2015

Genetic variation in *Miscanthus* X *giganteus* and the importance of estimating genetic distance thresholds for differentiating clones

Katarzyna Glowacka
University of Illinois at Urbana-Champaign, kglowacka2@unl.edu

Lindsay V. Clark
University of Illinois at Urbana-Champaign

Shivani Adhikari
University of Illinois at Urbana-Champaign

Junhua Peng
Huazhong Agricultural University

J. Ryan Stewart
Brigham Young University

See next page for additional authors

Authors

Katarzyna Glowacka, Lindsay V. Clark, Shivani Adhikari, Junhua Peng, J. Ryan Stewart, Aya Nishiwaki, Toshihiko Yamada, Uffe Jorgensen, Trevor R. Hodkinson, Justin Gifford, John A. Juvik, and Erik J. Sacks

GLOBAL CHANGE BIOLOGY
BIOENERGY

GCB Bioenergy (2015) **7**, 386–404, doi: 10.1111/gcbb.12166

Genetic variation in *Miscanthus* × *giganteus* and the importance of estimating genetic distance thresholds for differentiating clones

KATARZYNA GŁOWACKA[1,2], LINDSAY V. CLARK[1], SHIVANI ADHIKARI[1], JUNHUA PENG[3], J. RYAN STEWART[4], AYA NISHIWAKI[5], TOSHIHIKO YAMADA[6], UFFE JØRGENSEN[7], TREVOR R. HODKINSON[8], JUSTIN GIFFORD[1], JOHN A. JUVIK[1] and ERIK J. SACKS[1]

[1]*Energy Biosciences Institute, University of Illinois, 1206 W. Gregory Dr., Urbana, IL 61801, USA,* [2]*Institute of Plant Genetics, Polish Academy of Sciences, ul. Strzeszyńska 34, Poznań 60-479, Poland,* [3]*College of Plant Science and Technology, Huazhong Agricultural University, No. 1, Shizishan Street, Wuhan, Hubei 430070, China,* [4]*Plant & Wildlife Sciences, Brigham Young University, 167 WIDB, Provo, UT 84602, USA,* [5]*Field Science Center, Faculty of Agriculture, University of Miyazaki, 1-1 Nishi Gakuenkibanadai, Miyazaki-shi Miyazaki-ken 889-2192, Japan,* [6]*Field Science Center for Northern Biosphere, Hokkaido University, Kita 11 Nishi 10, Kita-ku, Sapporo, Hokkaido 060-0811, Japan,* [7]*Dept. of Agroecology, Aarhus University, Blichers Allé 20, Postboks 50, Tjele DK-8830, Denmark,* [8]*School of Natural Sciences, Trinity College Dublin, Dublin D2, Ireland*

Abstract

Miscanthus × *giganteus* (Mxg) is an important bioenergy feedstock crop, however, genetic diversity among legacy cultivars may be severely constrained. Only one introduction from Japan to Denmark of this sterile, triploid, vegetatively propagated crop was recorded in the 1930s. We sought to determine if the Mxg cultivars in North America were all synonyms, and if they were derived from the European introduction. We used 64 nuclear and five chloroplast simple sequence repeat (SSR) markers to estimate genetic similarity for 27 Mxg accessions from North America, and compared them with six accessions from Europe, including the species' type-specimen. A subset of accessions was also evaluated by restriction-site associated DNA sequencing (RAD-seq). In addition, we assessed the potential of new crosses to increase Mxg genetic diversity by comparing eight new triploid Mxg progeny grown from seed, along with samples of the parental species *M. sacchariflorus* and *M. sinensis*. Estimates of genotyping error rates were essential for distinguishing between experimental error and true genotypic differences among accessions. Given differences in estimated error rates and costs per marker for SSRs and RAD-seq, the former is currently more cost-effective for determining if two accessions are genetically identical. We concluded that all of the Mxg legacy cultivars were derived via vegetative propagation from a single genet. In contrast with the Mxg legacy cultivars, genetic similarity to the type-specimen of eight new triploid Mxg progeny ranged from 0.46 to 0.56. Though genetic diversity among the Mxg legacy cultivars is critically low, new crosses can provide much-needed variation to growers.

Keywords: genetic diversity, genotyping error, interspecific hybrids, *Miscanthus sacchariflorus*, *Miscanthus sinensis*, RAD-seq, SSR

Received 7 September 2013 and accepted 19 October 2013

Introduction

Miscanthus × *giganteus* (Mxg) is a nothospiecies, derived from *M. sacchariflorus* (Msa) and *M. sinensis* (Msi) (Hodkinson & Renvoize, 2001; Hodkinson *et al.*, 2002b). In this article, we adhere strictly to the International Code of Nomenclature for algae, fungi, and plants (2012), especially with regard to nothospecies; thus, all progeny derived from crossing Msa and Msi are by definition Mxg. In Japan, indigenous populations of tetraploid Msa and diploid Msi are common (Hirayoshi *et al.*, 1957; Adati & Shiotani, 1962; Nishiwaki *et al.*, 2011). In southern Japan, sympatric populations of Msa and Msi infrequently produce interspecific triploid progeny (i.e. Mxg; Nishiwaki *et al.*, 2011; Dwiyanti *et al.*, 2013). Notably, one such Mxg was introduced from Yokohama Japan to Denmark in 1935 by Aksel Olsen (Greef & Deuter, 1993; Linde-Laursen, 1993). During the last 20 years, Mxg has become an important feedstock

Correspondence: Erik J. Sacks, tel. 217-333-9327, fax 217-244-3637, e-mail: esacks@illinois.edu

crop for the emerging bioenergy industries in Europe and the United States (Scurlock, 1999; Clifton-Brown *et al*., 2004; Heaton *et al*., 2008; Somerville *et al*., 2010). The USDA Biomass Crop Assistance Program (BCAP) alone supported the planting target of ~20 000 acres of Mxg during 2011 and 2012 (K. Novak, personal communication). Under the BCAP program, Mxg 'Illinois' was the predominant cultivar planted, with a smaller area planted to 'Freedom'.

Numerous accessions of sterile triploid Mxg of unknown provenance, which we call legacy cultivars, have been named in Europe and North America for commercial sale and scientific research, but how much genetic diversity is represented by these accessions is a key question. All of the legacy Mxg in Europe and North America are cultivars that have been maintained by people in public and private gardens, and agricultural fields; Mxg is neither native nor naturalized on these continents. Greef *et al*. (1997) evaluated 31 European accessions of Mxg with amplified fragment length polymorphism (AFLP) markers but found only two distinct genotypes that were 98% similar, indicating a striking lack of genetic diversity. Given error rates commonly observed for AFLPs (Douhovnikoff & Dodd, 2003; Bonin *et al*., 2004; Lasso, 2008), it is doubtful that the two Mxg groups observed by Greef *et al*. (1997) were truly different genotypes, although distinguishing between genotyping error and mutation (i.e. a horticultural sport) is difficult (Pompanon *et al*., 2005; Cipriani *et al*., 2010). Such a lack of genetic diversity represents a significant risk to growers, as the emergence of a single virulent pest or disease could damage or destroy all commercial production. If Mxg in North America was obtained solely from Europe, then we would expect genetic diversity to be similarly lacking in the United States and Canada. Recently, however, Chouvarine *et al*. (2012) reported that they identified six distinct Mxg genotypes in the United States, including 'Illinois' and 'Freedom', using Illumina transcriptome sequencing. However, the pairwise differences observed by Chouvarine *et al*. (2012) were on the order of one single nucleotide polymorphism (SNP) per 10 000 bp, which is within the expected error rate for this technology (Hedges *et al*., 2009; Dewey *et al*., 2012; Nielsen *et al*., 2012).

Genotyping errors are inherent to all molecular marker systems and they limit the inferences that can be made about relationships among individuals (Bonin *et al*., 2004; Pompanon *et al*., 2005; Minoche *et al*., 2011). Precision in genotype calling is especially important when the objective is to distinguish among individuals that have similar multilocus genotypes due to close kinship (e.g. full sibs), or for the even more challenging case of distinguishing among different somatic mutant lineages that originated from a single genet. However, technical replication enables estimation of genotyping error rates, and subsequent consensus calling has proven to be a valuable approach for mitigating these errors (Zhang *et al*., 2006; Christelova *et al*., 2011).

A key question about Mxg genetic diversity is: Are the legacy cultivars in North America derived from different and/or additional introductions (i.e. different genets) than the European genotype? Were multiple genotypes of Mxg introduced or was there only one? To resolve the seemingly incongruous conclusions of Greef *et al*. (1997) and Chouvarine *et al*. (2012), we tested the hypothesis, that Mxg in North America was derived from the European genotype, by estimating genetic similarity among a broad sample of accessions from North America, and comparing them with previously studied accessions from Europe, including the type-specimen for the species. Additionally, we assessed the potential of new Mxg progeny, from planned interspecific crosses and collections in the wild (Japan), to increase genetic diversity of this newly important bioenergy crop. We also explored the limits of inference in modern molecular marker systems, such as simple sequence repeats (SSRs) and restriction-site associated DNA sequencing (RAD-seq), for distinguishing different multilocus DNA fingerprints in *Miscanthus*.

Materials and methods

Plant material

We studied 85 *Miscanthus* accessions (Table 1), including 50 Mxg, 28 Msa, and seven Msi. Two sugarcane cultivars were included as an out-group. We compared the Mxg type-specimen, '1993–1780' (Hodkinson & Renvoize, 2001; live plant provided by Hodkinson), with 32 Mxg accessions collected from nurseries and scientific institutions in North America and Europe, eight colchicine-induced polyploids (5x-6x) of Mxg 'Illinois' (Chae *et al*., 2013), and eight new triploid Mxg genotypes obtained from seed. Included among the Mxg accessions were the leading commercial legacy cultivars in America, 'Illinois' and 'Freedom', and the European genotype 'EMI-1', of which the latter was included in the AFLP study by Greef *et al*. (1997) and in a subsequent phenotypic study at five European locations (Clifton-Brown *et al*., 2001). Two accessions, 'Illinois 2' and 'Illinois-1-1 (Mxg1-1)', were obtained via callus culture, a process that has the potential to produce somaclonal variants. New triploid Mxg genotypes from seed included four full-sibs from a cross (Msa 'Bluemel Giganteus' × Msi var. *condensatus* 'Cabaret') made at the University of Illinois (Chae *et al*., 2013), 'Nagara' from a cross made by M. Deuter (Tinplant; http://www.tinplant-gmbh.de), and three natural hybrids ('Ogi63', 'Ogi79' and 'Ogi80') obtained via seeds collected from a wild tetraploid Msa population that was sympatric with a diploid Msi population in southern Japan (Nishiwaki *et al*., 2011; Dwiyanti *et al*., 2013).

Table 1 *Miscanthus* and sugarcane genotypes in this study: their provenance, DNA content and genetic similarity to the *M.* × *giganteus* type-specimen '1993–1780' based on binary data of 527 alleles from 64 nuclear simple sequence repeat (SSR) loci (359 alleles, when considering only *Miscanthus* and excluding sugarcane)

Name	Accession identifier	Wild (w)/ cultivar (cv)	Locality of collection	Country of collection	Provider and prior sources	2C-DNA content (pg)	Ploidy	# of presence–absence differences	Jaccard's similarity coefficient
***M.* × *giganteus* (3x), legacy cv's**									
'1993–1780'	1993–1780	cv	England	UK	T. R. Hodkinson, Trinity College Dublin	6.92	3x	0	1.00
'Acton, Ontario'	pf1-32	cv	ON	Canada	NEF			0	1.00
'Bluestem'	UI10-00118	cv	WA	USA	Bluestem Nursery	6.95	3x	0	1.00
'Danish'	N1	cv	Hornum	Denmark	S. Jeżowski, IPG ← M. Deuter, Tinplant			0	1.00
'EMI-1'	LASEI1	cv	Midtjylland	Denmark	U. Jørgensen, Aarhus Univ. ← Larsen, Silkeborg		3x††	0	1.00
'Flower Factory'*	UI10-00001	cv	WI	USA	The Flower Factory nursery	6.92	3x	0	1.00
'Freedom'	pf1-15	cv	MS	USA	NEF ← B. Baldwin, MSU			0	1.00
'Freedom'	UI10-00119	cv	MS	USA	T. Voigt, UI ← B. Baldwin, MSU	6.94§	3x	0	1.00
'Georgia'	NEF-6	cv	GA	USA	NEF ← ornamental garden in GA			0	1.00
'German'	N2	cv	Saxony-Anhalt	Germany	S. Jeżowski, IPG ← M. Deuter, Tinplant		3x**	2	0.98
'GG1'	UI12-00012	cv		USA	NEF ← Southern USA	6.95	3x	0	1.00
'GG2'	UI12-00013	cv		USA	NEF ← Southern USA	7.08	3x	0	1.00
'Gilded Tower'	UI10-00120	cv	NC	USA	Plant Delights Nursery Inc.	6.99§	3x	0	1.00
'G-Max'	UI11-00026	cv	IL	USA	T. Voigt, UI	6.68	3x	0	1.00
'Hortico'	UI10-00121	cv	ON	Canada	Hortico Nurseries Inc.	6.93§	3x	0	1.00
'Illinois'	pf1-16	cv	IL	USA	NEF ← UI			0	1.00
'Illinois'	UI10-00107	cv	IL	USA	Speedling Inc ← T. Voigt, UI ← Chicago Botanic Garden	6.86§	3x	0	1.00
'Illinois 1'	NEF-1	cv	IL	USA	NEF ← UI			0	1.00
'Illinois 2'	NEF-2	cv	IL	USA	NEF ← commercial tissue culture laboratory ← UI			0	1.00
'Illinois-1-1 (Mxg1-1)'	UI10-00108	cv	IL	USA	W. Chae & J. Juvik, UI ← callus culture	7.07	3x	0	1.00
'John Amos Gig'	pf1-13	cv	Wales	UK	NEF ← Aberystwyth Univ.			0	1.00
'Long's Garden'	UI10-00125	cv	IL	USA	Long's Garden Inc. nursery	6.99§	3x	0	1.00
'McGuigan'	pf1-30	cv	ON	Canada	NEF ← Kent County Nursery			0	1.00
'Michigan'	NEF-3	cv	MI	USA	NEF ← selection from a garden in northern MI			0	1.00
'Mississippi Clone'	NEF-5	cv	MS	USA	NEF ← B. Baldwin, MSU			0	1.00
'NEF-7'	NEF-7	cv	ON	Canada	NEF			1	0.99
'NEF-8'	NEF-8	cv	ON	Canada	NEF			2	0.98
'NEF-9'	NEF-9	cv	ON	Canada	NEF			1	0.99
'Polish'	N3	cv		Poland	S. Jeżowski, IPG ← Botanic Garden of IHAR Bydgoszcz			3	0.98
'Roland'	pf1-12	cv	ON	Canada	NEF			0	1.00
'Rosahar's Garden'	pf1-33	cv	ON	Canada	NEF			0	1.00
'Walla Walla'	UI10-00122	cv	WA	USA	Walla Walla Nursery Company Inc.	6.96§	3x	0	1.00
'Wisconsin'	NEF-4	cv	WI	USA	NEF ← selection from an abandoned nursery			0	1.00
***M.* × *giganteus* colchicine-induced polyploids**									
'Illinois-5x.02 (Mxg2x-2)'	UI10-00110	cv	IL	USA	W. Chae & J. Juvik, UI	12.65¶	5x	0	1.00

(continued)

Table 1 (continued)

Name	Accession identifier	Wild (w)/ cultivar (cv)	Locality of collection	Country of collection	Provider and prior sources	2C-DNA content (pg)	Ploidy	# of presence–absence differences	Jaccard's similarity coefficient
'Illinois-6x.01 (Mxg2x-1)'	UI10-00109	cv	IL	USA	W. Chae & J. Juvik, UI	14.18¶	6x	0	1.00
'Illinois-6x.03 (Mxg2x-3)'	UI10-00111	cv	IL	USA	W. Chae & J. Juvik, UI	13.98	6x	0	1.00
'Illinois-6x.06 (Mxg2x-6)'	UI10-00112	cv	IL	USA	W. Chae & J. Juvik, UI	13.43¶	6x	0	1.00
'Illinois-6x.07 (Mxg2x-7)'	UI10-00113	cv	IL	USA	W. Chae & J. Juvik, UI	12.53	6x	2	0.98
'Illinois-6x.08 (Mxg2x-8)'	UI10-00114	cv	IL	USA	W. Chae & J. Juvik, UI	13.52	6x	2	0.98
'Illinois-6x.09 (Mxg2x-9)'	UI10-00115	cv	IL	USA	W. Chae & J. Juvik, UI	13.43¶	6x	2	0.98
'Illinois-6x.10 (Mxg2x-10)'	UI10-00116	cv	IL	USA	W. Chae & J. Juvik, UI		6x	0	1.00
***M.* × *giganteus* (3x), new from seed (Msa × Msi)**									
'10UI-032.001'	10UI-032.001	cv	IL	USA	J. Gifford, J. Juvik, & E. Sacks, UI	6.82§	3x	77	0.49
'10UI-032.002'	10UI-032.002	cv	IL	USA	J. Gifford, J. Juvik, & E. Sacks, UI	7.02§	3x	66	0.56
'10UI-032.003'	10UI-032.003	cv	IL	USA	J. Gifford, J. Juvik, & E. Sacks, UI	6.91§	3x	71	0.54
'10UI-032.004'	10UI-032.004	cv	IL	USA	J. Gifford, J. Juvik, & E. Sacks, UI	7.08	3x	74	0.51
'Nagara'	M116	cv	Saxony-Anhalt	Germany	M. Deuter, Tinplant	6.79	3x	88	0.46
'Ogi63'	Ogi63	w	Miyazaki	Japan	A. Nishiwaki, Miyazaki Univ.	7.00‖	3x	88	0.47
'Ogi79'	Ogi79	w	Miyazaki	Japan	A. Nishiwaki, Miyazaki Univ.	6.70‖	3x	71	0.54
'Ogi80'	Ogi80	w	Miyazaki	Japan	A. Nishiwaki, Miyazaki Univ.	6.90‖	3x	79	0.49
***M.* × *giganteus* (2x)**									
var. *purpurascens* 'Herkules'†	UI10-00018	cv	MD	USA	Kurt Bluemel Inc. nursery	4.67	2x	91	0.41
***M. sacchariflorus* (4x)**									
'Bluemel Giganteus'	UI10-00117	cv	MD	USA	Kurt Bluemel Inc. nursery	8.65§	4x	81	0.45
'EMI-5'	MATEREC11	w	Gifu	Japan	U. Jørgensen, Aarhus Univ. ← M. Deuter, Tinplant		4x††	87	0.44
'Gifu-2010-011'	JPN-2011-003	w	Gifu	Japan	T. Yamada, Hokkaido Univ.	8.18	4x	94	0.41
'Gifu-2010-027'	JPN-2011-005	w	Gifu	Japan	T. Yamada, Hokkaido Univ.	8.25	4x	99	0.40
'Gotemba Gold'	UI11-00005	cv	OH	USA	Glasshouse Works nursery	8.55	4x	83	0.45
'Hokkaido U, Livestock Farm-2'	JPN-2011-008	w	Hokkaido	Japan	T. Yamada, Hokkaido Univ.	8.51	4x	96	0.38
'Hokkaido Univ-selection-1'	JPN-2011-011	w	Miyazaki	Japan	T. Yamada, Hokkaido Univ.	8.40	4x	86	0.46
'KMI-2011-001'	JPN-2011-001	w	Miyazaki	Japan	T. Yamada, Hokkaido Univ.	8.51	4x	96	0.40
'PF30150'	UI11-00032	w	Gifu	Japan	NEF	8.22	4x	90	0.43
'PF30151'	UI11-00033	w	Gifu	Japan	NEF	8.29	4x	93	0.41
'PF30152'	UI11-00034	w	Gifu	Japan	NEF	8.34	4x	86	0.45
'PF30153'	UI11-00035	w	Gifu	Japan	NEF	8.06	4x	97	0.39
'PF30154'	UI11-00036	w	Gifu	Japan	NEF	8.38	4x	99	0.39
'PF30155'	UI11-00037	w	Gifu	Japan	NEF	8.21	4x	84	0.42
'PF30156'	UI11-00038	w	Gifu	Japan	NEF	9.72	4x	84	0.45

(continued)

Table 1 (continued)

Name	Accession identifier	Wild (w)/ cultivar (cv)	Locality of collection	Country of collection	Provider and prior sources	2C-DNA content (pg)	Ploidy	# of presence–absence differences	Jaccard's similarity coefficient
'PF30157'	UI11-00039	w	Gifu	Japan	NEF	7.99	4x	99	0.39
'Tōhoku-2010-020'	JPN-2010-008	w	Akita	Japan	T. Yamada, Hokkaido Univ.	8.53	4x	93	0.40
'Tōhoku-2010-037'	JPN-2010-015	w	Yamagata	Japan	T. Yamada, Hokkaido Univ.	8.41	4x	89	0.41
***M. sacchariflorus* (2x) and colchicine-induced tetraploid**									
'Bluemel'	UI10-00006	cv	MD	USA	Kurt Bluemel Inc. nursery	4.29	2x	99	0.35
'Earthly Pursuits'	UI10-00007	cv	MD	USA	Earthly Pursuits nursery	4.29	2x	99	0.35
'Golf Course 2x'	UI11-00006	cv	MI	USA	J. Juvik, UI ← Michigan State Univ.	4.40§	2x	99	0.35
'Golf Course 4x'	UI11-00007	cv	IL	USA	J. Juvik, UI ← colchicine treatment of 2x callus culture	8.93	4x	99	0.35
'Hortico'	UI10-00008	cv	ON	Canada	Hortico Nurseries Inc.	4.47§	2x	99	0.35
'PMS-071'	PMS-071	w	Beijing	China	J. Peng, Wuhan Botanical Garden	4.36	2x	89	0.38
'PMS-075'	PMS-075	w	Beijing	China	J. Peng, Wuhan Botanical Garden	4.39	2x	83	0.41
'Robustus-Bluemel'	UI10-00009	cv	MD	USA	Kurt Bluemel Inc. nursery	4.39	2x	99	0.35
'Robustus-Earthly Pursuits'	UI10-00010	cv	MD	USA	Earthly Pursuits nursery	4.43§	2x	99	0.35
var. *lutarioriparius* 'PF30022'	UI11-00031	w		China	NEF	4.26§	2x	105	0.31
***M. sinensis* (2x)**									
var. *condensatus* 'Emerald Shadow'	UI10-00038	cv	MD	USA	Kurt Bluemel Inc. nursery	5.45	2x	96	0.39
var. *condensatus* 'Cabaret'	UI10-00012	cv	FL	USA	Emerald Coast Growers nursery	5.33	2x	96	0.39
var. *condensatus* 'Cosmo Revert'	UI10-00014	cv	CA	USA	Greenlee Nursery	5.26	2x	107	0.34
var. *condensatus* 'Cosmopolitan'	UI10-00015	cv	FL	USA	Emerald Coast Growers nursery	5.10	2x	107	0.34
var. *purpurascens*	UI10-00019	cv	MD	USA	Kurt Bluemel Inc. nursery	5.47	2x	107	0.31
'Strictus'	UI10-00092	cv	MD	USA	Kurt Bluemel Inc. nursery	5.62	2x	107	0.31
'Tripple Brook Farm'‡	UI10-00011	cv	MA	USA	Tripple Brook Farm Inc. nursery	5.60	2x	99	0.32
Control sugarcane lines									
SO8272	PI495670	cv		Malaysia		10.49	10x	242	0.10
IJ76-315	PI423343	cv		Indonesia		10.76	10x	251	0.10

*Originally named *M. floridulus* 'Flower Factory'.
†Originally named *M. sinensis* var. *purpurascens* 'Herkules'.
‡Originally named *M. sacchariflorus* 'Triple Brook Farm'.
§Chae *et al.* (2014).
¶Chae *et al.* (2013).
‖Nishiwaki *et al.* (2011).
**Głowacka *et al.* (2010).
††Clifton-Brown *et al.* (2001).
IPG, Institute of Plant Genetics, PAS; MSU, Mississippi State Univ.; NEF, New Energy Farms; UI, Univ. of Illinois; IHAR, Plant Breeding and Acclimatization Institute.

Accessions of Mxg's parental species, Msa and Msi, were included in the study to provide information on the potential diversity that could be expected from additional interspecific crosses. The Msa accessions included 18 tetraploids of Japanese origin, seven diploids obtained from nurseries in North America, and three diploids from China. The seven Msi accessions were ornamental cultivars available commercially in the United States. Included in the study were Msa 'Bluemel Giganteus' (4x) and Msi var. *condensatus* 'Cabaret' (2x), the parents of four new triploid Mxg genotypes produced at the University of Illinois.

DNA Extraction and SSR marker genotyping

Young leaves were lyophilized, then ground in a ball mill (Geno/Grinder 2000, SPEX SamplePrep, Metuchen, NJ, USA). Lyophilized samples were stored at −20 °C before and after milling. DNA was extracted from lyophilized, powdered leaf tissue by using a CTAB method modified from Kabelka *et al.* (2002). The accessions were initially screened with 85 SSRs (Table 2), including 79 nuclear (Hung *et al.*, 2009; James *et al.*, 2012; Swaminathan *et al.*, 2012) and six chloroplast markers (de Cesare *et al.*, 2010). A final set of 64 nuclear and five chloroplast SSRs were used for the analyses. Of the nuclear SSRs, 38 were previously mapped on Msi (one per chromosome arm for each of the 19 chromosomes in Msi; Swaminathan *et al.*, 2012; Table 2). PCR conditions were the same as described in Swaminathan *et al.* (2012), except that 0.25 μL of each 10 μM primer stock solution was used instead of 0.1 μL. Additionally, for chloroplast SSRs Sac-2, Sac-3, Sac-10, and Sac-13, the annealing temperatures 58 °C, 56 °C, 52 °C, and 62 °C were used respectively. Size separation of the PCR products was done on a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) with GeneScan 500 LIZ size standard at the Univ. of Illinois' Keck Center for Functional Genomics. Marker scoring was done using the STRand software v. 2.4.59 (http://www.vgl.ucdavis.edu/STRand; Toonen & Hughes, 2001). Amplicons between 75 and 350 bp were analyzed.

If entries differed from each other by less than five percent in their allelic profile but were not identical (i.e. were nearly identical), we suspected genotyping errors due to allele dropout (ADO) or false allele (FA). If a genotyping error was suspected, we implemented the following procedures sequentially until an error was identified or the difference was found to be repeatable: (i) technical replication via additional PCR followed by consensus calling, (ii) changing the fluorescent dye used, (iii) switching from our standard DNA polymerase (GoTaq Colorless Master Mix; Promega, Madison, WI, USA) to a high-fidelity enzyme (Phusion High-Fidelity DNA Polymerase; New England BioLabs, Ipswitch, MA, USA), and (iv) re-extraction of DNA followed by additional PCR. Genetic distance estimates in Table 1 were based on genotype calls that were validated with the multistage protocol described above. To further quantify sources of variation for genotyping errors, we also conducted a factorial experiment comparing two entries (Mxg 'Illinois' and Mxg 'Freedom'), 64 nuclear SSRs, the effects of new DNA extractions from different leaf samples within entry, and PCR reactions-electrophoretic separations within DNA extractions.

RAD-Seq genotyping

Twelve legacy Mxg cultivars, all eight colchicine-induced Mxg polyploids, two diploid Msa, one tetraploid Msa, and one Msi were included in three RAD-seq libraries. The type specimen Mxg '1993–1780', Msa (2x) 'Bluemel', and three of the Mxg hexaploids were replicated in two libraries. Additional individuals that were part of a separate study were also included in the libraries (data not shown). Sequencing-library preparation (http://openwetware.org/wiki/Sacks:RAD-seq) was based on the protocol of Poland *et al.* (2012), using 96 barcoded adapters from Thurber *et al.* (2013). Quantitative PCR and sequencing on an Illumina HiSeq 2000 with 100 bp single-end reads were performed at the University of Illinois Roy J. Carver Biotechnology Center DNA Sequencing Unit.

Flow cytometry

Nuclear DNA content and associated estimates of ploidy were obtained via flow cytometry using a protocol modified from Rayburn *et al.* (2009) with *Sorghum bicolor* 'Pioneer 8695' as the internal standard. DNA content of each entry was analyzed except those for which estimates were previously published or for which fresh leaf samples were unavailable for our study. Nuclei were analyzed using a flow cytometer Model LSRII (BD Biosciences, San Jose, CA, USA; Flow Cytometry Facility at the University of Illinois-Keck Biotechnology Center). To estimate pg of DNA per 2C nucleus, mean fluorescence of the analyzed sample G1 peak was divided by the fluorescence reading of the G0/G1 peak of sorghum, multiplied by 1.74 pg/2C (McMurphy & Rayburn, 1991).

Data analysis

Simple sequence repeat allelic data were recorded as binary scores. Genetic similarity among accessions was estimated using Jaccard's (1908) similarity coefficients and used for hierarchical clustering unweighted pair group method with arithmetic mean (UPGMA), revealed by NTSYS-pc v. 2.21m (Rohlf, 2002). Discriminant analysis of principal components (DAPC) was performed with the R package adegenet v. 2.15.1 (Jombart *et al.*, 2010) to visualize groups of genetically related individuals. For the factorial experiment, an ANOVA was performed with SAS procedure GLIMMIX (SAS 9.3, SAS Institute, Cary, NC, USA) using a binomial distribution to test sources of variation for genotyping errors.

For sequence analysis, the software Stacks version 0.9996 (Catchen *et al.*, 2011) and Bowtie2 version 2.0.2 (Langmead & Salzberg, 2012) were used. Sequences were split by barcode and trimmed to 80 nucleotides using Stacks, then aligned using Bowtie2 to the Mxg exome sequence published by Chouvarine *et al.* (2012). Similar results were obtained when aligning to the *S. bicolor* genome (data not shown). The *pstacks*, *cstacks*, and *sstacks* modules of Stacks were then used to detect variants, build a catalog of variants, and call genotypes, respectively, in the 29 samples. For genotype comparisons, we analyzed 2617 RAD tags (196 275 nucleotides) that aligned to the exome sequence and were present in all 29 samples. A custom R script

Table 2 85 simple sequence repeat (SSR) markers used to compare genetic diversity of *Miscanthus* entries in this study. Linkage group (if previously reported), references, repeat motif, dye used, and primer sequence are shown. Mean error per allele (MEPA) and mean error per locus (MEPL) were estimated via within accession replication for two Mxg accessions, Mxg '1993–1780' and Mxg 'Freedom' UI10-00119, in this study. Italicized SSRs were dropped from the study due to lack of informativeness or reproducibility

SSR name	LG*	Repeat motif	Dye†	Forward primer (5′–3′)	Reverse primer (5′–3′)	MEPA	MEPL
Chromosomal SSRs mapped on *M. sinensis* chromosomes (Swaminathan *et al.*, 2012)							
ESSR_009	1	GCG	VIC	GGGACCAAGATGACCACACC	CCTTGAGCGTGGCTTTCATC	0	0
GSSR_035	1	CGC	6-FAM	TAGTACTCACAGTCTCGCTCCTGG	GAGTGGTGGCCTTGAGCTTC	0	0
GSSR_037	2	GAA	PET	AACAGTGAAGCCTCCCTATCTTGA	TCATGGATTATGCATCACCTTCTT	0.094	0.125
GSSR_069‡	*2*	*GTG*	*NED*	*GAGGATCAGGAACTGGGAGTGAC*	*CTCCGATCTAATCCATTCCATCAC*		
ESSR_012	3	CGC	6-FAM	AACGAGGACAACGCCAACTC	TCGATCACGAGAGGAGAGCC	0	0
ESSR_017	3	CGGC	VIC	ATCTTCTGGGTGGGTCCGTT	CCGTTCACCTTAAACTGCGG	0.031	0.125
GSSR_015§	*4*	*CACG*	*NED*	*TACTAGCTAAACGGACGTGCAGTG*	*GAAGATTCTCTGTGCTGACTGGGT*		
GSSR_020	4	AT	PET	TTTCTGCACAAGAACACCAACATT	CAATGCAGTTTGAAAAGAGGAAGG	0	0
GSSR_046	5	CCAT	6-FAM	CAAGGGATACAGACCTCACCATTC	GGCTTCGGAGATCCCAATTAGA	0	0
GSSR_051	5	GGC	VIC	TTTCTTGTCCAAGAGATCAGGAGG	ACCTGCACACCGAGAACCAC	0.013	0.063
GSSR_070	6	TCG	PET	GATGGACAAGATCACCGAGAAGTT	GAGGACCCCTTCTTCTTGTCGT	0	0
GSSR_054	6	CTGTGC	NED	GGAGAGAAGACGAAGGAAAGGC	CAAGCTTTCACTGCATTTCAATTTC	0	0
ESSR_026	7	GT	VIC	CCCTCCTCTCGTCCTCTTCC	CCGGCGTCGATCTCTCTAAC	0	0
GSSR_042	7	AGGGG	6-FAM	GGGGTTTAGGGCTTTGGAAGA	CCGCTCCTCCTGTTACTTTTCTTT	0	0
ESSR_001	8	CATA	PET	CATCAAACAATCCGTACGTGGA	TGGCAGTAGCTCAGTTCAGGC	0	0
GSSR_048	8	GGGGAT	NED	GTGATGTGGATCTGAAAAGCAAGA	CAACCCTAACTTGTCCAAACCATC	0	0
GSSR_045	9	TTTGT	6-FAM	ATGTTTACTTCACCCTTTTTGGGC	CACGAAGTTTTTCGCGTTTCTTAC	0	0
GSSR_061	9	GAC	VIC	TATAGTATGGCCTCGTGCAAGTGA	TGTTACGATTTGCTAATCTCCCGT	0	0
GSSR_012	10	TGC	PET	TGCAAAAACACTCCAATGACTTGT	GTTCATGACACTTTGACTCGATGG	0	0
GSSR_044	10	TCT	NED	ATCCGAACAGTGTGGTCGTAACAT	CATGAGAGTAGAGCCGGTAGCAAT	0	0
GSSR_019§	*11*	*AG*	*6-FAM*	*TCAAGTTGGAGTTTCATCAGCATC*	*AGGTACGCAGGGACCACATATAAA*		
GSSR_021	11	TAC	VIC	GAAGCAACCAAGCACCAAAATC	CGCGAACGTACCTAAAAGTTTGAC	0.125	0.125
ESSR_008	12	CAAGCA	PET	CAGCCATTCAAGGACAAGCC	CACGCCAATCACGGTTATCA	0	0
GSSR_038	12	TTC	NED	AATTTTCTTTGGACCGAACCTCAT	TGGCAGCTACATCAGAAATCAAAA	0	0
ESSR_010	13	ATTCGA	6-FAM	CATGCAGAGGCTGGAGAAGG	CGGCAACCAATCCCTCGTAT	0	0
GSSR_050	13	CCT	VIC	GCACTGCACCTTAATTAATCCACC	CCTGCTCATGATGAATCAAGAATG	0	0
GSSR_075	14	GCACCG	NED	TGGACATTTACATCCACTTTGCAG	AGAGAAACAGGAGAGGCACTAGCA	0	0
GSSR_023	14/15	AGC	PET	ACAACCACAGGATTACACTGGACA	CAGCTTCAGCAGCAGTAGCAGTAG	0	0
ESSR_014	15	AAAG	6-FAM	GAAGCAACGCCTCACCAAAC	GGTGTGCTCCGGCTAGTGTT	0	0
GSSR_077	15	ATAGG	VIC	TGTAGAGGGTTGTGGATTGAGAAA	TCGTCAATTAATTATAGATGTTTGGTCA	0.031	0.063
GSSR_053‡	*16*	*CCA*	*PET*	*AGGAATGAGCTCGGTGATGCT*	*CTCACGATCGACGACTACGCT*		
GSSR_071	16	GTG	NED	TGTAACGGCCAACCTTCTGAGT	ATACAAAGCAGGTGATGAATGTCG	0	0
GSSR_005	17/16	CCG	6-FAM	ACCCAATAAGCACGGAACCCTA	AAGCGAGCGTACCTGTAGAGGAG	0	0
GSSR_064	17/16	GCT	VIC	GACAACCTCGACTCGTACAACGTA	AGAGTAGAGAACACGATTCACGGC	0	0
ESSR_028	18	GGA	PET	CCAAGCAGACCCCCAGTAGA	ATCGATCTCGCACAGCAGGT	0	0
GSSR_022	18	CAGCA	NED	ATTATTAAAATCCGGAGACCAGCC	ATTGAGGAGGTAGAAGAAGCCAGC	0	0
ESSR_024	19	CCTTT	6-FAM	CCCAACTCCCCTACTTGCCT	GCGGTCTCGGTCAAAGACAC	0	0
GSSR_039	19	CGG	VIC	GCAGTCTAGGATTGATCCGATGG	GGTGTCCACCATCATCGTCTG	0	0

(continued)

Table 2 (continued)

SSR name	LG*	Repeat motif	Dye†	Forward primer (5′–3′)	Reverse primer (5′–3′)	MEPA	MEPL
Sugarcane SSRs (James *et al.*, 2012)							
Mis.fluor.6§		*TTATA*	*VIC*	*TACGCGATTGTGTAAAGTACACCG*	*TAGCTAGCTCTCTCCTCATCGCTC*		
Mis.fluor.7		ATATA	PET	AGCATCAAGCACAATCCTCATTCT	TCAGGTACACTACTTGGTCTGTTTGTG	0.188	0.188
Mis.fluor.8		AATAT	NED	GACGAATTGACCGCCTACCTTTAT	ACACTCACTCAAGTGCCTTGCTTC	0	0
Mis.fluor.9		TTCTA	6-FAM	CGAATTGGTCAAGACTCTCCTGTT	AGCCTGAAGCAAATTCAATGAAAC	0	0
Mis.fluor.10		ATATT	VIC	CACGTACATTGCTAGCTGGAACC	ATTTGATCGTACTCGGAAGCGTAG	0	0
Mis.fluor.11		ATT	PET	ATCCGCTTCGACCTCTACATCAC	TGACAGGAGATGAAATGCATCGTA	0	0
Mis.fluor.12		AAT	NED	GGAATAATGTTTAGGTGGGGAAGC	TGCATAGTATCTGTCCACGTGAGATT	0	0
Mis.fluor.13¶		AAAT	6-FAM	CGCGCAACTTTCTTTCTTTTT	GGTCCATAGCCACTGCAATAAA		
Mis.fluor.14‡		*ATT*	*VIC*	*AAAGTTGGGTGAACAAATAAAATAAAAA*	*ATCGGGGAGCCTAAGTCATTT*		
Mis.fluor.15‡		*AAC*	*PET*	*CAACCTGAGCATACAGGCTAGACA*	*TGAACTGTGCATCAATGGTAATGA*		
Mis.fluor.16		AGATC	NED	CAAGCAAAGTTGCGTTTACTCTCT	CCATGCTATCATTAGTGCAGCTC	0.125	0.125
Mis.fluor.17‡		*GGACG*	*6-FAM*	*TAGCTTACTTTGACGGTGCTCGAT*	*ATGTATGATGGTGCCTGGTGC*		
Mis.fluor.18		AGAGA	VIC	ATGTCCAGCAAGGAGGGAAAG	TCTCGATCAGGAGAAGAAGCCTTA	0	0
Mis.fluor.19		AATCA	PET	TAGAAGCACGTGTTCGCGATG	GCTGCTAGCGATCGAGCTGAC	0	0
Mis.fluor.21		GCC	FAM	CTTGCCTGCTTCGGCATCTT	GACGGTCTCACTCTCACCATCATC	0.063	0.063
Mis.fluor.22		GCTGTT	VIC	GTTGTCCTTCCCTTTGGGTTG	GATCTATCAGAGAAGTCCCAGCGA	0	0
Mis.fluor.23		CCTGGA	PET	CAAACACCGTCGTGTTACTCCTC	TCAGGACTTTCTTCGTCAGGATTC	0	0
Mis.fluor.24§		*TGTC*	*NED*	*GTTTTCTTCACCCACAGCATTGAC*	*CAACCAACTATCGCGGTTGC*		
Mis.fluor.25		AATA	6-FAM	AAGATCTCACATGGTTATGTTTTGACA	GGCGAGACAGAGTCATTTTTCTTT	0.063	0.063
Mis.fluor.26		CCGTC	VIC	AAACAGAAGTCTACGTGGAGGTGG	GAAATAATGGAGGGAGGGAGGAT	0	0
Mis.fluor.27		GAGAG	PET	GGGGATATATAAATGAGGATGGCG	CCTGTCCTGACTCCTCTTCTGTTC	0	0
Mis.fluor.28		CTGAT	NED	TATATGGCACGGTGCAGTAACATT	TATATAGGATATCCGGCCGTGTGA	0	0
Mis.fluor.29		GATCG	6-FAM	ACGCACTTCAGACCTCAGTCAGTT	GAGAGACCACCCGATCCCAG	0.094	0.188
Mis.fluor.30		GGGAA	VIC	GTTATAGGCCGGGATAAACAATGG	ATGTTTACACATATGCCCTCGCTC	0	0
Mis.fluor.31		TCATG	PET	GCCTAGTCTAGCCGGACAGTATGA	GTTGTGTTGAATTAAGTTGCACGG	0	0
Mis.fluor.32‖		*TGCGT*	*NED*	*AGCAGTGCAGGTTGTAGCAGC*	*TCCATCCATCTTCTTCTCCGATTA*		
Mis.fluor.33‡		*GAGAG*	*6-FAM*	*GCGAGATCTCAGTTCGTTGGTAAT*	*TTTCTCAGTTCTTATGTTTAAGCCAGC*		
Mis.fluor.35‡		*TGCTCC*	*PET*	*CAACAAGTTGCTGTGTTGACGTT*	*ATATCACATCGGACTATCGGAGGA*		
Mis.fluor.36‡		*ACGAAG*	*NED*	*GAAGGAGAGATCATGTCTTGGCTC*	*CTGTTCTACTTGCAGCTTCCTTCC*		
Mis.fluor.37		GTCT	6-FAM	TCAAAACTGAATGCAGGCAATAGA	TACATAGCCTGAAAGCAACGGGTA	0	0
Mis.fluor.38‡		*ATTA*	*VIC*	*GACCGAAAAGAAAACCACCAAAAT*	*TTCAGAAAATTAAGGCCACGTGAT*		
Mis.fluor.39		CTTCGT	PET	TCAGCTCTTTCCAGCATTTGTACC	CAAGTCGTTGCTGGCGAAAG	0	0
Mis.fluor.40		TGATGG	NED	TGTTGTATGGAGTGAGGTGAGGAA	TGTCCAAATTTTAAGCAAAGCACA	0	0
Mis.fluor.41		TGT	6-FAM	GATGCGCAGTTGTTCTCTCATTAT	CAAATATCTCCAGGAACAGCATGA	0.250	0.250
Mis.fluor.42		TG	VIC	ATTTTGAAATAAGAAAGACGGCCA	ACACACAAACACACTCATCATTCG	0	0
Mis.fluor.44		GA	NED	AACGTGTCGTAAGGTTTGGTTGTT	TGCTGCAGGTCCGTACTATACATC	0	0
Mis.fluor.45		CGTGAT	6-FAM	AGAGGAGATGTTGGAAGGTACACG	TAGTTGGCTCTTGCACACCTGTAA	0	0
Mis.fluor.46		CCACGG	VIC	GACCAGGGCAATAAGCACAAC	ATATCGAGATGCCTACGAGAAAGG	0	0
Mis.fluor.47		TATC	PET	ATGGATTTGGCTAGTTTGCATTGT	GGTCTGAGGTTGGGTAGGAGTTTT	0	0
Mis.fluor.48‡		*CT*	*NED*	*TGGAATTAGTCTTTTCAACCAACCA*	*GGTAAAGACCCAAATTACTGGTGATG*		

(continued)

Table 2 (continued)

SSR name	LG*	Repeat motif	Dye†	Forward primer (5′–3′)	Reverse primer (5′–3′)	MEPA	MEPL
***Miscanthus* SSRs designed by PCR isolation of microsatellite arrays (PIMA) (Hung *et al.*, 2009)**							
M&M21		CT	6-FAM	GCATATATTGACCTATGTGTG	TAGCTTCGTTCCATCTCCAT	0	0
Chloroplast SSRs mined from sugarcane complete chloroplast genome sequences (de Cesare *et al.*, 2010)							
Sac-2§		$(T)_{11}/(A)_{21}$	*6-FAM*	*ACCTTTCCCTGCATTAGGCA*	*AGACCACGACTGATCCTCAA*		
Sac-3		T	VIC	ACCAAAGAAAAGAAGGACGC	TCTACGGATAGGGACTCTAT		
Sac-10		T	PET	TAAAACGGGCATTCCTACGC	CCGAAGCATAAACAAAGACAG		
Sac-13		T	NED	GTGCTTCGAGGGCGCAAATT	CAGAGCGGAGTAGAGCAGTT		
Sac-17		A	PET	TACACCGGACGCTCCTGTCAAA	TTGCCCCTCTTGCATGTACT		
Sac-26		T	NED	GAGTGTGTGCGAGTTGTCTA	ACCAATGAATCGCGAAATGC		
Average						0.017	0.022

*Linkage group in *M. sinensis* map (Swaminathan *et al.*, 2012).
†Applied Biosystems (ABI) fluorescent dyes: 6-FAM®, VIC™, PET™ and NED™.
‡No product.
§Product not reproducible.
¶No product produced in the two Mxg entries studied to estimate genotyping error rates.
‖Monomorphic product in all *Miscanthus* accessions evaluated.

was written to calculate nucleotide dissimilarity (SNPs per base pair) between samples. Only the last 75 nucleotides of each 80 nucleotide locus were used, given that the first five nucleotides were part of the restriction site and therefore invariable. If a pair of samples each had more than one haplotype at a locus, the haplotypes were matched to each other to find the minimum nucleotide dissimilarity. A UPGMA tree was calculated using the R package phangorn (Schliep, 2011).

Results

Genetic diversity in Mxg

Genetic similarity to the type-specimen, '1993–1780', of the triploid Mxg legacy cultivars and the Mxg 'Illinois' colchicine-induced polyploids, based on nuclear SSRs, ranged from 0.98 to 1 (Table 1, Fig. 1). Mxg 'Illinois', 'Freedom', and 'EMI01' were identical to the type-specimen (Table 1). 'Illinois' and 'Freedom' are the predominant Mxg commercial cultivars in the U.S. and 'EMI01' has been studied extensively in Europe. Only five of 32 Mxg legacy cultivars and three of the eight colchicine-induced Mxg polyploids differed from the type-specimen for at least one nuclear SSR allele (1–3 presence-absence differences; Table 1). Among the eight Mxg legacy cultivars and colchicine-induced Mxg polyploids that differed from the type-specimen, pairwise nuclear genetic similarities ranged from 0.96 to 1, with the three colchicine-induced polyploids identical to each other (Fig. 1, Table 3). In the UPGMA and DAPC cluster analyses, all of the Mxg legacy cultivars, including the type-specimen, and the colchicine-induced Mxg polyploids formed a single cluster (Figs 1 and 2). Consistent with the nuclear DNA markers, all of the 41 Mxg legacy cultivars and colchicine-induced polyploids, had identical chloroplast SSR profiles (Fig. 3). Genome sizes for the Mxg legacy cultivars ranged from 6.68 ± 0.11 to 7.07 ± 0.07 which was similar to the ~7.0 pg observed in previously studies of 'Illinois' (Table 1; Rayburn *et al.*, 2009; Chae *et al.*, 2013).

In contrast with the Mxg legacy cultivars, each of the eight new from seed triploid Mxg progeny differed from the type-specimen and from each other. Genetic similarity to the type-specimen of the eight new triploid Mxg progeny ranged from 0.46 to 0.56 based on nuclear SSRs (Table 1, Fig. 1). Among the new Mxg progeny themselves, pairwise nuclear genetic similarities ranged from 0.41 to 0.75; '10UI-032.001' and 'Ogi80' were the least similar to each other, whereas the full sibs '10UI-032.002 and '10UI-032.003' were the most similar to each other (Fig. 1, Table 3). Nuclear genetic similarities among the four full sibs in the 10UI-032 family ranged from 0.64 to 0.75. In the UPGMA and DAPC cluster analyses, two clusters were formed by the new from

seed triploid Mxg entries, with one cluster consisting of the four 10UI-032 full sibs, and the other consisting of 'Ogi63', 'Ogi79', 'Ogi80', and 'Nagara' (Figs 1 and 2). Two chloroplast SSR profiles were observed for the new Mxg entries, with one consisting of the four 10UI-032 full sibs and its maternal parent Msa 'Bluemel Giganteus' (4x), and the other consisting of 'Ogi63', 'Ogi79', 'Ogi80', and 'Nagara' M116 (Fig. 3), which was consistent with the UPGMA and DAPC nuclear SSR analyses (Fig. 2). The chloroplast SSR profiles for the new Mxg entries differed from the Mxg type-specimen's profile. Genome sizes for the eight new Mxg progeny ranged from 6.70 to 7.08 pg, indicating that they were likely triploids, as expected (Table 1). The diploid Mxg var. *purpurascens* 'Herkules' was originally and erroneously named Msi but it clustered with triploid Mxg entries in the UPGMA and the DAPC analyses (Figs 1 and 2), and it had a genome size of 4.67 pg (Table 1), which was consistent with it being a hybrid between diploid Msa and Msi parents.

Genetic diversity in Mxg's parental species, Msa and Msi

Our modest samples of the Mxg parental species, Msa and Msi, were more genetically differentiated from the Mxg type-specimen than the new from seed Mxg genotypes. Genetic similarity to the Mxg type-specimen, '1993–1780', of the tetraploid Msa genotypes, diploid Msa genotypes, and Msi gentoypes ranged from 0.38 to 0.46, 0.31 to 0.41, and 0.31 to 0.39, respectively (Table 1). As expected, genetic similarity of the triploid Mxg to Msa was typically greater than to Msi, because triploid Mxg has two copies of the genome from Msa and only one copy from Msi. Among the tetraploid Msa genotypes, pairwise genetic similarities ranged from 0.44 to 0.96 (Table S1). All seven of the diploid Msa accessions obtained from US nurseries were genetically identical to each other or nearly so (0.98–1.0 similarity, Fig. 1). In contrast, the three Msa diploids from China were dissimilar to each other and to the accessions sold by US nurseries (Table S1, Fig. 1). Pairwise genetic similarities among the nonidentical diploid Msa (including those from China and the US nursery genotype) ranged from 0.51 to 0.64, with Msa var. *lutarioriparius* being the most differentiated (Table S1, Fig. 1). In the DAPC analysis, three clusters were formed by the Msa accessions, with one group consisting of only diploids, another group consisting of only tetraploids, and the third group including both diploids and tetraploids (Fig. 2).

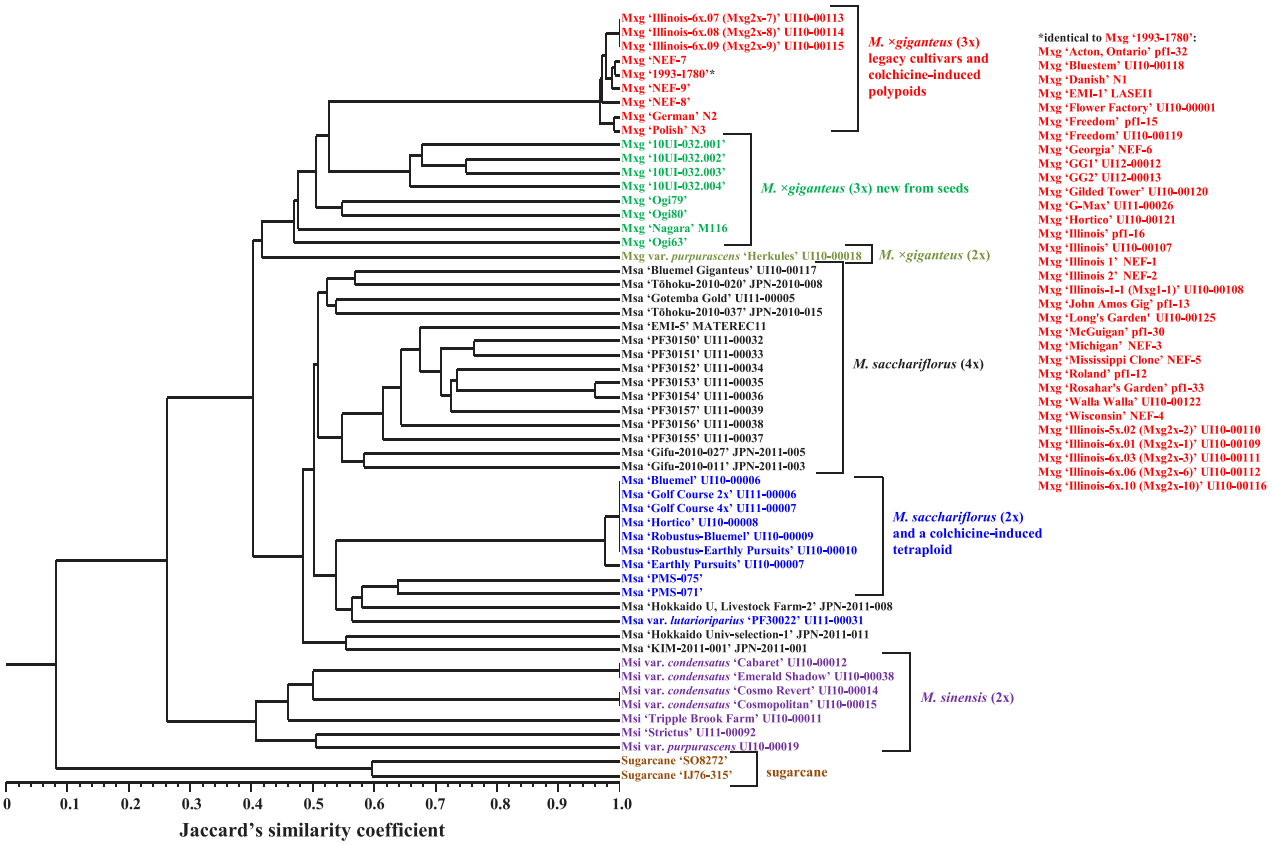


Fig. 1 UPGMA cluster analysis of 85 *Miscanthus* accessions and two sugarcane lines based on Jaccard's similarity coefficients calculated from binary data of 527 alleles from 64 nuclear SSR loci revealed by software NTSYS-pc v. 2.21m.

Table 3 Matrix of Jaccard's similarity coefficients for all *Miscanthus* × *giganteus* accessions in the study that differed from the type-specimen '1993–1780'

No	Accession name, identifier and ploidy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Mxg 'German' N2 (3x)	1.00																
2	Mxg 'NEF-7' NEF-7 (3x)	0.98	1.00															
3	Mxg 'NEF-8' NEF-8 (3x)	0.97	0.98	1.00														
4	Mxg 'NEF-9' NEF-9 (3x)	0.98	0.98	0.98	1.00													
5	Mxg 'Polish' N3 (3x)	0.99	0.97	0.96	0.97	1.00												
6	Mxg 'Illinois-6x.07 (Mxg2x-7)' UI10-00113 (6x)	0.97	0.98	0.97	0.98	0.96	1.00											
7	Mxg 'Illinois-6x.08 (Mxg2x-8)' UI10-00114 (6x)	0.97	0.98	0.97	0.98	0.96	1.00	1.00										
8	Mxg 'Illinois-6x.09 (Mxg2x-9)' UI10-00115 (6x)	0.97	0.98	0.97	0.98	0.96	1.00	1.00	1.00									
9	Mxg '10UI-032.001' 10UI-032.001 (3x)	0.49	0.48	0.47	0.48	0.49	0.49	0.49	0.49	1.00								
10	Mxg '10UI-032.002' 10UI-032.002 (3x)	0.56	0.56	0.55	0.56	0.56	0.57	0.57	0.57	0.68	1.00							
11	Mxg '10UI-032.003' 10UI-032.003 (3x)	0.54	0.54	0.53	0.54	0.54	0.55	0.55	0.55	0.68	0.75	1.00						
12	Mxg '10UI-032.004' 10UI-032.004 (3x)	0.51	0.51	0.51	0.51	0.51	0.52	0.52	0.52	0.64	0.65	0.68	1.00					
13	Mxg 'Nagara' M116 (3x)	0.46	0.46	0.46	0.46	0.46	0.47	0.47	0.47	0.48	0.53	0.54	0.47	1.00				
14	Mxg 'Ogi63' Ogi63 (3x)	0.47	0.48	0.48	0.48	0.46	0.48	0.48	0.48	0.44	0.47	0.48	0.45	0.45	1.00			
15	Mxg 'Ogi79' Ogi79 (3x)	0.54	0.54	0.54	0.55	0.54	0.55	0.55	0.55	0.46	0.51	0.52	0.51	0.50	0.51	1.00		
16	Mxg 'Ogi80' Ogi80 (3x)	0.49	0.49	0.49	0.48	0.48	0.49	0.49	0.49	0.41	0.47	0.45	0.47	0.44	0.46	0.55	1.00	
17	Mxg var. *purpurascens* 'Herkules' UI10-00018 (2x)	0.41	0.42	0.42	0.41	0.41	0.42	0.42	0.42	0.37	0.44	0.45	0.43	0.43	0.38	0.44	0.42	1.00

Genome size for the tetraploid Msa accessions ranged from 7.99 to 9.72 pg, and for the diploid Msa accessions ranged from 4.29 to 4.47 pg (Table 1).

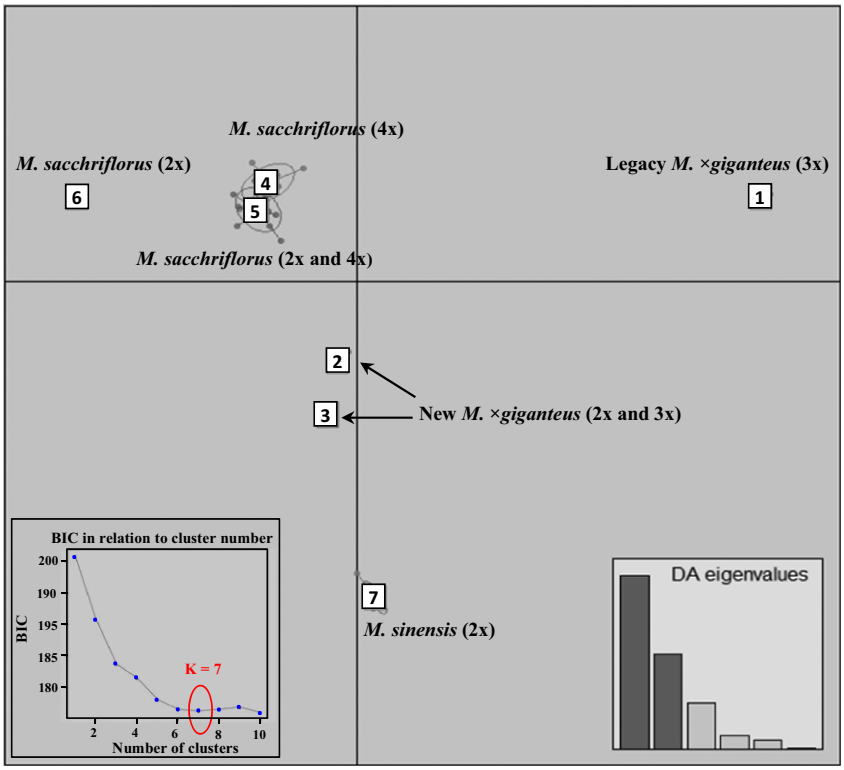
Pairwise genetic similarities among the Msi accessions, excluding duplicate genotypes, ranged from 0.37 to 0.50. Simple sequence repeat marker data, indicated that 'Emerald Shadow' is a green revertant of the variegated 'Cabaret', and confirmed that 'Cosmo Revert' is an all-green form of the variegated 'Cosmopolitan'. In the UPGMA and DAPC cluster analyses, all of the Msi accessions formed a single cluster (Figs 1 and 2). Genome size for the Msi accessions ranged from 5.10 to 5.62 pg (Table 1).

SSR performance

Although all of the 85 SSRs used in the current study were previously reported to amplify *Miscanthus* DNA (Hung *et al.*, 2009; de Cesare *et al.*, 2010; James *et al.*, 2012; Swaminathan *et al.*, 2012), we observed that 11 produced no product, four gave only unreproducible products, and one was monomorphic (Table 2). From the 64 informative nuclear SSR loci, 359 alleles were observed in *Miscanthus*, and 527 alleles were observed when both *Miscanthus* and sugarcane were considered. Nine Msa-specific alleles and five Msi-specific alleles were identified (Table 4). Interestingly, the Msa-specific alleles were observed in both diploid as well as tetraploid accessions. Mis.Fluor.32 produced a 138 bp product in all 85 *Miscanthus* accessions but was absent from the two sugarcane lines, which makes this amplicon potentially useful for distinguishing *Miscanthus* from *Saccharum*, and identifying their intergeneric hybrids.

SSR repeatability

In the factorial experiment on genotyping error, ADOs were observed but FAs were not. ANOVA results indicated highly significant differences among SSRs for genotyping error (Table 5). Allele within SSR, Accession, SSR × Accession, Allele within SSR × Accession, and DNA extraction within Accession were all nonsignificant (Table 5). The variance component estimate for DNA extraction within Accession was an order of magnitude smaller than the variance associated with PCR reactions within DNA extraction × Accession (0.001 in contrast to 0.015). The two accessions in this experiment, Mxg 'Illinois' and Mxg 'Freedom', were subsequently found to have identical multilocus genotypes, which limited inferences about the effect of accession on genotyping error rate but lent confidence to assessment of the other sources of variation studied. Of the 64 informative nuclear SSR loci evaluated in the factorial experiment, 11 had error rates greater than zero (Table 2).



1) All 33 legacy cultivars of *M.* ×*giganteus* (3x) and 8 colchicine-induced *M.* ×*giganteus* polypoids
2) New *M.* ×*giganteus* (3x): '10UI-032.001', '10UI-032.002', '10UI-032.003', '10UI-032.004'
3) New *M.* ×*giganteus* (3x): 'Ogi63', 'Ogi79', 'Ogi80', 'Nagara' M116;
M. ×*giganteus* (2x): var. *purpurascens* 'Herkules' UI10-00018
4) *M. sacchriflorus* (4x): 'EMI-5' MATEREC11, 'PF30150' UI11-00032, 'PF30151' UI11-00033, 'PF30152' UI11-00034, 'PF30153' UI11-00035, 'PF30154' UI11-00036, 'PF30155' UI11-00037, 'PF30156' UI11-00038, 'PF30157' UI11-00039
5) *M. sacchriflorus* (2x): 'PMS-075', 'PMS-071', var. *lutarioriparius* 'PF30022' UI11-00031;
M. sacchriflorus (4x): 'Bluemel Giganteus' UI10-00117, 'Gifu-2010-027' JPN-2011-005, 'Gifu-2010-011' JPN-2011-003, 'Gotemba Gold' UI11-00005, 'Hokkaido U, Livestock Farm-2' JPN-2011-008, 'Hokkaido Univ-selection-1' JPN-2011-011, 'KIM-2011-001' JPN-2011-001, 'Tōhoku-2010-020' JPN-2010-008, 'Tōhoku-2010-037' JPN-2010-015
6) *M. sacchriflorus* (2x) and colchicine-induced tetraploid: 'Bluemel' UI10-00006, 'Golf Course 2x' UI11-00006, 'Golf Course 4x' UI11-00007, 'Hortico' UI10-00008, 'Robustus-Bluemel' UI10-00009, 'Robustus-Earthly Pursuits' UI10-00010, 'Earthly Pursuits' UI10-00007
7) All *M. sinensis* (2x) accessions

BIC –Bayesian Information Criterion

Fig. 2 Discriminant analysis of principal components (DAPC) of 85 *Miscanthus* accessions based on binary data of 359 alleles from 64 nuclear SSR loci revealed by R package adegenet v. 2.15.1. K = 7 clusters selected based on the Bayesian Information Criterion, and DAPC eigenvalues are depicted in the enclosed barplot.

Among the SSR loci, the mean error rate per allele averaged 0.017 and the mean error rate per locus averaged 0.022 (Table 2).

In our evaluation of 64 nuclear SSRs to compare 32 legacy cultivars and eight colchicine-induced polyploids of Mxg with the Mxg type-specimen '1993–1780', we observed 49 cases of putative ADO and two cases of FAs on 25 SSRs (Table 6). Six of the 11 SSR loci that had a nonzero error rate in the factorial experiment, also had genotyping errors in the larger comparison of Mxg accessions (Table 5 and Table S2). Nearly half of the ADOs were corrected by technical replication (conducting another PCR) and nearly half again were corrected by changing the fluorescent dye used (PET was more error-prone than the others), for a cumulative correction rate of 0.71 (Table 6). Subsequently, another fifth of the ADOs were corrected by using high-fidelity DNA polymerase and an additional quarter was corrected by re-extraction of DNA followed by additional PCR, for a total cumulative correction rate of 0.82.

Efficacy of high-throughput sequencing for fingerprinting cultivars

Pairs of technical replicates from the Mxg type specimen and hexaploid Mxg did not cluster together in

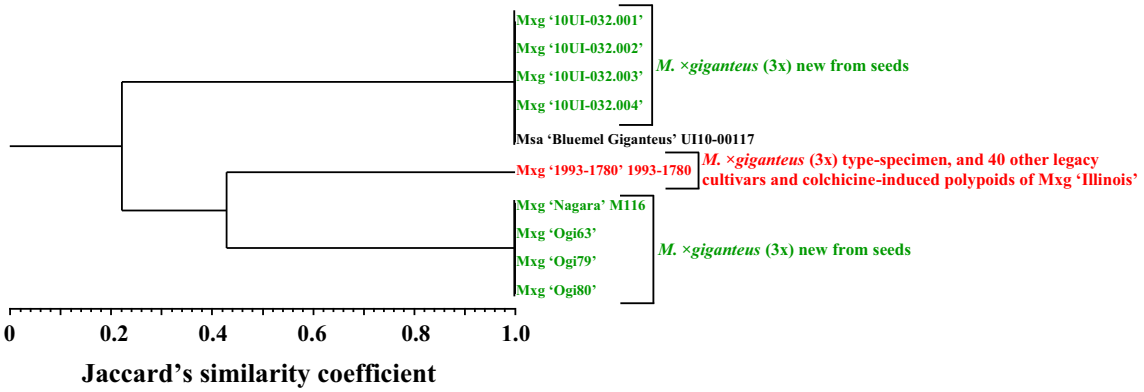


Fig. 3 UPGMA cluster analysis based on five chloroplast SSR loci for 48 *Miscanthus* × *giganteus* accessions and the female parent of the 10UI-032 cross, *M. sacchariflorus* 'Bluemel Giganteus' (4x), with Jaccard's similarity coefficients calculated from binary data of 10 alleles and revealed by software NTSYS-pc v. 2.21m.

UPGMA based on RAD-seq data (Fig. S1), and were sufficiently spread out to indicate that all legacy Mxg and polyploid Mxg were genetically indistinguishable. The average nucleotide dissimilarity between pairs of Mxg replicates, which is an estimate of the error rate, was 0.00536 ± 0.00014. At a threshold of 0.00563 (mean + 2SE) for distinguishing true genetic difference from error rate, two of the hexaploids ('Mxg 2x-3' and 'Mxg 2x-10') would be considered distinct from each other and from legacy Mxg, and all legacy Mxg tested would be considered identical (Table S2). However, these two hexaploid lines also had the lowest read counts of the dataset (Table S2), decreasing the chances that all of their alleles would be captured in sequencing and therefore decreasing their apparent similarity to any genetically identical accessions.

Table 4 List of *Miscanthus sacchariflorus* and *M. sinensis* specific alleles

SSR name	Allele (bp)	Specificity	
		M. sacchriflorus	*M. sinensis*
GSSR_023	163	+	–
ESSR_001	84	+	–
ESSR_008	114	+	–
ESSR_014	152	+	–
ESSR_017	167	+	–
ESSR_024	186	+	–
Mis.Fluor.10	158	+	–
Mis.Fluor.11	76	+	–
Mis.Fluor.40	158	+	–
GSSR_051	145	–	+
ESSR_024	191	–	+
Mis.Fluor.21	97	–	+
Mis.Fluor.31	275	–	+
Mis.Fluor.40	157	–	+

"+" = allele present in all tested accessions of the species.
"–" = allele absent in all tested accessions of the species.

Clustering patterns between diploid Msa, tetraploid Msa, Msi, and Mxg were similar to those found with SSR data (Figs 1 and 4). The nucleotide dissimilarity between the pair of diploid Msa replicates was 0.00421, reflecting the improved sequence coverage of a diploid relative to triploids (or doubled triploids) for a given sequencing depth.

Empirical and theoretical data suggest that a minimum depth of 20–30 sequencing reads per locus would be needed to capture all alleles of Mxg at least 95% of the time (Fig. 4). Given that *Miscanthus* has undergone recent genome duplication and that Mxg is triploid, we would expect triploid Mxg to have six copies of each locus (up to six alleles) in contrast with a maximum of two alleles expected in a typical diploid organism (or four alleles in diploid *Miscanthus* because of the genome duplication). For the four pairs of Mxg technical replicates, the median read depth was 29; 2029 RAD tags (77.5%) had at least one sample with a read depth below 20, and 1291 RAD tags (49.3%) had at least one sample with a read depth below 10, indicating that many alleles were not captured by RAD-seq. Using data

Table 5 Analysis of variance for sources of genotyping error (error per allele) associated with SSRs on *Miscanthus*. DNA was twice extracted for each of two accessions, Mxg '1993–1780' and Mxg 'Freedom' UI10-00119, and then PCR amplifications with 63 SSR primer pairs were repeated four times for each extraction. For the each DNA extraction all the steps of technical replication were performed independently

Source of variation	df	*P*-value
SSR	62	<0.0001
Allele (SSR)	59	0.9848
Accession	1	0.9001
SSR × Accession	62	1.0000
Allele(SSR) × Accession	59	1.0000
DNA extraction (Accession)	2	0.9356

Table 6 Putative allele dropouts (ADOs) and false alleles (FAs) of 64 nuclear SSRs observed on 32 legacy cultivars and eight colchicine-induced polyploids of *Miscanthus* × *giganteus* relative to the *M.* × *giganteus* type-specimen, '1993–1780'. Confirmation and correction genotyping errors is shown for the following four procedures, which were applied sequentially: repetition of PCR (Rep), change of fluorescent dye (Dye), use of high fidelity polymerases (HF polymerase), and DNA extraction from new collected tissue (Re-extraction)

				Allele		ADO/FA correction			
SSR	Dye used	Allele scored in Mxg '1993–1780'	Accession identifier	Dropped	False	Rep	Dye	HF polymerase	Re-extraction
GSSR_035	FAM	165; 171	UI10-00125	165; 171		Y			
GSSR_037	PET→FAM	145; 148	UI10-00125	145; 148		N	Y		
	PET→FAM	145; 148	UI11-00026	145; 148		N	Y		
GSSR_020	PET→FAM	117; 126	UI10-00118	117; 126		N	N	N	Y
GSSR_051	VIC	141; 145; 146; 150; 153	N1	146		Y			
	VIC		N2	146		Y			
	VIC		N3	146		Y			
	VIC		NEF-6	146		Y			
GSRR_054	NED	171	UI11-00026	171		Y			
	NED		UI10-00118	171		Y			
	NED		NEF-6	171		Y			
	PET→FAM		NEF-9	171	176	N/N	N/N	Y/N	
GSRR_048	NED→FAM	177; 183	UI10-00118	177; 178		N	N	N	Y
GSRR_044	NED→FAM	118	UI10-00125	118		N	Y		
	NED→FAM		UI11-00026	118		N	Y		
	NED→FAM		UI10-00118	118		N	Y		
	NED→FAM		UI10-00120	118		N	Y		
GSRR_045	FAM	122	UI10-00118	122		Y			
	FAM→VIC		UI11-00026	122		N	Y		
	FAM→VIC		UI10-00125	122		N	Y		
GSSR_061	VIC	128	UI12-00013	128		Y			
ESSR_008	PET	108; 114; 120; 125	UI10-00118	114		Y			
	PET		UI10-00125	108		Y			
GSSR_038	NED	138	UI10-00121	138		Y			
GSRR_077	VIC→FAM	113; 116	UI10-00125	113		N	N	Y	
	VIC→FAM		UI10-00108	113		N	Y		
ESRR_028	PET→FAM	130	NEF-7	130	133	N/N	N/N	Y/N	
GSSR_022	NED→FAM	153; 166; 174; 198	UI10-00113	174		N	N	N	N
	NED→FAM		UI10-00114	174		N	N	N	N
	NED→FAM		UI10-00115	174		N	N	N	N
Mis.fluor.7	PET	203	UI10-00125	203		Y			
	PET		UI10-00111	203		Y			
Mis.fluor.9	FAM→VIC	100; 182; 187	N2	187		N	N	N	
	FAM→VIC		N3	187		N	N	N	
Mis.fluor.19	PET→FAM	110	UI11-00026	110		N	Y		
	PET→FAM		UI10-00125	110		N	Y		
Mis.fluor.22	VIC→FAM	157	UI10-00118	157		N	Y		
	VIC→FAM		UI10-00125	157		N	Y		
Mis.fluor.25	FAM	136	UI11-00026	136		Y			
Mis.fluor.26	VIC	134	UI11-00026	134		Y			
Mis.fluor.27	PET	135	UI10-00118	135		Y			
Mis.fluor.29	FAM	173; 178	UI11-00026	173; 178		Y			
	FAM		UI10-00125	173; 178		Y			
Mis.fluor.31	PET	153	UI10-00118	153		Y			
	PET		UI11-00026	153		Y			

(continued)

Table 6 (continued)

SSR	Dye used	Allele scored in Mxg '1993–1780'	Accession identifier	Allele		ADO/FA correction			
				Dropped	False	Rep	Dye	HF polymerase	Re-extraction
Mis.fluor.37	FAM→VIC	106; 135; 144; 158	UI10-00113	106		N	N	N	N
	FAM→VIC		UI10-00114	106		N	N	N	N
	FAM→VIC		UI10-00115	106		N	N	N	N
Mis.fluor.39	PET→FAM	90; 134; 140; 154	NEF-8	140; 154		N	N	N	
ADO correction rate						0.45	0.48	0.21	0.25
Cumulative ADO correction						0.45	0.71	0.78	0.82

Y, yes; N, no.

from the four pairs of Mxg technical replicates, loci were separated into groups based on minimum read depth and dissimilarity between pairs of technical replicates. Dissimilarity reached a minimum when loci had a minimum read depth of 20–30, then increased again for loci with a higher read depth (Fig. 4), possibly because loci with a minimum read depth greater than 30 represented repetitive regions in the genome.

When all 24 sequenced accessions were considered, the median depth per accession per RAD tag in our study was 23 and the fifth percentile was 4. To get 95% of RAD-tags to a minimum of 20–30 reads per individual, we would therefore have to increase our depth five to seven times. This could possibly be accomplished with fewer than four additional sequencing runs if the distribution of reads among individuals and loci was random from run to run.

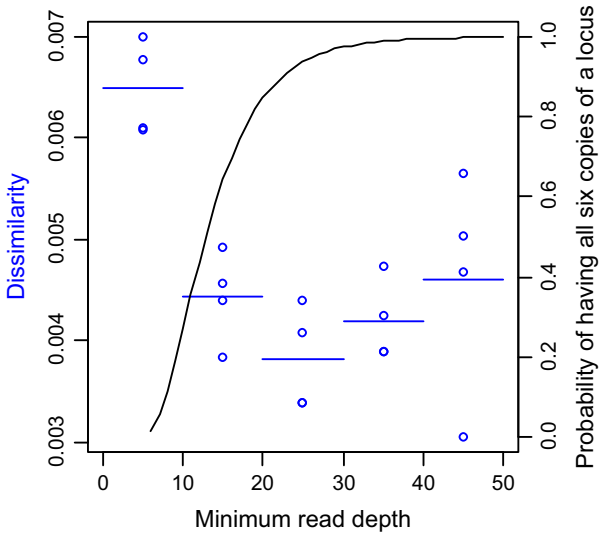


Fig. 4 Nucleotide dissimilarities between Mxg technical replicates in RAD-seq, and probability of capturing each of six copies of a locus, vs. minimum read depth per locus. Horizontal lines represent averages across the four pairs of Mxg replicates. Curve representing the probability of sampling all six copies of a locus in Mxg was calculated using the formula $p = \frac{6! * \{ {r \atop 6} \}}{6^r}$ where r is the number of reads, p is the probability of the set of reads including each of six locus copies at least once, and curly brackets indicate the Stirling number of the second kind.

Discussion

Past and present Mxg genetic diversity

From the marker data, we deduce that all of the Mxg legacy cultivars from Europe and North America that we tested were derived via vegetative propagation from a single genet. Thus, the Mxg clone that Aksel Olsen introduced from Japan to Denmark in the 1930s was subsequently distributed throughout Europe and North America, and given new cultivar names during that process.

We also identified somatic mutants (i.e. horticultural sports) of the Mxg clone originally imported to Europe. For example, three of the colchicine-induced hexaploids of Mxg 'Illinois' ('Illinois-6x.07 (Mxg2x-7)', 'Illinois-6x.08 (Mxg2x-8)', 'Illinois-6x.09 (Mxg2x-9)') differed from the type-specimen (and 'Illinois') by the same two SSR presence-absence differences but were identical to each other, indicating that all three hexaploids were likely derived from the same mutated callus cell *in-vitro*. Given that colchicine is a known mutagen (Tiwari & Khanorkar, 1984; Luckett, 1989), the production of some somatic mutant plants from colchicine-treated callus would be expected. Similarly, although Mxg 'Gilded Tower' had an identical SSR profile as the type-specimen, the former had longitudinally yellow-striped variegated leaves in contrast with the solid green leaves of the latter, indicating that 'Gilded Tower' has a mutation for yellow-stripe (possibly a chimera) but is part of the same somatic lineage as all of the other legacy cultivars. Failure of the tested markers to detect the mutation for yellow-striped leaves in 'Gilded Tower' highlights the difficulty of tagging single mutation differences among

otherwise isogenic lines. For the five Mxg legacy cultivars that had 98–99% genetic similarity to the type-specimen, we would be hesitant to rule out genotyping error and conclude that they are truly somatic mutants because these particular accessions came to us as one-time shipments of leaf samples only, and we were unable to obtain new leaf samples that would have allowed us to more fully test reproducibility of the one to three differing alleles.

For long-lived, vegetatively propagated, heterozygous crops, such as grape (This *et al.*, 2006; Cipriani *et al.*, 2010) and sugarcane (Schenck *et al.*, 2004), it has been common for a single cultivar (i.e. a multilocus genotype) to be given multiple names during distribution to new locations and over time, and for horticultural sports to be selected either inadvertently or purposely. Thus, the history of *Miscanthus* in Europe and North America during the last century has been typical of vegetatively propagated crops.

Our results, which indicated a near-absence of genetic diversity among the Mxg legacy cultivars in Europe and North America, were consistent with the results of Greef *et al.* (1997) and Hodkinson *et al.* (2002a). In contrast with our results, Chouvarine *et al.* (2012) reported considerable genetic differentiation based on exome sequencing among all six North American Mxg accessions that they studied, including 'Illinois' and three accessions that represented the cultivar 'Freedom'. However, Chouvarine *et al.* (2012) did not provide estimates of the genotyping error rate, via technical replication, associated with their method. Given that the frequency of pairwise differences observed by Chouvarine *et al.* (2012) was within typical error rates based on other sequencing studies (Hedges *et al.*, 2009; Dewey *et al.*, 2012; Nielsen *et al.*, 2012), and that our study found no nuclear or cytoplasmic genetic differences between 'Illinois' and 'Freedom', it is likely that the variation observed by Chouvarine *et al.* (2012) represented genotyping error, rather than true genetic differences among the accessions. Additionally, exome sequencing may have also introduced error due to differences in gene expression among their samples. Nevertheless, given the limitations associated with the small but nonzero error rates of exome sequencing and SSRs, we cannot rule out the possibility of one or more somatic mutations among the Mxg accessions tested by Chouvarine *et al.* (2012) or between 'Freedom' and 'Illinois' in our study.

Potential for future Mxg genetic diversity

New Mxg genotypes from seed (i.e. sexual reproduction) represent the future of *Miscanthus* as a feedstock crop for bioenergy. Although none of the eight new from seed triploid Mxg genotypes that we studied have yet been commercialized in the United States, they demonstrate that substantial genetic diversity of Mxg can be obtained if plant breeders make crosses between selected tetraploid Msa and diploid Msi parents and/or collect allotriploid seed from wild sympatric populations of these parental species. For example, the two most genetically similar Mxg full sibs that we studied, '10UI-032.002' and '10UI-032.003', had 11 times more presence–absence SSR differences between themselves, than the most diverged putative somatic mutant Mxg legacy cultivar compared with the type-specimen. As expected, genetic differentiation among Mxg individuals derived from different crosses was greater than comparisons within crosses (Fig. 1), and cytoplasmic genetic diversity was also obtained from the new crosses (Fig. 3). Moreover, the high levels of genetic diversity in our modest sampling of Mxg's parental species, Msa and Msi, indicates that even greater gains in diversity can be made from future crosses. Selection within Msa and Msi for adaptation, yield, and interspecific combining ability will lead to further improved cultivars of Mxg that will provide value to growers. Although point mutations can also provide valuable variation for clonally propagated horticultural crops, we would not expect such variation to be sufficient to substantially reduce the risk of catastrophic crop failure due to disease or insect pressure, or to enable rapid breeding for increased yield and improved adaptation to a diversity of environments. While mutation breeding has had successes (Jain, 2005) and could be a useful complementary approach for breeding Mxg, we expect that harnessing the great natural diversity available in Mxg's parental species would be the most advantageous strategy for improving this new crop. We envision that within the next decade, tens or hundreds of new Mxg triploid cultivars derived primarily from controlled crosses will become available to farmers. Although vegetative propagation of triploid Mxg is expensive relative to seed propagation, the low risk of invasiveness associated with the sterile triploids makes them the preferred deployment option in the short to medium term for areas where *Miscanthus* is not native. In the long term, seed-propagated Mxg cultivars at the tetraploid or diploid level that have been bred for reduced dispersal via nonshattering seed and/or sterility will provide a more economical option still, and with low risk of invasiveness (i.e. Mxg will become a fully domesticated crop).

Marker utility and limitations

As a new crop, the development of molecular markers for *Miscanthus* has occurred mostly within the last 5 years. A number of recent studies have identified SSR primers that amplify *Miscanthus* DNA (Hung *et al.*,

2009; de Cesare *et al.*, 2010; Zhao *et al.*, 2011; James *et al.*, 2012; Jiang *et al.*, 2012; Swaminathan *et al.*, 2012; Yu *et al.*, 2013), however, additional testing of newly identified SSRs is typically needed to determine their utility and fidelity. Out of the 85 SSRs previously reported to amplify *Miscanthus* DNA that we tested, we confirmed that 64 nuclear and five chloroplast markers were informative. The average SSR error rates we observed were similar to those reported in other studies (Broquet & Petit, 2004; Pompanon *et al.*, 2005; Zhang *et al.*, 2006). Moreover, some of the SSRs we tested had greater fidelity (lower error rates) than others. We observed no errors for 39 of the informative SSRs that we tested (Tables 2 and 6), which should serve as useful set of validated markers for future studies of *Miscanthus*.

The SSR markers used in the current study were more than sufficient for differentiating between Mxg full-sibs. However, obtaining study-specific and marker-specific genotyping error rates is essential for correctly interpreting comparisons of closely related individuals, such as clonal lineages. Similar conclusions have been reached for other clonally propagated crops, such as cacao and *Musa* (Zhang *et al.*, 2006; Christelova *et al.*, 2011). In particular, the strategy that we employed, of running additional PCR reactions for only the nearly identical entries, was also suggested by Zhang *et al.* (2006) and Christelova *et al.* (2011) and as a cost-effective approach for distinguishing between SSR genotyping errors and somatic mutant lineages.

Estimating genetic distance thresholds to differentiate clones is an important use of molecular markers. If SSR genotyping errors were not corrected for comparisons among the Mxg legacy cultivars, we found that a genetic similarity threshold of ~0.98 would be indicative of genotypes that may in fact be identical. However, with our multistage marker validation protocol, all fully validated Mxg legacy accessions had identical (100%) marker profiles. Based on our data for somatic mutants induced by colchicine, we found that a similarity threshold of 0.98 for validated marker genotypes would indicate that the accessions are nearly isogenic. Among the Mxg full-sibs, however, validated marker genotypes resulted in genetic similarity values that ranged from 0.41 to 0.75; thus, a threshold below ~0.75 would indicate different genets from the same pair of parents. There is more than a 20 point difference between the thresholds indicative of somatic mutations of a common genet from the threshold indicative of full siblings. Given such a large spread in genetic distance thresholds, there can be little confusion as to which Mxg genotypes are somatic mutations derived from Aksel Olsen's introduction and which are from new crosses.

Compared with SSRs, we found that RAD-seq had a high error rate relative to signal of true genetic differentiation (Figs 1 and 4); the ratio of average dissimilarity among legacy Mxg (error) to average dissimilarity between legacy Mxg and Msi var. *condensatus* 'Cabaret' was 0.02 with SSRs and 0.52 with RAD-seq. Sequencing errors inherent in Illumina technology partially explain differences between technical replicates, but more importantly the read depth was too low to guarantee sequencing of all alleles. Read depth was not reported by Chouvarine *et al.* (2012), making it difficult to determine if the differences they saw between Mxg accessions were due to a similar issue; however, their tree depicting relationships between Mxg and Msi accessions had similar relative branch lengths to ours using RAD-seq, and they reported that only 14.64% of all reads were in contigs found in all seven samples, implying a high missing data rate. Our cost for RAD-seq was approximately $20/sample not including labor, but to capture all alleles at 95% of loci, we would need to increase our read depth several-fold (Fig. 4), increasing the cost proportionally. Thus, in our study, SSRs are currently more cost-effective for determining if two accessions are genetically identical or not, whereas high-throughput sequencing is a better choice for analyses in which thousands of markers are desired but a relatively high error rate is acceptable (e.g. broad surveys of genetic diversity).

Genetic variation within the Mxg legacy cultivars is critically low, even taking into account somatic mutants. If large areas were planted to only this one genotype, the risk associated with a pest or disease outbreak would be high. Moreover, the lack of genetic diversity limits the geography to which this crop is best adapted. In the United States, Mxg 'Illinois' appears to be best adapted to the moist, mid-latitude areas of USDA hardiness zone 7, such as those in Kentucky and Tennessee. In the southern coastal plain of the United States, Mxg 'Illinois' flowers too early (in mid-summer) to maximize yield potential, whereas the late October flowering time in the Midwest United States results in better adaptation for yield in this more northern environment. However, in hardiness zone 6 and colder, winterkill and winter damage can occur on new Mxg 'Illinois' plantings during the establishment year. Similarly, Clifton-Brown *et al.* (2001) observed that Mxg did not overwinter well in Denmark and Sweden during the establishment year. Thus, there is currently a great need for additional Mxg genetic diversity that can be deployed to farmers' fields in support of the expanding bioenergy industry. The development of new triploid Mxg cultivars from crosses between selected parents can meet this need.

Acknowledgments

We wish to thank the Energy Biosciences Institute for funding this research. We also thank Dean Tiessen and Gerald van Koeverden of New Energy Farms, Stanisław Jeżowski of the Institute of Plant Genetics, Polish Academy of Science, and Brian Baldwin of Mississippi State University for providing materials used in this study.

References

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

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetic studies. *Molecular Ecology*, **13**, 3261–3273.

Broquet T, Petit E (2004) Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, **13**, 3601–3608.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait, (2011) *Stacks*: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

de Cesare M, Hodkinson TR, Barth S (2010) Chloroplast DNA markers (cpSSRs, SNPs) for *Miscanthus*, *Saccharum* and related grasses (Panicoideae, Poaceae). *Molecular Breeding*, **26**, 539–544.

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

Chae WB, Hong SAEJ, Gifford JM, Rayburn AL, Sacks EJ, Juvik JA (2014) Plant morphology, genome size, and SSR markers differentiate five taxonomic groups among accessions in the genus *Miscanthus*. *Global Change Biology Bioenergy*, doi: 10.1111/gcbb.12101.

Chouvarine P, Cooksey AM, McCarthy FM, Ray DA, Baldwin BS, Burgess SC, Peterson DG (2012) Transcriptome-based differentiation of closely-related *Miscanthus* Lines. *PLoS ONE*, **7**, e29850.

Christelova P, Valarik M, Hribova E, Van dhI, Channeliere S, Roux N, Dolezel J (2011) A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB PLANTS*, **2011**, plr024. doi: 10.1093/aobpla/plr024.

Cipriani G, Spadotto A, Jurman I *et al*. (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theoretical and Applied Genetics*, **121**, 1569–1585.

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, 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.

Dewey F, Pan S, Wheeler MT, Quake SR, Ashley EA (2012) DNA sequencing: clinical applications of new DNA sequencing technologies. *Circulation*, **125**, 931–944.

Douhovnikoff V, Dodd RS (2003) Intra-clonal variation and a similarity threshold for identification of clones: application to *Salix exigua* using AFLP molecular markers. *Theoretical and Applied Genetics*, **106**, 1307–1315.

Dwiyanti MS, Rudolph A, Swaminathan K *et al*. (2013) Genetic analysis of putative triploid *Miscanthus* hybrids and tetraploid *M. sacchariflorus* collected from sympatric populations of Kushima, Japan. *Bioenergy Research*, **6**, 486–493.

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.

Greef JM, Deuter M (1993) Syntaxonomy of *Miscanthus* × *giganteus* Greef et Deu. *Angewandte Botanik*, **67**, 87–90.

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.

Heaton E, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of *Miscanthus*. *Global Change Biology*, **14**, 1–15.

Hedges D, Burges D, Powell E *et al*. (2009) Exome sequencing of a multigenerational human pedigree. *PLoS ONE*, **4**, e8232.

Hirayoshi I, Nishikawa K, Kubono M, Murase T (1957) Cyto-genetical studies on forage plants (VI) On the chromosome number of Ogi (*Miscanthus sacchariflorus*). *Research Bulletin of the Faculty of Agriculture, Gifu University*, **8**, 8–13.

Hodkinson TR, Renvoize S (2001) Nomenclature of *Miscanthus* × *giganteus* (*Poaceae*). *Kew Bulletin*, **56**, 759–760.

Hodkinson TR, Chase MW, Renvoize SA (2002a) Characterization of a genetic resource collection for *Miscanthus* (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. *Annals of Botany*, **89**, 627–636.

Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA (2002b) 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.

Hung K-H, Chiang T-Y, Chiu C-T, Hsu T-W, Ho C-W (2009) Isolation and characterization of microsatellite loci from a potential biofuel plant *Miscanthus sinensis* (Poaceae). *Conservation Genetics*, **10**, 1377–1380.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bulletin de la Société vaudoise des sciences naturelles*, **44**, 223–270.

Jain SM (2005) Major mutation-assisted plant breeding programs supported by FAO/IAEA. *Plant Cell, Tissue and Organ Culture*, **82**, 113–123.

James BT, Chen C, Rudolph A *et al*. (2012) Development of microsatellite markers in autopolyploid sugarcane and comparative analysis of conserved microsatellites in sorghum and sugarcane. *Molecular Breeding*, **30**, 661–669.

Jiang JX, Wang ZH, Tang BR, Xiao L, Ai X, Yi ZL (2012) Development of novel chloroplast microsatellite markers for *Miscanthus* species (Poaceae). *American Journal of Botany*, **6**, e230–e233.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Kabelka E, Franchino B, Francis DM (2002) Two Loci from *Lycopersicon hirsutum* LA407 Confer Resistance to Strains of *Clavibacter michiganensis* subsp. *Michiganensis*. *Phytopathology*, **9**, 504–510.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Lasso E (2008) The importance of setting the right genetic distance threshold for identification of clones using amplified fragment length polymorphism: a case study with five species in the tropical plant genus *Piper*. *Molecular Ecology Resources*, **8**, 74–82.

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

Luckett DJ (1989) Colchicine mutagenesis is associated with substantial heritable variation in cotton. *Euphytica*, **42**, 177–182.

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

International code of nomenclature for algae, fungi and plants (Melbourne code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011 (2012) Regnum Vegetabile, 154 (eds McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Marhold K, Prado J, Prud'homme vRW, Smith GF, Wiersema JH, Turland NJ), 240 p. Koeltz Scientific Books. ISBN 978-3-87429-425-6.

Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, **12**, R112.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE*, **7**, e37558.

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.

Poland JA, Brown PJ, Sorrells ME, Jannik J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

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

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

Schenck S, Crepeau MW, Wu KK, Moore PH, Yu Q, Ming R (2004) Genetic diversity and relationships in native Hawaiian *Saccharum officinarum* sugarcane. *Journal of Heredity*, **95**, 327–331.

Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.

Scurlock JMO (1999) *Miscanthus: A Review of European Experience with a Novel Energy Crop*. ORNL Technical Memorandum TM-13732, 18 pp. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Somerville C, Youngs H, Taylor C, Davis SC, Long SP (2010) Feedstocks for lignocellulosic biofuels. *Science*, **329**, 790–792.

Swaminathan K, Chae WB, Mitros T *et al*. (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers reveals recent allo-tetraploidy. *BMC Genomics*, **13**, 142–159.

This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. *Trends in Genetics*, **22**, 511–519.

Thurber CS, Ma JM, Higgins RH, Brown PJ (2013) Retrospecitve genomic analysis of sorghum adaptation to temperate zone grain production. *Genome Biology*, **14**, R68.

Tiwari SP, Khanorkar SM (1984) Colchicine-induced true breeding miniature mutant in groundnut. *Current Science*, **53**, 1262–1263.

Toonen RJ, Hughes S (2001) Increased throughput for fragment analysis on an ABI PRISM automated sequencer using a membrane comb and STRand software. *BioTechniques*, **31**, 1320–1324.

Yu J, Zhao H, Zhu T, Chen L, Peng J (2013) Transferability of rice SSR markers to *Miscanthus sinensis*, a potential biofuel crop. *Euphytica*, **191**, 455–468.

Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA (2006) Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. *Crop Science*, **46**, 2084–2092.

Zhao H, Yu J, You FM, Luo M, Peng J (2011) Transferability of microsatellite markers from *Brachypodium distachyon* to *Miscanthus sinensis*, a potential biomass crop. *Journal of Integrative Plant Biology*, **53**, 232–245.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1 UPGMA tree calculated from nucleotide dissimilarity from RAD-seq data aligned to 196,275 nucleotides in the *M.* × *giganteus* transcriptome sequence. Technical replicates are indicated by letters A and B.

Table S1 Matrix of Jaccard's similarity coefficients for *M. sacchariflorus* accessions.

Table S2 Read counts and nucleotide dissimilarities of Mxg samples using RAD-seq data.