamentals) and planted both in the field and in the greenhouse. Within this collection, most are of *M. sinensis* with only a few diploid and tetraploid *M. sacchariflorus* and *M. x giganteus* accessions (W. Chae *et al*., in preparation).

Potential strategies for *Miscanthus* crop improvement include the resynthesis of new triploid *M. x giganteus* genotypes by conventional hybridization between tetraploid *M. sacchariflorus* and diploid *M. sinensis*, intra- and inter-specific hybridization among diploid compatible *Miscanthus* species followed by successive selection, and by manipulation of ploidy levels to circumvent reproductive barriers (Sacks *et al*., 2012). Although prevalent in Japan (Nishiwaki *et al*., 2011), few tetraploid accessions of *M. sacchariflorus,* one of the presumed parents of *M. x giganteus*, are currently available in the United States. Therefore, producing tetraploid *M. sacchariflorus* plants would be desirable to resynthesize new genotypes of triploid *M. x giganteus* germplasm. The production of triploid plants can also circumvent invasive issues associated with the two fertile seed-bearing parental species (Quinn *et al*., 2010).

Certain diploid *Miscanthus* species can produce fertile seed from interspecific hybridization and efforts to produce such hybrids have been made for decades. In Japan, hybrid populations from crosses between *M. sinensis* and *M. floridulus* (Adati & Shiotani, 1962) and between *M. sinensis var condensatus* and *M. tinctorius* (Hirayoshi *et al*., 1959) were produced.

Recently, several populations from crosses between *M. sinensis* and *M. sacchariflorus* have been produced by Mendel Biotechnology, Inc. (Hayward, California, USA) and by our research group at UIUC. There is also potential to generate progeny from crosses between plants with different ploidy levels. Triploid *Miscanthus* hybrids were produced from conventional crosses in Japan (Hirayoshi *et al*., 1960) and recently, we have generated new triploid *M. x giganteus* plants from the cross of diploid *M. sinensis* by tetraploid *M. sacchariflorus*.

Whole genome duplication via polyploidization is thought to have occurred in the genus *Miscanthus* in the process of or after divergence from the closely related Saccharinae clade (Swaminathan *et al*., 2012) fewer than 3 million years ago (Paterson *et al*., 2010). Also, the genus *Miscanthus* contains accessions with varying ploidy levels in the wild. Genetic and epigenetic changes associated with genome doubling in *Miscanthus* could result in phenotypic variation for biomass characteristics and delayed flowering time. The objective of this study was to produce synthetic autopolyploids from clonal accessions of three *Miscanthus* species to compare cytogenetic and phenotypic differences among polyploid lines and their progenitor genotypes.

4.3. Materials and Methods

Nomenclature

To define terms used in this article a ‘line’ refers to plants generated from the same callus that represents an independent event of genome doubling. A ‘regenerated’ and ‘oryzalin treated and regenerated’ line refers to control plants regenerated from callus without and with oryzalin treatment, respectively, and have the same ploidy level as the diploid and triploid progenitors. A ‘synthetic polyploid’ represents plants regenerated from callus treated with

oryzalin and having a doubled genome compared to the corresponding progenitor. Thus, synthetic autopolyploids of *M. sinensis* and *M. sacchariflorus* are tetraploid. In *M. x giganteus*, synthetic polyploids include both hexaploid and putative pentaploid lines unless otherwise specified.

Plant materials, explant tissues, and tissue culture media

Seeds (Jelitto Staudensamen GmbH, Schwarmstedt, Germany) or immature inflorescence tissues from eight genotypes of *M. sinensis*, *M. sacchariflorus* and *M. x giganteus* (Table 4.1) were sterilized by immersion in 0.5% NaOCl solution for 10 min and rinsed three times with sterilized double distilled (dd) H_2O. Immature inflorescences (2 to 4 cm long) were taken from the terminal nodes of culms following removal of sheath leaves just before flag leaves became visible. Sterilized seeds and immature inflorescence explants (2 to 3 mm) were then transferred to callus induction medium consisting of MS basal salts and MS vitamins (Murashige & Skoog, 1962) with 13.6 mM of 2,4-dichlorophenoxyacetic acid (2,4-D), 0.44 mM of 6-benzylaminopurine (BAP), 2.88 gL^{-1} of proline, 30 gL^{-1} of sucrose, and 750mgL^{-1} $MgCl_2 \cdot 6H_2O$ (Petersen, 1997). All media were supplemented with 2 g L^{-1} PhytagelTM agar (Sigma-Aldrich, St Louis, MO, USA), and adjusted to pH 5.5 prior to autoclaving. Cultures were incubated in darkness at 27±2°C and subcultured at two week intervals for the first month followed by three week intervals thereafter.

Production of synthetic autopolyploids

Three months after culture initiation, calli were treated with 10 µM of oryzalin (Sigma-Aldrich, St Louis, MO, USA) in liquid callus maintenance medium having the same composition

as the callus induction medium aforementioned excluding BAP. The concentration and exposure duration of oryzalin treatment was determined based on the results of Yu *et al*. (2009). Filter-sterilized oryzalin solutions were added after autoclaving the medium. 1–2 mm pieces of calli from each genotype were transferred into 50 mL of liquid callus maintenance medium in a 250mL Erlenmeyer flask. The flasks were shaken at 120 rpm at 27±2°C in the dark for 36 hours. After treatment, calli were transferred to solid callus maintenance medium and subcultured at two week intervals. Forty days after oryzalin treatment, callus survival rate was recorded (Table 4.1) and calli were transferred to regeneration medium consisting of MS basal salts and MS vitamins with 1.3 mM of NAA, 22 mM of BAP, 20 gL^{-1} of sucrose, and 750mgL^{-1} of $MgCl_2 \cdot 6H_2O$. Regeneration media was supplemented with 3 gL^{-1} of PhytagelTM agar and adjusted to pH 5.5 before autoclaving. The growth conditions were 16 h of cool white fluorescent light (40 microEinsteins s^{-1} m^{-2}) at 27±2°C. Calli were maintained for 70 days in regeneration medium, and regenerated shoots were transferred into MS basal medium (Murashige & Skoog, 1962) and maintained for 40 - 60 days for root induction. Plantlets with roots were transferred to plastic pots (700cm^3) containing Metro-Mix 500 (Sun Gro Horticulture, Canada Ltd.) in the greenhouse at UIUC. The temperature of the greenhouse was maintained at 28°± 2°C day (14 hr)/22°± 2°C night (10 hr), with supplemental lighting provided from 6 am to 8 pm if light intensities fell below 2670 microEinsteins s^{-1} m^{-2}.

Determination of plant ploidy levels

Three to four months after transplanting in the greenhouse, regenerated plants were screened for putative polyploids by measuring leaf guard cell length. Leaves from all oryzalin-treated plants and corresponding regenerated progenitors were collected for comparison of

guard cell length using microphotography of epidermal impressions as described by Rayburn *et al*. (2009). The plants possessing average stomata lengths 20% or larger than progenitors were considered putative synthetic polyploids (Yu *et al*., 2009). These lines were then subjected to flow cytometric analysis to measure their nuclear DNA (nDNA) content using the modified protocol of Yu *et al*. (2009). Ploidy levels were confirmed by comparing the peaks in the nDNA histogram between those of the progenitor (external standard) and putative polyploid lines. Briefly, young leaf tissue (1 cm^2) from newly emerging shoots from both a treated and control plant was chopped in a petri-dish containing 10 ml extraction buffer consisting of 13% (v/v) hexylene glycol, 10 mM Tris–HCl (pH 8.0), and 10 mM $MgCl_2$ with 200 µl of 25% Triton X. The samples were filtered through a 50-µm nylon mesh (Partec GmbH, Gorlitz, Germany) into a labeled test tube and kept on ice throughout. Following filtration, samples were centrifuged for 25 min at 300 xg at 4°C. The supernatant was then aspirated, and nuclei are resuspended in 300 µl of propidium iodide (PI, Sigma-Aldrich, St Louis, MO, USA). Then the solution was transferred to a 1.5 mL micro-centrifuge tube and incubated for 20 min at 37°C. After incubation, 300 µL of PI salt was added to each sample. Samples are then briefly vortexed, placed on ice, and stored at 4°C for at least 1 h. Nuclei were analyzed using flow cytometer Model LSRII (BD Biosciences, Flow Cytometry Facility at the University of Illinois-Keck Biotechnology Center). The excitation wavelength was set at 488 nm and a 570 nm emission filter was used. A minimum of 20,000 nuclei per sample were analyzed.

Chromosome counting of root tips

Chromosome numbers were counted for three progenitors and derived synthetic polyploid lines and one putative pentaploid line, Mxg5x-2 (Table 4.2). Root tips 1–2 cm in

length were excised and soaked in 0.05% 8-hydroxyquinoline for mitotic inhibition. After three hours, the root tips were rinsed in ddH_2O for 5 min and stored in 3:1 (v/v) 100% ethanol/acetic acid. The roots were stored at room temperature for four days and then stored at 4°C until use. Fixed root tips were rinsed in ddH_2O, hydrolyzed in 5 N HCl for 45 min and placed in Feulgen's stain for 2 h. Root tips were then rinsed in ddH_2O and a drop of 1% acetocarmine was added to the root tip. A cover slip was placed over the tissue and gently tapped with a dissecting needle to disperse the tissue. The slide was then flamed over an alcohol burner, and direct pressure was applied to the slide. The slides were then viewed using an Olympus BX61 microscope (Olympus America Inc., Melville, NY, USA). Photographs of chromosome spreads were taken using an Olympus U-CMAD3 camera and chromosome counts conducted on the clearest preparations.

Determination of 2C DNA contents in synthetic autopolyploid lines

The 2C-values of nDNA in five genome doubled lines for each of *M. sinensis* 'Grosse fontaine,' *M. sacchariflorus* 'Golf course' and *M. x giganteus* 'Illinois' and their diploid or triploid progenitors (Table 4.2) were measured by flow cytometry with sorghum as an internal standard using the modified protocol of Rayburn *et al*. (2009). The protocol is the same as described above with the replacement of the progenitor control with sorghum as the internal standard being co-chopped with the *Miscanthus* lines. The 2C value of nDNA content of the sorghum line was calibrated at 1.74 pg using the maize genotype W-22 as a calibration standard which is reported to have 5.35 pg DNA/2C (McMurphy & Rayburn, 1991). Mean fluorescence of the *Miscanthus* G1 peak is divided by the fluorescence reading of the internal standard, multiplied by 1.74 pg/2C, and expressed in pg/2C nucleus. Since sample G2/G1 peak ratio is

typically slightly less than the expected value of 2.0 (Wood & Todd, 1979; Watson, 1991), samples with sorghum G2/G1 peak ratios outside the range of 1.94–2.03 were excluded. Sample target peaks in DNA histograms that were not symmetrical or where coefficients of variation exceeded 5% were also excluded. For each accession, 3 - 5 samples were examined and samples that did not meet these criteria were discarded until three acceptable replications (one leaf per replication) were recorded.

Phenotypic evaluations of synthetic autopolyploid lines

Phenotypic evaluations of synthetic polyploid lines were conducted in the greenhouse under conditions described above. The investigation was performed on fifteen synthetic polyploids and their three progenitors: *M. sinensis* 'Grosse fontaine,' *M. sacchariflorus* 'Golf course' and *M. x giganteus* 'Illinois' (Table 4.3). Synthetic polyploid *M. x giganteus* 'Illinois' plants produced from Yu *et al*. (2009) were used in the evaluation with the addition of Mxg6x-20, a newly added line from the aforementioned tissue culture procedure (Table 4.1 and 4.3). Each line was divided into single stems with rhizome portions (*M. x giganteus* and *M. sacchariflorus*) of equivalent size and weight and planted in 10 L pots on October 7, 2010. A randomized complete block design with four replications was used with each block containing 21 lines from the three species (one regenerated control line, one oryzalin-treated and regenerated control line and 5 synthetic polyploid lines for each species). Phenotypic data were recorded every week from December 2010 to June 2011 for flowering time, plant height, stem diameter, and the number of tillers. Plant height was measured from ground to highest point and the stem diameter was taken at one-third of mature plant height from the base.

Field evaluation for phenotypic variation among *M. x giganteus* lines was conducted on the Energy Bioscience Institute farm at UIUC. Five regenerated control triploid lines, five oryzalin-treated and regenerated control triploid lines and eight synthetic hexaploid *M. x giganteus* lines (Yu *et al*. 2009) were utilized for this investigation. Transplants were prepared as described above and planted on June 16th, 2010 in a randomized complete block design with three replications with 1.5m spacing. A block contained 18 lines with 3 clones per each line (18 x 3 = 54 plants). Data on tiller number, plant height and stem diameter were collected October 17th, 2011. Phenotypic data was collected as described above.

Statistical analysis

To estimate overall phenotypic effect of neopolyploidization, t-tests were performed on the phenotypic data between calculated means from 2 progenitor lines and those from 5 synthetic polyploid lines of each species (Table 4.3). Analysis of variation (ANOVA, General Linear Model) was performed on the 2C value of nDNA content and the phenotypic data among polyploids and progenitors with significant differences between mean values were determined by the Student's t-test. For field data for synthetic polyploids of *M. x giganteus*, the means of each line were grouped into three categories including regenerated triploid controls, oryzalin-treated regenerated triploid controls and synthetic hexaploid lines, and box plots generated from the data. All statistical analyses were performed and box plots were generated using JMP software (SAS Institute, Cary, NC).

4.4. Results

Calli induction and survival and shoot regeneration

All explants from the eight genotypes of the three species produced calli on media with the same composition. Calli from each explant source were observed to proliferate after oryzalin treatments for 2 days. The survival rates of oryzalin-treated calli after 40 days of culture on solid callus maintenance medium differed depending on explant source and species. The survival rates of seed-derived calli were much higher than those of immature inflorescence-derived calli in *M. sinensis* (Table 4.1). Among immature inflorescence-derived calli, *M. x giganteus* 'Illinois' showed the highest and *M. sinensis* 'Undine' the lowest survival rates. This could be due to the differential response of explants and genotypes to the callus proliferation medium.

The number of regenerated shoots was counted at 70 days after culture on the regeneration medium. Approximately half of the calli transferred to the regeneration medium across all genotypes displayed necrosis and did not produce shoots. Shoot producing calli were usually compact and white embryogenic-like calli (Kim *et al*., 2010). Calli from *M. sinensis* 'Undine,' *M. sacchariflorus* 'Blue stem' and 'Robustus' did not produce any regenerated shoots (Table 4.1). This indicates variation among genotypes in response to shoot regeneration media, suggesting different accessions require different media composition for shoot differentiation as has been observed in previous studies (Petersen *et al*., 2002; Wang *et al*., 2011).

Determination of ploidy level and nDNA contents of synthetic polyploids

The oryzalin-treated regenerated plants with average stomata size 20% or larger than their progenitor plants were selected as putative synthetic polyploid plants (Figure 4.1). nDNA content of all selected plants was approximately double that of their progenitor counterparts. For

oryzalin-treated regenerated plants from immature inflorescence-derived calli, polyploidization rates were 44.9% for *M. sinensis* 'Grosse fontaine,' 39.0% for *M. sacchariflorus* 'Golf course' and 30.4% for *M. x giganteus* 'Illinois' (Table 4.1). One synthetic polyploid line per species was selected randomly for chromosome counting. Nuclei of synthetic polyploids from *M. sinensis*, *M. sacchariflorus* and *M. x giganteus* were observed to contain 76, 76, and 114 chromosomes, respectively, confirming successful genome doubling in all three *Miscanthus* species as counts matched the expected number of chromosomes (Figure 4.2 and Table 4.2).

Five independently produced, synthetic polyploid lines from each species were randomly selected and 2C nDNA content of each determined by flow cytometry as described above. The DNA histograms of regenerated control and synthetic polyploid lines of each *Miscanthus* species are shown in Figure 3. Coefficient of variations (standard deviation/mean ×100) of nDNA content of all lines analyzed did not exceed 5% (data not shown). All of the synthetic polyploid lines from *M. sinensis* and *M. sacchariflorus* have the approximately doubled mean nDNA content of progenitor lines, 8.96 ± 0.09 pg and 10.71 ± 0.24 pg, respectively; however, a synthetic polyploid line from *M. x giganteus* had lower 2C nDNA content and chromosome numbers than expected (Table 4.2). This line (Mxg5x-02, Table 4.2) was selected for chromosome counting and was observed to harbor 95 chromosomes (Table 4.2, Figure 4.2g) and therefore, is considered a putative pentaploid line (5x, 5 × 19 = 95 chromosomes).

The effect of synthetic polyploidization

The effect of polyploidization on phenotypic performance in synthetic polyploid lines differed among three *Miscanthus* species grown in the greenhouse. Simple t-tests between the means of progenitor and polyploid lines showed that the number of tillers ($p < 0.001$) and plant

height ($p < 0.01$) were significantly different in *M. sacchariflorus*. The number of tillers ($p < 0.05$) and flowering time ($p < 0.05$) in *M. sinensis*, and stem diameter ($p < 0.05$) in *M. x giganteus* were also significantly different between synthetic polyploid and progenitor lines (Table 4.3). Overall, the number of tillers was reduced in all synthetic polyploid lines. However, in *M. sinensis* and *M. x giganteus*, plant tiller numbers in the synthetic polyploid lines, GF4x-11 and Mxg6x-20, are comparable to that in progenitor lines, respectively (Table 4.4). Plant heights of synthetic polyploid lines of *M. sinensis* and *M. x giganteus* tended to be smaller than that of regenerated progenitors (Table 4.3); nevertheless, some synthetic polyploid lines displayed similar plant heights to regenerated progenitors (Table 4.4, Figure 4b, c). Polyploidization did not affect flowering time in *M. sacchariflorus,* but delayed flowering in *M. sinensis* and *M. x giganteus*. Flowering in three synthetic polyploid lines of *M. x giganteus* was significantly delayed, from 2 to 17 weeks compared to their progenitors (Table 4.4). Although there were no clearly superior synthetic polyploid lines compared to regenerated lines in terms of biomass characteristics, the phenotypes of GF4x-11 and Mxg6x-20 are comparable to their diploid and triploid progenitors, *M. sinensis* and *M. x giganteus*, respectively (Table 4.4).

Field evaluation of biomass accumulation traits in *M. x giganteus* confirms that genome doubling can generate phenotypic variation. Since these lines are the first generation after genome doubling, oryzalin-treated and regenerated control lines (triploid) were also used in order to investigate the effect of the antimitotic agent on the phenotype under field conditions. Synthetic polyploid (hexaploid) lines of *M. x giganteus* showed greater phenotypic variation for all three traits investigated than the regenerated and oryzalin-treated and regenerated control lines (Figure 4.5). The flowering time could not be investigated since most synthetic polyploid lines of *M. x giganteus* did not flower before senescence in the field due to the onset of chilling

fall temperatures (Figure 4.4c). The means of plant height and tiller number were significantly reduced in synthetic polyploid *M. x giganteus,* but the differences in plant height between synthetic polyploids and their progenitors was greater since most synthetic polyploid lines did not flower in the field, possibly due to slower vegetative growth. The mean stem diameter of synthetic polyploid lines increased slightly compared to their corresponding progenitor lines. These results were acquired from single year observation of a second year old field plot after planting and thus may not represent the phenotypic differences in fully established plant stands. Future research will be conducted over multiple years with mature stands. The interpretation of these results is also somewhat limited since we used only 10 progenitor and 8 synthetic polyploid lines; however, the pattern of greater phenotypic variation among synthetic polyploids was consistent with greenhouse observation.

4.5. Discussion

We produced synthetic polyploid plants of three *Miscanthus* species which represents the first report of artificial production of tetraploid plants from diploid *M. sacchariflorus*. Tissue culture produced calli from two different explant tissues and from various genotypes of three *Miscanthus* species on the same medium composition. While calli survived and proliferated after oryzalin treatment, some genotypes did not regenerate shoots (Table 4.1). This implies that different *Miscanthus* accessions require modification of the composition of shoot regeneration medium for recalcitrant genotypes. The high survival rates of calli after 10 μM of oryzalin treatment seen by Yu *et al*. (2009) was also observed in this experiment. Polyploidization rates ranged between 30.4 and 73.5% which are comparable to the highest polyploidization rates reported in other studies (Peterson *et al*., 2003; Yu *et al*., 2009; Głowacka *et al*., 2010).

The guard cell size is commonly used for estimating ploidy level in closely related species since it is often significantly larger in polyploids than in the diploid progenitors (Tate *et al*., 2005). The usefulness of guard cell length in distinguishing different ploidy levels has been demonstrated in other grass species such as barley (Borrino & Powell, 1988), rye grass (Speckmann *et al*., 1965), *Paspalum glaucescens* (Gramineae; Paniceae) (Pozzobon & Valls, 2000), wheat (Khazaei *et al*., 2010) and even in fossilized plants (Masterson, 1994). In our study, the preliminary screening of oryzalin-treated plants by comparison of stomata size with their corresponding progenitors was successful since all selected plants were confirmed as synthetic polyploids by flow cytometry and chromosome counts (Figures 4.1, 2, 3).

A putative pentaploid plant was obtained from oryzalin treatment to calli of triploid *M. x giganteus* "Illinois'. Odd ploidy level plants (triploids) were also observed in genome doubling of diploid *M. sinensis* by colchicine and oryzalin treatment (Peterson *et al*., 2003). Although the mechanism is not understood, genome doubling by antimitotic agents can result in aneuploidy and odd numbers of ploidy levels. This may be explained by residual antimitotic agent activity in cells in later cycles of mitosis after genome doubling. The concentration of residual antimitotic agent may vary among individual genome doubled cells and inhibit microtubule polymerization to various degrees in dividing cells, resulting in chromosomal aberration due to vagrant and/or laggard chromosomes in dividing cells (Sharma, 1990). The chromosomal aberration can result in a wide range in chromosome numbers in aneuploid cells and euploid cells containing either even or, by chance, odd ploidy levels. In successive generations, there may be tendency for the euploid genome cells to survive in a callus cell colony since aneuploidy may be deleterious to cell fitness due to changes in the copy number of structural genes (Torres *et al*., 2008) while euploid cells will maintain balanced genomes. As a result, a callus cell colony could be

composed primarily of euploid cells, with either an even or odd ploidy level, although portion of odd ploidy calli should be very small. In previous work, only nine triploids were obtained out of 1,377 regenerated plants (0.7%) following colchicine and oryzalin treatments of diploid *M. sinensis* (Peterson *et al*., 2003).

Synthetic polyploid lines in our study differed from their corresponding progenitors in phenotypic traits such as increased cell size (Figure 4.1) and stem diameter (Table 4.3). Similar results were observed in colchicine-induced polyploid *Miscanthus* species, which showed increased cell size and stem diameter (Głowacka *et al*., 2010). Earlier studies in other species revealed common phenotypic changes associated with neopolyploidy: coarser, thicker and larger leaves and larger reproductive organs (flowers and seeds) (Ramsey and Schemske 2002 and references therein). The 'gigas' characteristics of neopolyploids such as increased cell size, larger leaves, enlarged reproductive organs and robust stems are possibly due to an increased DNA content (Randolph, 1941).

Delayed flowering and reduced tiller number and plant height were observed in synthetic polyploids of all three *Miscanthus* species compared to their progenitors (Table 4.3) as was previously observed in *M. sinensis* (Głowacka *et al*., 2010). The delayed flowering, and reduced tillering and plant height are presumably due to delayed development in these polyploid lines. A positive correlation has been observed between genome size and the cell volume which is negatively correlated with the cell division rate (Bennett & Leitch, 2005). Polyploid cells are larger to accommodate larger genomes requiring more time to replicate with reduced cell cycle rates, thus delaying development and flowering (Noggle, 1946; Stebbins, 1971; Te Beest *et al*., 2012).

Allopolyploidy generally induces a greater variation in gene expression than autopolyploidy due to the combined effects of dosage changes and interactions between two genetically distinct genomes (Chen, 2007). The dramatic variation in flowering time evident in synthetic autoallopolyploids of *M. x giganteus*, which was not observed in the two synthetic 'autopolyploids' of *M. sinensis* and *M. sacchariflorus* (Table 4.4) may be due to differential transcriptional or post-transcriptional gene regulation following allopolyploidization. Variation in gene transcriptional regulation may be due to either genetic or epigenetic changes. Genomic rearrangement in chromosomal regions where important flowering genes are located may be responsible for the flowering time variation in hexaploid *M x giganteus* lines. Non-reciprocal transposition of the *FLOWERING LOCUS C* (*FLC*) gene resulted in flowering time variation among resynthesized *Brassica* allopolyploids (Schranz & Osborn, 2000; Pires *et al*., 2004). Genomic rearrangements may also be associated with the activation of transposable elements. Polyploidization has been observed to activate transposable elements (Kashkush *et al*., 2002) generating indels within and/or near coding sequences as well as resulting in the breakage or rearrangement of chromosomes. In *M. x giganteus*, repetitive sequences related to either transposable elements or centromeric repeats comprise 95% of genome (Swaminathan *et al*., 2010). Epigenetic changes such as variation in DNA methylation and histone modification can also account for phenotypic variation in synthetic polyploids (Otto, 2007). Differential post-transcriptional regulation in allopolyploids by small RNAs (Ha *et al*., 2009; Ng *et al*., 2012) could also be associated phenotypic variation among synthetic polyploid *M. x giganteus* lines.

The genus *Miscanthus* experienced a paleopolyploidization event and chromosome fusion after its divergence from closely related genera and before species diversification (Kim *et al*., 2012; Ma *et al*., 2012; Swaminathan *et al*., 2012) The variation in flowering time presumably

due to polyploidy, generation of interploid hybrids (Nishiwaki *et al*., 2011) and/or the paleopolyploidzation event in *Miscanthus* may help explain the wide geographical and latitudinal distribution of this genus. The differences in flowering time could create reproductive isolation of new polyploids from their progenitors (Hegarty & Hiscock, 2008) and allow neopolyploids to enter new ecological niches (Otto, 2007). The natural allopolyploid *Arabidopsis suecica* displays later flowering than the artificially generated allopolyploid (Wang *et al*., 2006) suggesting that phenotypic change related to allopolyploid events is subject to natural selection (Chen, 2007). Variation in the geographic distribution of diploid and tetraploid cytotypes of *Anthoxanthum alpinum* also appears to relate to differences in flowering phenology (Felber, 1988; Tate *et al*., 2005).

Synthetic polyploids provide many opportunities for crop improvement in a conventional breeding program. Interspecific hybrids between distant taxa (different species or genera) are often sterile because of the failure chromosomes to properly pair during meiosis. Genome doubling is a technique to restore the fertility of these interspecific hybrids by providing homologous chromosomal duplicates (Ramsey & Schemske, 2002). Fertile allopolyploids have been produced from genome doubling of sterile hybrids of related plant species (Thomas, 1993; Nimura *et al*., 2006). Sterility associated with triploidy can also be overcome by genome doubling. Restoration of pollen viability and seedset was observed from oryzalin-induced hexaploids derived from sterile triploid interspecific rose hybrids (Kermani *et al*., 2003). Loss of self-incompatibility after polyploidization events has been observed in other species (Miller & Venable, 2000) and could overcome self-incompatibility in *Miscanthus* species, opening opportunities to uncover genetic variability in this obligate outcrossing genus (Sacks *et al*., 2012). These opportunities can be applied to *Miscanthus* breeding programs, where natural (Sobral *et*

al., 1994) and artificial intergeneric hybrids (Park *et al*., 2011) exist and where self-incompatibility presents barriers to gene flow and the generation of inbred lines.

The establishment of triploid *M.* ×*giganteus* production fields requires vegetative propagation via rhizome divisions or tissue culture, which would be expensive relative to seed propagation if the latter option were made available (Lewandowski *et al*., 2000). For rhizome harvest, cleaning, separation and replanting, it has been estimated that one hectare of mature *M.* ×*giganteus* (3 or more years old) will provide sufficient rhizomes for the planting of approximately only ten hectares of new production (T. Voigt, personal communication). This is a significant constraint in the development of a bioenergy cropping system that will need to have the capacity to rapidly upscale to fulfill large production requirements. In contrast, seed-based propagation of *Miscanthus* should be much cheaper, scalable to meet production needs, and utilize sexual hybridization and selection to generate diverse and improved germplasm for commercial production. Individual culm inflorescences of *Miscanthus* accessions are known to produce hundreds or even thousands of seeds.

The impetus for the seed propagation of *Miscanthus*, however, impacts on its potential invasiveness. The putative parental species of *M. x giganteus*, *M. sacchariflorus* and *M. sinensis*, are not native to the United States and the former is designated as invasive in several states with the latter considered as putatively invasive (Quinn *et al*., 2010). The production of allotriploid seeds from crosses between diploid and tetraploid *M. sinensis* and *M. sacchariflorus* can circumvent invasive issues by generating sterile non-invasive plants. Our recently developed polyploids derived from *M. sinensis* and *M. sacchariflorus* could be used to broaden the genetic base of *M. x giganteus*, thereby allowing for further exploitation of inter-specific heterosis. If generation of triploid and pentaploid seeds (from crosses between hexaploid *M. x giganteus* and

tetraploid accessions of *M. sinensis* and *M. sacchariflorus*) is feasible on a scale amenable to commercial seed production, this could result in seed-propagated production of cultivars that are potentially sterile.

Genome doubling could be also used for circumventing invasive issues by delaying flowering time. Most of our hexaploids did not complete flowering under the environmental condition found in Central Illinois although viable pollen was produced from these hexaploid lines in the greenhouse (Figure 4.4). Late flowering in *Miscanthus* is associated with higher biomass yields (Clifton-Brown *et al*., 2001). Seeds of late-flowering and fertile polyploids could be harvested in southern latitudes and planted in temperate areas where growing seasons are not long enough for the initiation or completion of the flowering process. Such a production system would reduce risks of *Miscanthus* invasion.

The synthetic polyploid lines from this study can also be used to study neopolyploidization; an advantage over natural polyploids from which their corresponding progenitors are often unknown. Investigating the association between transcriptional and post-transcriptional regulation on phenotypic variation could help to reveal the mechanisms underlying neopolyploidization in *Miscanthus* and other species. Studies are currently underway to detect changes in gene expression and investigate variation in small RNA mediated gene silencing between synthetic polyploids and their corresponding progenitors.

Table 4.1. Survival of calli after oryzalin treatment and plant regeneration of *Miscanthus* accessions and their ploidy levels.

Species	Genotype	Plant materials	No. callus treated /survived (%)	No. plants regenerated		
				Total	Di- or triploids	Synthetic polyploids (%)
M. sinensis	Early hybrid	Seed	378/322 (85.2)	162	43	119 (73.5)
	New hybrid	Seed	378/370 (97.9)	30	20	10 (33.3)
	Grosse fontaine	Immature inflorescence	347/73 (21.0)	69	28	31 (44.9)
	Undine	Immature inflorescence	597/102 (17.1)	-	-	-
M. sacchariflorus	Blue Stem	Immature inflorescence	128/79 (61.7)	-	-	-
	Golf course	Immature inflorescence	318/250 (78.6)	59	36	23 (39.0)
	Robustus	Immature inflorescence	171/142 (83.4)	-	-	-
M. x giganteus	Illinois	Immature inflorescence	678/629 (92.8)	194	135	59 (30.4)

Table 4.2. 2C values of nuclear DNA content and chromosome numbers of regenerated and synthetic autopolyploid lines used for greenhouse investigation.

Species	Ploidy	Entry	Sources of callus	nDNA Contents (pg) [z]	No. chr.
M. sacchariflorus 'Golf course'	2x	GC2xR-01	Regenerated	4.44 ± 0.05 b	38
	2x	GC2xO-02	Orizalin-treated	4.47 ± 0.01 b	-
	4x	GC4x-02	Orizalin-treated	9.00 ± 0.34 a	76
	4x	GC4x-04	Orizalin-treated	9.03 ± 0.14 a	-
	4x	GC4x-05	Orizalin-treated	9.02 ± 0.42 a	-
	4x	GC4x-07	Orizalin-treated	8.87 ± 0.28 a	-
	4x	GC4x-10	Orizalin-treated	8.85 ± 0.06 a	-
M. sinensis 'Grosse fontaine'	2x	GF2xR-06	Regenerated	5.28 ± 0.10 c	38
	2x	GF2xO-01	Orizalin-treated	5.42 ± 0.06 c	-
	4x	GF4x-02	Orizalin-treated	10.72 ± 0.34 ab	-
	4x	GF4x-03	Orizalin-treated	11.10 ± 0.24 a	-
	4x	GF4x-04	Orizalin-treated	10.55 ± 0.34 b	76
	4x	GF4x-05	Orizalin-treated	10.47 ± 0.41 b	-
	4x	GF4x-11	Orizalin-treated	10.73 ± 0.14 ab	-
M. x giganteus 'Illinois'	3x	Mxg3xR-02	Regenerated	7.00 ± 0.06 d	57
	3x	Mxg3xO-03	Orizalin-treated	6.75 ± 0.24 d	-
	6x	Mxg6x-01	Orizalin-treated	14.18 ± 0.26 a	-
	5x	Mxg5x-02	Orizalin-treated	12.65 ± 0.59 c	95*
	6x	Mxg6x-04	Orizalin-treated	13.43 ± 0.32 b	114*
	6x	Mxg6x-07	Orizalin-treated	13.43 ± 0.63 b	-
	6x	Mxg6x-20	Orizalin-treated	13.70 ± 0.07 ab	-

[z] means levels with different letters are significantly different at $P < 0.001$.

* Some measurements of the number of chromosomes varied among counts due to overlaying chromosomes.

Table 4.3. Phenotypic differences in number of tillers, plant height, flowering time and stem diameter between the means of regenerated lines and five synthetic autopolyploid lines of the three *Miscanthus* species.

Lines	No. tillers	Plant height (cm)	Flowering (Week)	Stem diameter (mm)
M. sacchariflorus				
Golf course 2x	30.0 ± 3.61 ***	139.5 ± 5.33 **	9.2 ± 0.29 NS	2.6 ± 0.10 NS
Golf course 4x	10.1 ± 4.91	120.0 ± 6.31	9.7 ± 0.83	2.6 ± 0.27
M. sinensis				
Grosse fontaine 2x	27.7 ± 4.65 *	204.3 ± 11.22 NS	14.7 ± 0.29	4.8 ± 0.10 NS
Grosse fontaine 4x	17.3 ± 5.42	159.4 ± 34.57	15.5 ± 0.38 *	4.9 ± 0.64
M. x giganteus				
Illinois 3x	28.3 ± 1.53 NS	204.7 ± 14.16 NS	15.8 ± 0.29 NS	4.9 ± 0.42
Illinois 6x	16.1 ± 10.22	153.4 ± 40.14	20.9 ± 6.92	5.9 ± 0.45 *

t-tests were performed between calculated means from 2 progenitor lines and those from the synthetic polyploid lines of each species. *, ** and *** indicate significant differences based on t-test at <0.05, <0.01, and <0.001, respectively. NS indicates no significant differences.

Table 4.4. Phenotypic variations in number of tillers, plant height, flowering time and stem diameter of regenerated and synthetic autopolyploid lines in three *Miscanthus* species.

Ploidy	Lines	No. tillers	Plant height (cm)	Flowering (Week)	Stem diameter (mm)
Miscanthus sacchariflorus 'Golf course'					
2x	GC2xR-01	33.7 ± 1.53 a[z]	132.5 ±18.42 a	8.7 ± 0.58 a	2.6 ± 0.19 a
2x	GC2xO-02	26.3 ± 8.02 b	146.5 ±10.27 a	9.7 ± 0.58 a	2.7 ± 0.00 a
4x	GC4x-02	4.7 ± 1.53 d	113.5 ±16.91 a	10.0 ± 2.00 a	2.2 ± 0.69 a
4x	GC4x-04	11.3 ± 3.21 cd	120.0 ±14.93 a	9.7 ± 1.15 a	2.4 ± 0.51 a
4x	GC4x-05	13.0 ± 1.00 c	117.3 ±37.51 a	10.3 ± 1.53 a	2.6 ± 0.69 a
4x	GC4x-07	16.0 ± 4.58 c	130.4 ±8.92 a	10.3 ± 2.08 a	2.9 ± 0.19 a
4x	GC4x-10	5.3 ± 3.21 d	119.0 ±12.40 a	8.3 ± 0.58 a	2.8 ± 0.38 a
Miscanthus sinensis 'Grosse fontaine'					
2x	GF2xR-06	25.7 ± 4.73 a	194.3 ± 13.38 ab	14.7 ± 0.58 a	4.6 ± 0.51 b
2x	GF2xO-01	29.7 ± 5.86 a	214.2 ± 12.78 a	14.7 ± 0.58 a	5.0 ± 0.33 ab
4x	GF4x-02	15.7 ± 2.52 cd	164.7 ± 29.58 c	15.7 ± 1.15 a	5.9 ± 0.51 a
4x	GF4x-03	18.0 ± 4.00 bc	173.6 ± 13.22 bc	15.3 ± 0.58 a	5.2 ± 0.84 ab
4x	GF4x-04	18.7 ± 5.03 bc	174.0 ± 11.43 bc	15.3 ± 0.58 a	4.8 ± 1.07 ab
4x	GF4x-05	9.7 ± 0.58 d	99.1 ± 11.64 d	16.0 ± 1.41 a	4.4 ± 0.69 b
4x	GF4x-11	24.7 ± 2.31 ab	185.8 ± 1.94 bc	15.0 ± 0.00 a	4.3 ± 0.33 b
Miscanthus x giganteus 'Illinois'					
3x	Mxg3xR-02	21.3 ± 3.06 b	195.2 ± 7.13 a	15.7 ± 0.58 d	5.1 ± 0.69 a
3x	Mxg3xO-03	35.3 ± 4.62 a	214.2 ± 45.34 a	16.0 ± 0.00 d	4.7 ± 0.33 a
6x	Mxg6x-01	10.7 ± 3.06 c	92.7 ± 20.91 c	33.0 ± 0.00 a	5.8 ± 0.70 a
5x	Mxg5x-02	16.7 ± 3.21 bc	175.7 ± 34.37 ab	20.0 ± 1.73 b	6.4 ± 1.39 a
6x	Mxg6x-03	9.3 ± 4.62 c	132.5 ± 29.58 bc	18.0 ± 1.41 c	5.6 ± 1.84 a
6x	Mxg6x-07	10.3 ± 5.03 c	188.8 ± 38.88 a	16.7 ± 1.15 cd	6.3 ± 0.00 a
6x	Mxg6x-20	33.7 ±11.59 a	177.4 ± 10.65 ab	16.7 ± 0.58 cd	5.4 ± 1.17 a

[z] means levels within a column with a different letters are significantly different at $P < 0.001$.

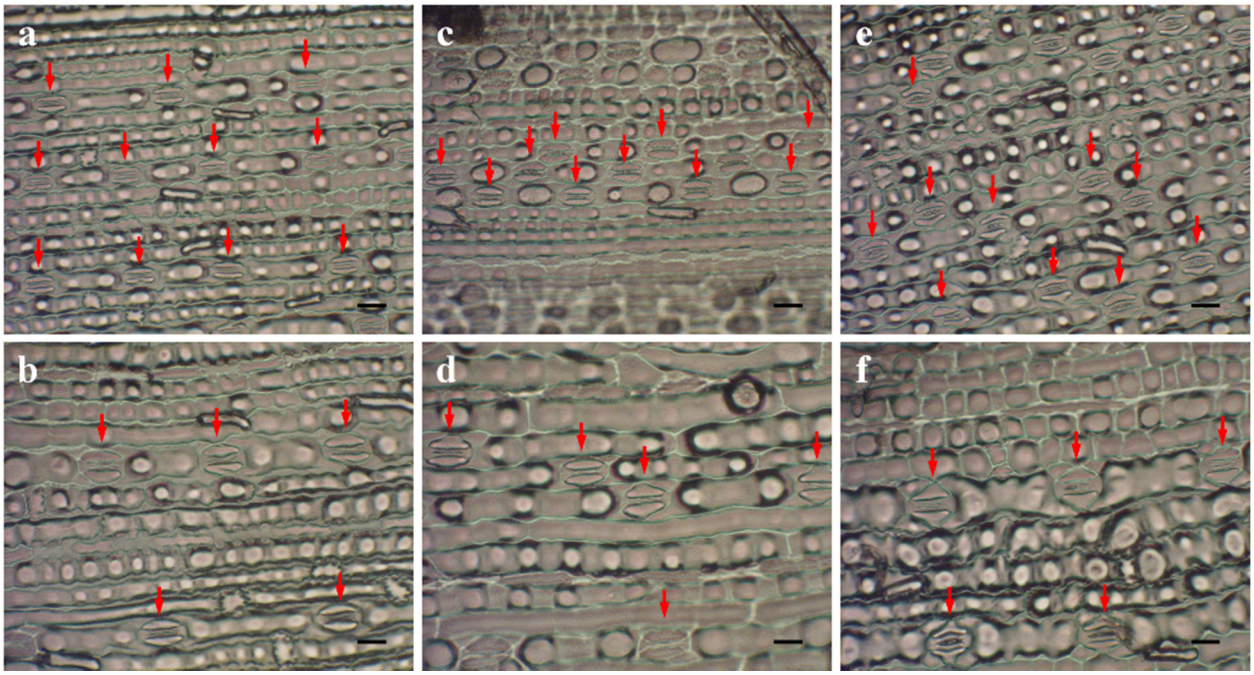

Figure 4.1. Leaf surface of progenitor (upper) and synthetic autopolyploid plants (below) in 3 *Miscanthus* species. Scale bars are 20 µm in length and arrows point to guard cells. (a) and (b) *M. sacchariflorus* 'Golf course' diploid and synthetic polyploid; (c) and (d) *M. sinensis* 'Grosse fontaine' diploid and synthetic polyploid; (e) and (f) *M. x giganteus* 'Illinois' triploid and synthetic polyploid.

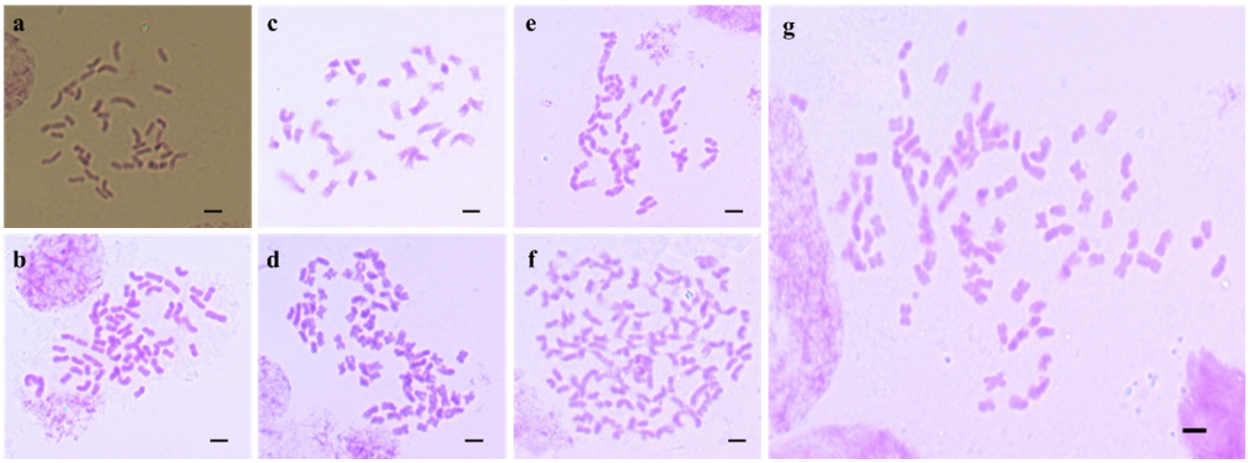

Figure 4.2. Chromosomes of regenerated and synthetic polyploid plants in 3 *Miscanthus* species. Scale bars = 10 µm. (a) *M. sacchariflorus* 'Golf course' diploid (2n=2x=38) and (b) tetraploid (2n=4x=76); (c) *M. sinensis* 'Grosse fontaine' diploid (2n=2x=38) and (d) tetraploid (2n=4x=76); (e) *M. x giganteus* 'Illinois' triploid (2n=3x=57) and (f) hexaploid (2n=6x=114); (g) *M. x giganteus* 'Illinois' pentaploid (5x=95).

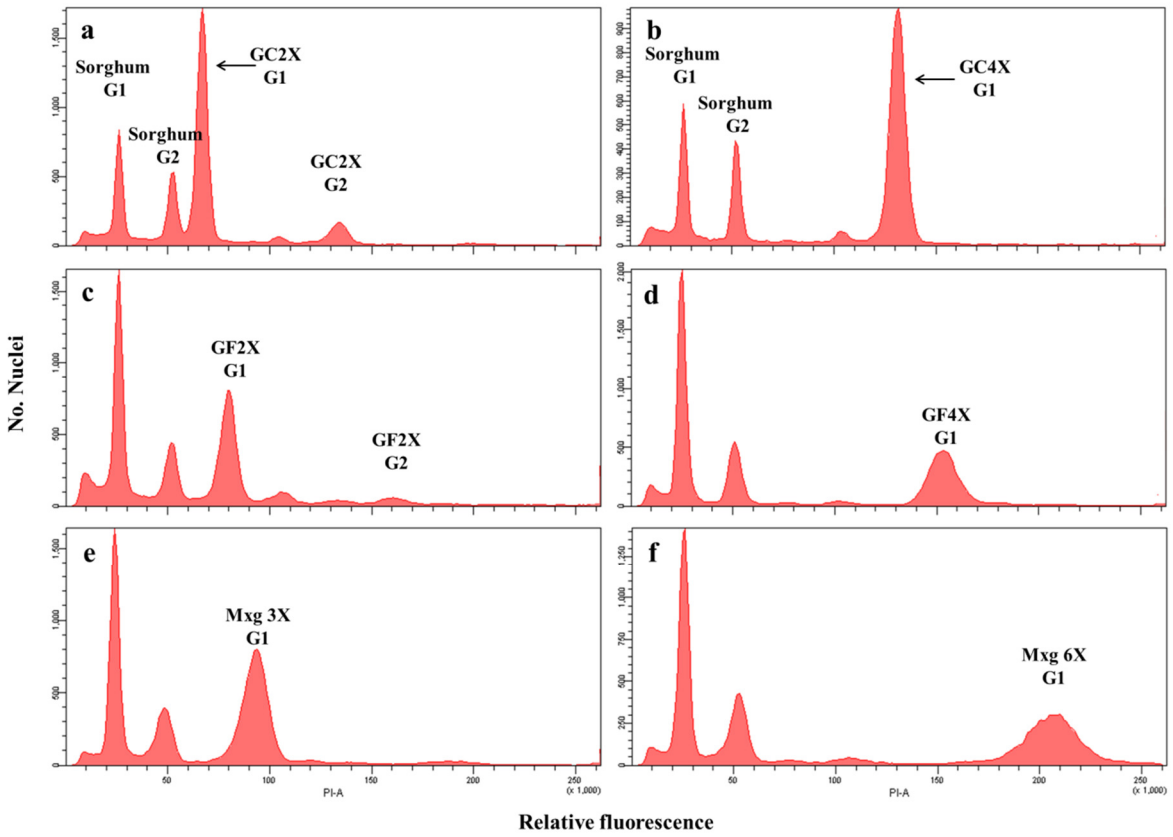

Figure 4.3. Histogram of nuclei extracted from leaf tissue of regenerated controls and chromosome doubled polyploid lines from explants of immature inflorescence tissue of three *Miscanthus* species. Nuclei were stained with propidium iodide and sorghum was used as an internal standard. *M. sacchariflorus* 'Golf course' 2x (a) and 4x (b); *M. sinensis* 'Grosse fontaine'2x (c) and 4x (d); *M. x giganteus* 'Illinois'3x (e) and 6x (f).

Figure 4.4. Synthetic polyploid plants in the field and greenhouse. (a) Synthetic tetraploid (left) and regenerated diploid control (right) plants of *M. sacchariflorus*; (b) Synthetic tetraploid (right) and regenerated diploid control (left) plants of *M. sinensis*; (C) Synthetic hexaploid (left) and regenerated triploid control (right) plants of *M. x giganteus.* Photos for field and greenhouse plants were taken at October and at May, 2011, respectively.

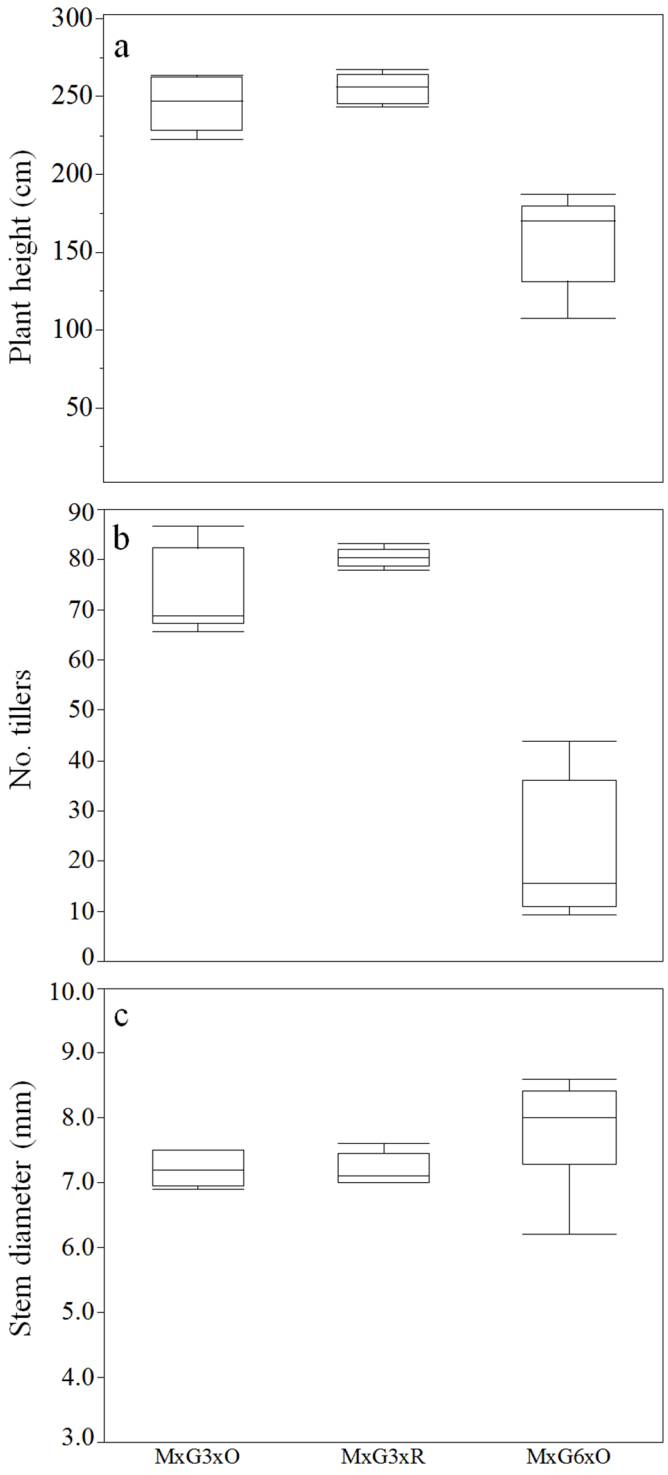

Figure 4.5. Box plot showing median (line), interquartile range (boxes), and 5% to 95% percentile (whiskers) for plant height (A), the number of tillers (B) and stem diameter (C) in regenerated control (triploid, Mxg3xR), oryzalin-treated and regenerated control (triploid, Mxg3xO) and synthetic autopolyploid lines (hexaploid, Mxg 6xO) of *M. x giganteus* grown in field condition of second year field plots.

LITERATURE CITED

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.

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* **168**, 2217–2226.

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, **8**, 135–141.

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

Arro A, Veremis JC, Kimbeng CA, Botanga C (2006) Genetic diversity and relationships revealed by AFLP markers among *Saccharum spontaneum* and realted species and genera. *Journal American Society Sugar Cane Technologists*, **26**, 101–115.

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martin A (2002) Preliminary genetic linkage map of *Miscanthus sinensis* with RAPD markers. *Theoretical and Applied Genetics*, **105**, 946–952.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Cambell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanic Garden*, **82**, 247–277.

Bennet MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. *Annals of Botany*, **76**, 113–176.

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London B*, **274**, 227–274.

Bennett MD, Leitch IJ (2005) Genome size evolution in plants. In: *The evolution of the genome* (ed Gregory TR), pp. 89-162. Elsevier Academic Press, San Diego, California.

Birol I, Jackman SD, Nielsen CB *et al*. (2009) De novo transcriptome assembly with ABySS. *Bioinformatics*, **25**, 2872–2877.

Boggs CL (2001) Species and speciation. In : *Internationnal Encyclopedia of the Social & Behavioral Sciences* (eds Smelser NJ, Baltes PB), pp.14855–14861.

Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J (2009) The "inner circle" of the cereal genomes. *Current Opinion in Plant Biology*, **12**, 119–125.

Borrino EM, Powell W (1988) Stoamtal guard cell length as an indicator of ploidy in microspore-derived plants of barley. *Genome*, **30**, 158–160.

Brand MH (1999) Small divisions of ornamental grasses produce the best growth following direct potting. *Hortscience*, **34**, 1126–1128.

Brandes E (1956) Origin, dispersal and use in breeding of the Melanesian garden sugarcane and their derivatives, *Saccharum officinarum* L. *Proceedings of the International Society of Sugar Cane Technologist*s, **9**, 709–750.

Bradshaw JD, Prasifka JR, Steffey KL., Gray ME (2010) First report of field populations of two potential aphid pests of the bioenergy crop *Miscanthus x giganteus*. *The Florida Entomologist*, **93**, 135–137.

Burner DM (1991) Cytogenetic analyses of sugarcane relatives (Andropogoneae: Saccharinae). *Euphytica*, **54**, 125-133.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity*, **84**, 441-451.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Carroll A, Somerville C (2009) Cellulosic Biofuels. *Annual Review of Plant Biology*, **60**, 165-182.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540-552.

Chakravarti A, Lasher LK, Reefer JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics*, **128**, 175–182.

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

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology*, **58**, 377-406.

Christian DG, Bullard MJ, Wilkins C (1997) The agronomy of some herbaceous crops grown for energy in southern England. *Aspects of Applied Biology, Biomass and Energy Crops*, **49**, 41–52.

Christian DG, Haase E (2001) Agronomy of *Miscanthus*. *In*: *Miscanthus for Energy and Fibre* (eds Jones MB, Walsh M), pp. 21–45. James & James, London.

Clayton WD, Renvoize SA (1986) Genera Graminum, grasses of the world. *Kew Bulletin Additional Series XIII*.

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

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), pp273–294. 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, 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.

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Science*, **160**, 1115–1123.

Daniels J, Roach BT (1987) Taxonomy and evolution. *In: Sugarcane improvement through breeding* (ed Heinz DJ), pp 7–84. Elsevier, New York.

Deuter M (2000) Breeding approaches to improvement of yield and quality in *Miscanthus* grown in Europe. EMI Project, Final report, pp. 28–52.

D'Hont A, Glaszmann JC (2001) Sugarcane genome analysis with molecular markers: a first decade of research. *International Society of Sugar Cane Technologists*, **24,** 556-559.

D'Hont A, Rao PS, Feldmann P., Grivet L, Islam-Faridi N, Taylor P, Glaszmann JC (1995) Identification and characterization of sugarcane intergeneric hybrids, *Saccharum officinarum* × *Erianthus arundinaceus*, with molecular markers and DNA in situ hybridization. *Theoratical and Applied Genetics*, **91**, 320–326.

D'Hont A, Ison D, Alix K, Roux C (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome*, **225**, 221–225.

Dolezel J (1997) Application of flow cytometry for the study of plant genomes. *Journal of Applied Genetics*, **38**, 285–302.

Dolezel J, Bartos J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size in trout and human. *Cytometry A*, **51**, 127–128.

Dolezel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry *Nature protocols*, **2**, 2233–2244.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics*, **42**, 443–461.

Draper J, Mur L, Jenkins G (2001) *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiology*, **127**, 1539–1555.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Feldman M, Levy AA (2009) Genome evolution in allopolyploid wheat–a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics*, **36**, 511–518.

Felber F (1988) Phénologie de la floraison de populations diploïdes et tétraploïdes d'*Anthoxanthum alpinum* et d'*Anthoxanthum odoratum*. *Canadian Journal of Botany*, **66**, 2258–2264.

Gale MD, Devos KM (1998) Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences*, **95**, 1971–1974.

Ganal MW, Altmann T, Röder M (2009) SNP identification in crop plants. *Current Opinion in Plant Biology*, **12**, 211–217.

Głowacka K, Jeżowski S, Kaczmarek Z (2010) In vitro induction of polyploidy by colchicine treatment of shoots and preliminary characterisation of induced polyploids in two *Miscanthus* species. *Industrial Crops and Products*, **32**, 88–96.

Grant V (1981) Plant Speciation, 2nd ed. New York: Columbia Univ. Press.

Greef JM, Deuter M, Jung C, Schondelmaier J (1997) Genetic diversity of European *Miscanthus* species revealed by AFLP fingerprinting. *Genetic Resources and Crop Evolution*, **44**, 185–195.

Greilhuber J, Dolezel J, Lysak M, Bennet MD (2005) The origin, evolution and proposed stabilization of the terms 'Genome size' and C-value' to describe nuclear DNA contents. *Annals of Botany*, **95**, 255–260.

Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, **96,** 241–252.

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science*, **80**, 524–535.

Ha M, Lu J, Tian L *et al*. (2009) Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences*, **106**, 17835–17840.

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of *Miscanthus*. *Global Change Biology*, **14**, 2000–2014.

Heaton EA, Dohleman FG, Miguez AF *et al*. (2010) *Miscanthus*: A promising biomass crop. *Advances in Botanical Research*, **56**, 75–137.

Heaton E, Voigt T, Long SP (2004) A quantitative review comparing the yields of two candidate C_4 perennial biomass crops in relation to nitrogen temperature and water. *Biomass and Bioenergy*, **27**, 21–30.

Hegarty MJ, Hiscock SJ (2008) Genomic clues to the evolutionary success of polyploid plants. *Current Biology*, **18**, R435–R444.

Hernández P, Dorado G, Laurie DA, Martín A, Snape JW (2001) Microsatellites and RFLP probes from maize are efficient sources of molecular markers for the biomass energy crop *Miscanthus*. *Theoretical and Applied Genetics*, **102**, 616–622.

Hirayoshi I, Nishikawa K, Hakura A (1960) Cytogenetical studies on forage plants Ⅷ. 3x- and 4x-hybrids raised from the cross *Miscanthus sinensis var. condensatus* x *M. sacchariflorus*. *Research bulletin of the Faculty of Agriculture, Gifu University,* **12**, 82–88.

Hirayoshi I, Nishikawa K, Kubono M, Sakaida T (1959) Cyto-genetical studies on forage plants. VII. Chromosome conjugation and fertility in *Miscanthus* hybrids including *M. sinensis M. sinensis var. condensatus* and *M. tinctorius*. *Research bulletin of the Faculty of Agriculture, Gifu University,* **11**, 86–91.

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, Revoize 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 (2001) Nomenclature of *Miscanthus x giganteus* (Poaceae). *Kew Bulletin*, **56**, 759–760.

Hodkinson TR, Renvioze SA, Chase MW (1997) Systematics of *Miscanthus*. *In: Biomass and Bioenergy crops* (eds Bullard M J *et al*.) Aspects of Biology 49. Association of Applied Biologist. Warwick, UK. pp 189–197.

Honda M (1930) Monographia Poacearum Japonicarum, bambusoideis exclusis. *Journal of Faculty of Science, Imperial University of Tokyo, Section III, Botany* **3,** 1–432.

Hsiao C, Chatterton NJ, Asay KH, Jensen KB (1994) Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome*, **37**, 112–120.

Huelsenbeck J, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

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

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, **44**, 223–270.

Jakob K, Zhou F, Paterson AH (2009) Genetic improvement of C_4 grasses as cellulosic biofuel feedstocks. *In Vitro Cellular and Developmental Biology-Plant*, **45**, 291–305.

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*, DOI: 10.1007/s11032-011-9651-1

Jones N, Ougham H, Tomas H (1997) Markers and mapping : we are all geneticists now. *New Phytologist*, **137**, 165–177.

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, **160,** 1651–1659.

Kent WJ (2002) BLAT–The BLAST-like alignment tool. *Genome Research*, **12**, 656–664.

Kermani MJ, Sarasan V, Roberts AV, Yokoya K, Wentworth J, Sieber VK (2003) Oryzalin-induced chromosome doubling in Rosa and its effect on plant morphology and pollen viability. *Theoretical and Applied Genetics*, **107**, 1195–1200.

Khanna M, Dhungana B, Clifton-Brown J (2008) Costs of producing miscanthus and switchgrass for bioenergy in Illinois. *Biomass and Bioenergy*, **32**, 482–493.

Khazaei H, Monneveux P, Hongbo S, Mohammady S (2010) Variation for stomatal characteristics and water use efficiency among diploid, tetraploid and hexaploid Iranian wheat landraces. *Genetic Resources and Crop Evolution*, **57**, 307–314.

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.

Kim HS, Zhang G, Juvik JA, Widholm JA (2010) *Miscanthus* × *gigant*eus plant regeneration: effect of callus types, ages and culture methods on regeneration competence. *Global Change Biology Bioenergy*, **2**, 192–200.

Knapik EW, Goodman A, Atkinson OS *et al*. (1996) A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms. *Development*, **123**, 451–460.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, **25**, 2283–2285.

Laat D, Godhe W, Vogelzang MJ (1987) Determination of ploidy of single plants and plant populations by flow cytometry. *Plant Breeding*, **99**, 303–307.

Lafferty J, Lelley T (1994) Cytogenetic studies of different *Miscanthus* species with potential for agricultural use. *Plant Breeding*, **113**, 246–249.

Langmead B (2010) Aligning short sequencing reads with Bowtie. *Current Protocol in Bioinformatics*, **11**, 1–24.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.

Lee YN (1964a) Taxonomic studies on the genus *Miscanthus*: relationships among the section, subsection and species, part 1. *Journal of Japanese Botany*, **39**, 196–205.

Lee YN (1964b) Taxonomic studies on the genus *Miscanthus*: relationships among the section, subsection and species, part 2. *Journal of Japanese Botany*, **39**, 257–265.

Lee YN (1964c) Taxonomic studies on the genus *Miscanthus*: relationships among the section, subsection and species, part 3. *Journal of Japanese Botany*, **39**, 257–265.

Lee YN (1993) Manual of the Korean grasses. Ewha Wonmans University Press, Seoul, Korea.

Levin, DA (2002) The role of chromosomal change in plant evolution. pp14–16. Oxford University press. Inc. New York.

Levy AA, Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon alloploidization. *Biological Journal of the Linnean Society*, **82**, 607–613.

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, Scurlock JMO, Lindvall E, Christou M (2003) The development and current status of potential rhizomatous grasses as energy crops in the US and Europe. *Biomass and Bioenergy*, **25**, 335–361.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Linde-Laursen IB (1993) Cytogenetic analysis of *Miscanthus 'Giganteus'*, an interspecific hybrid. *Hereditas*, **119**, 297–300.

Liu B, Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution*, **29**, 365–379.

Luo MC, Deal KR, Akhunov ED *et al*. (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proceedings of the National Academy of Science*, **106**, 15780–15785.

Lysak MA, Dolezel J (1998) Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia*, **51**, 123–132.

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, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biology*, **9**, 13.

Mallet J (2007) Hybrid speciation. *Nature*, **446,** 279–283.

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of Angiosperms. *Science*, **264**, 421–-424.

Matzke MA, Matzke AJM (1998) Polyploidy and transposons. *Trends in Ecology and Evolution*, **13**, 241.

Mayr E (1942) Systematics and the origin of species. New York: Columbia University Press.

McLaughlin SB, Walsh ME (1998) Evaluating environmental consequences of producing herbaceous crops for bioenergy. *Biomass and Bioenergy*, **14**, 317–324.

McMurphy LM, Rayburn AL (1991) Genome size variation in maize populations selected for cold tolerance. *Plant Breeding*, **106**, 190–195.

Mehra PN, Sharma ML (1975) Cytological studies in some central and eastern Himalayan grasses: 1 the Andropogoneae. *Cytologia*, **40**, 61–74.

Miller JS, Venable DL (2000) Polyploidy and the evolution of gender dimorphism in plants. *Science*, **289**, 2335–2338.

Ming R, Liu SC, Lin YR *et al*. (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics*, **150**, 1663–1682.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiologia Plantarum*, **15**, 473–497.

Naidu SL, Moose SP, AL-Shoaibi AK, Raines CA, Long SP (2003) Cold tolerance of C_4 photosynthesis in *Miscanthus x giganteus*: adaptation in amounts and sequence of C_4 photosynthetic enzymes. *Plant Physiology*, **132**, 1688–1697.

Nelson JM, Lane B, Freeling M (2002) Expression of a mutant maize gene in the ventral leaf epidermis is sufficient to signal a switch of the dorsoventral axis. *Development*, **129**, 4581–4589.

Ng DW, Lu J, Chen ZJ (2012) Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Current Opinion in Plant Biology*, **15**, 154–161.

Nimura M, Kato J, Horaguchi H, Mii M, Sakai K, Katoh T (2006) Induction of fertile amphidiploids by artificial chromosome-doubling in interspecific hybrid between *Dianthus caryophyllu*s L. and *D. japonicus* Thunb. *Breeding Science*, **56**, 303–310.

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.

Noggle GR (1946) The physiology of polyploidy in plants. I. Review of the literature. *Lloydia*, **9**, 153–173.

Nolte AW, Tautz D (2010) Understanding the onset of hybrid speciation. *Trends in Genetics*, **26**, 54–58.

Ohri D (1998) Genome size variation and plant systematic. *Annals of Botany*, **82 (supplement A)**, 75–83.

Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, Tobias CM (2010) Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics*, **185**, 745–760.

Otto SP (2007) The evolutionary consequences of polyploidy. *Cell*, **131**, 452–462.

Ozkan H, Tuna M, Arumuganathan K (2003) Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. *Journal of Heredity*, **94**, 260–264.

Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. *New Phytologist*, **186**, 5–17.

Park J, Yu Q, Gracia NS, Acuna GM, da Silva JA (2011) Development of new intergeneric cane hybrids, Miscanes, as a source of biomass feedstock for biofuel production. *Plant & Animal Genomes XIX Conference January 15-19, Town & Country Convention Center, San Diego, CA*

Paterson AH, Bowers JE, Bruggmann R *et al*. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Paterson AH, Freeling M, Tang H, Wang X (2010) Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*, **61**, 349–372.

Petersen KK (1997) Callus induction and plant regeneration in *Miscanthus x ogiformis* Honda 'Giganteus' as influenced by benzyladenine. *Plant Cell, Tissue and Organ Culture*, **49**, 137–140.

Petersen KK, Hagberg P, Kristiansen K, Forkmann G (2002) In vitro chromosome doubling of *Miscanthus sinensis*. *Plant Breeding*, **121**, 445–450.

Petersen KK, Hagberg P, Kristiansen K (2003) Colchicine and oryzalin mediated chromosome doubling in different genotypes of *Miscanthus sinensis*. *Plant Cell, Tissue and Organ Culture*, **73**, 137–146.

Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biological Journal of the Linnean Society*, **82**, 675–688.

Pozzobon MT, Valls JFM (2000) Cytogeography and variation of stomatal size of *Paspalum Glaucescens* (*Gramineae*; *Paniceae*) in Southern Brazil. *Euphytica*, **116**, 251–256.

Prasifka JR, Bradshaw JD, Meagher RL, Nagoshi RN, Steffey KL, Gray ME (2009) Development and feeding of fall armyworm on *Miscanthus x giganteus* and switchgrass. *Journal of Economic Entomology*, **102**, 2154**–**2159.

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus *Sorghum* (Poaceae). *Annals of Botany*, **95**, 219–227.

Quinn LD, Allen DJ, Stewart JR (2010) Invasiveness potential of *Miscanthus sinensis*: implications for bioenergy production in the United States. *Global Change Biology Bioenergy*, **2**, 310–320.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, **5**, 94–100.

Ragauskas AJ *et al*. 2006. The path forward for biofuels and biomaterials. *Science* **311**, 484–489.

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, **33**, 589–639.

Randolph LF (1941) An evaluation of induced polyploidy as a method of breeding crop plants. *The American Naturalist*, **75**, 347–363.

Rayburn AL, Crawford J, Rayburn CM, Juvik JA (2009) Genome size of three *Miscanthus* species. *Plant Molecular Biology Reporter*, **27**, 184–188.

Rayburn AL, McCloskey R, Tatum TC, Bollero GA, Jeschke MR, Tranel PJ (2005) Genome size analysis of weedy *Amaranthus* species. *Crop Sciences*, **45**, 2557–2562.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systemics*, **28**, 359–389.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.

Rieseberg LH, Willis JH (2007) Plant speciation. *Science*, **317**, 910–914.

Rohlf F J (2002) NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System, Version 2.10z. Exeter Software, Setauket, New York, USA.

Sacks EJ, Juvik JA, Lin Q, Stewart JR, Yamada T (2012) The gene pool of *Miscanthus* species and its improvement. In: *Genomics of the Saccharinae* (ed Paterson AH) in press.

Salse J, Chagué V, Bolot S *et al*. (2008) New insights into the origin of the B genome of hexaploid wheat: evolutionary relationships at the SPA genomic region with the S genome of the diploid relative *Aegilops speltoides*. *BMC Genomics*, **9**, 555.

Salas Fernandez MG, Becraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. *Trends in Plant Sciences*, **14**, 454–461.

Scally L, Hodkinson TR, Jones MB (2001) Origins and taxonomy of *Miscanthus*. In: *Miscanthus for energy and fibre*. (eds Jones MB, Walsh M), pp. 1–9. James & James, London, UK.

Schranz ME, Osborn TC (2000) Novel flowering time variation in the resynthesized polyploid *Brassica napus*. *Journal of Heredity*, **91**, 242–246.

Sharma CBSR (1990) Chemically induced aneuploidy in higher plants. *Mutagenesis*, **5**, 105–126.

Sobral BWS, Braga DPV, LaHood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* Griseb. subribe of the *Andropogoneae* Dumort. tribe. *Theoretical and Applied Genetics*, **87**, 843–853.

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annual Review of Plant Biology*, **60**, 561–588.

Speckmann GJ, Post J, Dijkstra H (1965) The length of stomata as an indicator for polyploidy in rye-grasses. *Euphytica*, **14**, 225–230.

Spencer JL, Raghu S (2009) Refuge or reservoir? The potential impacts of the biofuel crop *Miscanthus x giganteus* on a major pest of maize. *PLoS ONE*, **4**, e8336.

Sreenivasan TV, Ahlowalia BS, Heinz DJ (1987) Cytogenetics. In: *Sugarcane improvement through breeding* (ed Heinz DJ), pp. 211-253. Elsevier Academic Press, Amsterdam, The Netherlands, Developments in Crip Science, no. 11.

Stebbins GL (1971) Chromosomal evolution in higher plants. London: Addision-Wesley.

Sun Q, Lin Q, Yi Z, Yang Z, Zhou F (2010) A taxonomic revision of *Miscanthus s. l.* (Poaceae) from China. *Botanical Journal of the Linnean Society*, **164**, 178–220.

Sun Y, Skinner DZ, Liang GH, Hulbert SH (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, **89**, 26–32.

Swaminathan K, Chae W, Mitros T *et al*. (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics* **13**, 142.

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 doi:10.1186/gb-2010-11-2-r12.

Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. *Genome Research*, **14**, 1916–1923.

Swift H (1950) The constancy of deoxyribose nucleic acid in plant nuclei. *Genetics*, **36**, 123–127.

Tate JA, Soltis DE, Soltis PS. 2005. Polyploidy in plants. In: *The evolution of the genome* (ed Gregory TR), pp. 371–426. Elsevier Academic Press, San Diego, California.

Tatum TC, Nunez L, Kushad MM, Rayburn AL (2006) Genome size variation in pumpkin (*Cucurbita sp*.). *Annals of Applied Botany*, **149**, 145–151.

Te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2012) The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**, 19-45.

Thiel T, Graner A, Waugh R, Grosse I, Close TJ, Stein N (2009) Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. *BMC Evolutionary Biology*, **9**, 209.

Thomas H (1993) Chromosome manipulation and polyploidy. In *Plant Breeding: Principles and Prospects* (eds Hayward MD, Bosemark NO, Romagosa T), pp. 79–92. Chapman and Hall Ltd, London.

Thompson JD, Lumaret R (1992) The evolutionary dynamics of polyploid plants: origins, establishment, and persistence. *Trends in Ecology and Evolution*, **7**, 302–307.

Torres EM, Williams BR, Amon A (2008) Aneuploidy: cells losing their balance. *Genetics*, **179**, 737–746.

Van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Research*, **93**, 343–349.

Van Ooijen JW, Voorrips RE (2001) JoinMap® version 3.0: software for the calculation of genetic linkage maps. Wageningen: Plant Research International.

Wang D, Portis AR, Moose SP, Long SP (2008) Cool C_4 photosynthesis: pyruvate Pi dikinase expression and activity corresponds to the exceptional cold tolerance of carbon assimilation in *Miscanthus x giganteus*. *Plant Physiology*, **148**, 557–567.

Wang J, Tian L, Lee HS, Chen ZJ (2006) Nonadditive regulation of *FRI* and *FLC* loci mediates flowering-time variation in Arabidopsis allopolyploids. *Genetics*, **173**, 965–974.

Wang X, 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.

Watson JV (1991) Introduction to flow cytometry. pp. 210–212. Cambridge University Press Cambridge, UK.

Wendel JF (2000) Genome evolution in polyploids. *Plant Molecular Biology*, **42**, 225–249.

Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ (2003) High-resolution genotyping by amplicon Melting analysis using LCGreen. *Clinical Chemistry*, **49**, 853–860.

Wolffe AP, Matzke MA (1999) Epigenetics: regulation through repression. *Science*, **286**, 481–486.

Wood JCS, Todd P (1979) Analysis of cellular DNA distribution. I. G2/G1 peak channel ratios. *Cell Biochemistry and Biophysics* **1**, 211–218.

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.

Zhang J, Nagai C, Yu Q, Pan Y, Ayala-Silva T, Schnell RJ, Comstock JC, Arumuganathan AK, Ming R (2012) Genome size variation in three Saccharum species. *Euphytica*, **185**, 511–519.