

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications from USDA-ARS / UNL Faculty

U.S. Department of Agriculture: Agricultural
Research Service, Lincoln, Nebraska

2011

Hierarchical classification of switchgrass genotypes using SSR and chloroplast sequences: ecotypes, ploidies, gene pools, and cultivars

J. E. Zalapa

University of Wisconsin, jezalapa@wisc.edu

D. L. Price

University of Wisconsin

S. M. Kaeppler

University of Wisconsin, smkaeppl@wisc.edu

C. M. Tobias

USDA-ARS, christian.tobias@ars.usda.gov

M. Okada

USDA-ARS

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/usdaarsfacpub>

Zalapa, J. E.; Price, D. L.; Kaeppler, S. M.; Tobias, C. M.; Okada, M.; and Casler, M. D., "Hierarchical classification of switchgrass genotypes using SSR and chloroplast sequences: ecotypes, ploidies, gene pools, and cultivars" (2011). *Publications from USDA-ARS / UNL Faculty*. 1306.
<https://digitalcommons.unl.edu/usdaarsfacpub/1306>

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Agricultural Research Service, Lincoln, Nebraska at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications from USDA-ARS / UNL Faculty by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

J. E. Zalapa, D. L. Price, S. M. Kaeppler, C. M. Tobias, M. Okada, and M. D. Casler

Hierarchical classification of switchgrass genotypes using SSR and chloroplast sequences: ecotypes, ploidies, gene pools, and cultivars

J. E. Zalapa · D. L. Price · S. M. Kaeppler ·
C. M. Tobias · M. Okada · M. D. Casler

Received: 10 August 2010 / Accepted: 22 October 2010 / Published online: 23 November 2010
© Springer-Verlag (outside the USA) 2010

Abstract Switchgrass (*Panicum virgatum* L.) is an important crop for bioenergy feedstock development. Switchgrass has two main ecotypes: the lowland ecotype being exclusively tetraploid ($2n = 4x = 36$) and the upland ecotype being mainly tetraploid and octaploid ($2n = 8x = 72$). Because there is a significant difference in ploidy, morphology, growth pattern, and zone of adaptation between and within the upland and lowland ecotypes, it is important to discriminate switchgrass plants belonging to different genetic pools. We used 55 simple sequence repeats (SSR) loci and six chloroplast sequences to identify patterns of variation between and within 18 switchgrass cultivars representing seven lowland and 11 upland cultivars from different geographic regions and of varying ploidy levels.

We report consistent discrimination of switchgrass cultivars into ecotype membership and demonstrate unambiguous molecular differentiation among switchgrass ploidy levels using genetic markers. Also, SSR and chloroplast markers identified genetic pools related to the geographic origin of the 18 cultivars with respect to ecotype, ploidy, and geographical, and cultivar sources. SSR loci were highly informative for cultivar fingerprinting and to classify plants of unknown origin. This classification system is the first step toward developing switchgrass complementary gene pools that can be expected to provide a significant heterotic increase in biomass yield.

Communicated by M. Frisch.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1488-1) contains supplementary material, which is available to authorized users.

J. E. Zalapa (✉)
USDA-ARS, Vegetable Crops Research Unit,
Department of Horticulture,
University of Wisconsin, Madison, WI, USA
e-mail: jezalapa@wisc.edu

D. L. Price · S. M. Kaeppler
Department of Agronomy,
University of Wisconsin, Madison, WI, USA

M. D. Casler
USDA-ARS, US Dairy Forage Research Center,
Madison, WI, USA

C. M. Tobias · M. Okada
USDA-ARS, Western Regional Research Center,
Albany, CA, USA

Introduction

Switchgrass (*Panicum virgatum* L.) is an important crop used as a bioenergy feedstock. It is water and nutrient-efficient, high in biomass yield, tolerant to many biotic and abiotic stresses, and adapted across a wide geographic range, including marginal lands (Sanderson et al. 2007). The broad geographic adaptation of switchgrass in North America is reflected by a wide array of phenotypic variability with some very strong associations between phenotype and environment (Casler and Boe 2003; Casler et al. 2007a).

Switchgrass is an outcrossing, perennial C4 grass, mostly self-incompatible with two phenotypically distinct ecotypes across its distribution: upland (e.g., cultivars Blackwell and Summer) and lowland (e.g., cultivars Alamo and Kanlow) ecotypes. The upland ecotype is usually associated with xeric to mesic habitats and northern latitudes and the lowland ecotype is prevalent in more hydric habitats and southern latitudes (Casler et al. 2010). Variation among and within ecotypes is believed to be related mainly

to latitude of origin and is usually reflected in reproductive phenology and winter-hardiness (Casler et al. 2007b). For example, southern switchgrass germplasm tends to flower later in the season, allowing more time for vegetative growth and therefore increased yields. Thus, lowland cultivars tend to be larger (≥ 2.7 m, tall) and produce more biomass than the comparatively shorter (≤ 2.4 m, tall) upland cultivars. Switchgrass can be divided into a ploidy series from $2n = 2x = 18$ to $2n = 12x = 108$ (Nielson 1944) with possible ploidal variation including aneuploids (Costich et al. 2010). Lowland ecotypes are tetraploid ($2n = 4x = 36$) and upland ecotypes are mainly tetraploid and octaploid ($2n = 8x = 72$) (Hopkins et al. 1996; Hultquist et al. 1996, 1997; Lu et al. 1998; Costich et al. 2010).

The upland and lowland ecotypes have been clearly assigned in distinct genetic clusters based on polymorphisms in chloroplast sequences (Hultquist et al. 1996; Missaoui et al. 2006), random amplified polymorphic DNA (RAPD; Gunter et al. 1996), a nuclear gene encoding plastid acetyl-CoA carboxylase (Huang et al. 2003), and restriction fragment length polymorphism (RFLP) markers (Missaoui et al. 2006). Little or no differentiation has been detected within ecotypes and ploidy levels in such studies. More recently, simple sequence repeats (SSR) have been identified in expressed sequence tags and genomic sequences that provide highly variable, reproducible marker systems for genotyping switchgrass (Tobias et al. 2005, 2006, 2008; Narasimhamoorthy et al. 2008). Switchgrass SSR diversity studies have utilized a limited number of lowland or upland cultivars represented by a few individuals per cultivar and have not identified clear differences beyond ecotype differentiation (Narasimhamoorthy et al. 2008; Cortese et al. 2010).

Because there are significant differences in morphology, growth pattern, and adaptation zones among switchgrass accessions, it is important to be able to easily discriminate between genetic pools with differential environmental adaptation. The reported presence of hybrid vigor in upland \times lowland crosses and the possibility that the two ecotypes may act as natural heterotic groups (Martinez-Reyna and Vogel 2008; Vogel and Mitchell 2008) creates a further need for accurate and efficient discrimination between switchgrass genetic pools.

The genetic differentiation of switchgrass cultivars should be feasible because they represent well-differentiated genetic pools from a wide range of ecosystems and are not far removed from wild populations. Most cultivars are simple seed increases of prairie-remnant populations or were developed from a limited number of breeding cycles, no more than two or three generations removed from the wild (Sanderson et al. 2007; Casler and Boe 2003; Casler 2010). Moreover, since early adoption and deployment of switchgrass for bioenergy production has been based on a

limited number of existing cultivars (Casler 2010), a greater knowledge of the genetic structure of switchgrass genetic pools will be essential for the preservation of genetic diversity during breeding. Information regarding the amount of genetic diversity present among switchgrass cultivars will be valuable to understand the diversity in natural populations, in breeding programs, and in public germplasm collections. Characterization of genetic diversity will be essential to select the most diverse genotypes for germplasm improvement and hybridization to develop superior cultivars, maintain genetic integrity of the cultivars, conserve germplasm for regional use, and for genetics and genomics projects including population structure, gene flow, and mapping studies.

The aim of this study was to use SSR markers and chloroplast DNA (cpDNA) sequences to study the genetic variation among and within switchgrass cultivars, which represent the upland and lowland ecotypes, multiple ploidy levels, and a broad geographic range. The specific objectives of this study were to: (1) discriminate upland and lowland ecotypes from multiple ecological regions, (2) identify the potential for SSR markers to discriminate ploidy levels, (3) identify patterns of DNA polymorphisms related to geographic origin and ecological regions, and (4) identify the potential of SSR markers as a tool to fingerprint specific switchgrass cultivars.

Materials and methods

Plant materials

Eighteen cultivars were chosen for this study based on a broad range of diversity and the availability of viable seed (7 lowland and 11 upland; Table 1). All the 18 cultivars can be considered ecotypes since they are simple seed increases of prairie-remnant populations, represent no more than two or three generations removed from the wild, or are close derivatives of ecotypes, having undergone only one or two generations of selection, not sufficient to change their region of adaptation. Thus, all 18 cultivars are still highly representative of local germplasm from their respective regions. Seeds from each cultivar were obtained from the USDA National Genetic Resources Program (NGRP; <http://www.ars-grin.gov/npgs/index.html>), and a total of 1,200 (50–100/cultivar) single seeds were planted. The resulting seedlings yielded 370 transplants (single genotypes) that were raised in the greenhouse from which 184 samples were selected to represent 18 cultivars (Table 1). The goal was to represent each cultivar with at least eight plants, but this was not accomplished in some cases due to reduced germination rates and/or seedling mortality. Additionally, leaves from eight switchgrass plants of uncertain origin

Table 1 Cultivars used to represent three major switchgrass taxa: lowland tetraploid, upland tetraploid, and upland octaploid, including number of plants (*n*), DNA content, and inferred ploidy

Cultivar	Origin	Ecotype ^a	<i>n</i>	DNA content		Ploidy
				Mean	SD	
Alamo	Texas	Lowland	16	2.91	0.04	4×
Kanlow	Oklahoma	Lowland	16	2.76	0.04	4×
Miami	Florida	Lowland	2	2.88	0.03	4×
SG5	Unknown	Lowland ^b	9	3.15	0.06	4×
Stuart	Florida	Lowland	2	2.94	0.03	4×
Timber	North Carolina	Lowland	9	2.85	0.03	4×
Wabasso	Florida	Lowland	6	2.84	0.10	4×
Dacotah	North Dakota	Upland	16	2.72	0.05	4×
Summer	Nebraska	Upland	16	2.77	0.09	4×
Blackwell	Oklahoma	Upland	16	5.57	0.11	8×
Carthage	North Carolina	Upland	8	5.75	0.14	8×
Cave-in-Rock	Illinois	Upland	16	5.88	0.11	8×
Forestburg	South Dakota	Upland	17	5.63	0.05	8×
Pathfinder	Nebraska	Upland	9	5.90	0.10	8×
Shelter	West Virginia	Upland	5	5.79	0.13	8×
Sunburst	South Dakota	Upland	9	5.78	0.07	8×
Shawnee	Illinois	Upland	4	5.49	0.11	8×
Trailblazer	Nebraska	Upland	8	5.91	0.15	8×
Unknown	Wisconsin	Ambiguous ^c	8	3.59	0.46	4×, 8×

^a Ecotype previously described in the switchgrass literature

^b Exhibits phenotypic characteristics very similar to the ecotype designation

^c Exhibits characteristics difficult to classify

from the USDA-ARS breeding program in Madison, WI were included in the study for classification purposes.

Flow cytometry

Leaves from individual plants of each cultivar were prepared and analyzed separately using the CyStain PI Absolute P kit (Partec, Swedesboro, NJ, USA). Solution suspensions of nuclei were prepared in a Petri dish by finely scoring fresh leaf samples (100 mg) for 60–120 s with a sharp razor blade in 2 ml of CyStain PI Absolute P lysis buffer. Our homogenization procedure involved scoring rather than chopping since we found this method to be relatively gentle and allowed the release of nuclei while leaving the leaf tissue largely intact to avoid extra cellular debris. After ice incubation for 5 min, the solution was filtered using a 20 µm Celltrics filter (Partec, Swedesboro, NJ, USA) and then centrifuged for 5 min at 200 RCF. Afterward, the supernatant was discarded, and the cells were gently resuspended using 500 µl of CyStain PI Absolute P staining buffer and stained using 25 µl of a CyStain PI Absolute P propidium iodide and RNase solution. The stained nuclei samples were incubated for 45 min at 37°C before flow cytometric analyses and then analyzed using a FACScalibur flow cytometer (Becton–Dickinson, San Jose, CA, USA) equipped with an argon laser (488 nm).

The mean DNA content per plant cell for each sample was analyzed using ModFit software (Verity Software

House, Topsham, ME, USA). Mean DNA content per plant cell was based on a minimum of 5,000 nuclei count. Ploidy levels of the various cultivar samples were estimated by comparing the relative DNA content of cultivars to a single Alamo sample, a known tetraploid. Ploidy levels were then determined from the ratio of the control Alamo sample and unknown ploidy cultivars samples using the following equation: mean peak value for the unknown ploidy cultivars sample divided by the mean peak level for the Alamo standard multiplied by 3.0 (the mean DNA content of switchgrass reported by Hultquist et al. 1996). Samples with a mean DNA content, 2.7–3.2 pg, were deemed tetraploid and samples with 4.8–6.0 pg were classified as octaploid (Hopkins et al. 1996; Hultquist et al. 1996, 1997; Lu et al. 1998; Costich et al. 2010).

Primer selection

Initially, 2,351 SSR primer pairs developed at the USDA-ARS, Western Regional Research Center, Genomics and Gene Discovery Unit were tested (Tobias et al. 2008) using one plant from Kanlow (lowland type) and one plant from Summer (upland type). Subsequently, the best 476 primer pairs were chosen based on amplification quality and polymorphic alleles between the Summer and Kanlow plants. These primer pairs were evaluated using a small panel of diverse switchgrass genotypes (one plant each from Alamo, Dacotah, Cave-in-Rock, Timber, Wabasso, Stuart, Shelter,

and Forestburg). These genotypes were chosen to represent extreme phenotypic diversity within switchgrass, while minimizing the number of genotypes for this preliminary screen of SSR markers. We identified a total of 55 SSR loci with optimal amplification characteristics, easily scorable bands, absence of spurious peaks, and potentially useful in the differentiation of the two switchgrass ecotypes, ploidy levels, and cultivars within ecotypes (Table 2; Online Resource 1). All 55 SSR loci produced amplicons of the expected size according to Tobias et al. (2008).

DNA isolation and PCR

Total genomic DNA was isolated from approximately 0.5 cm² of leaf tissue using a DNeasy kit (QIAGEN, Valencia, CA, USA). The M13 universal primer (M13 forward, 5'-CACGACGTTGTAAAACGAC-3') was labeled either with carboxyfluorescein (FAM) or hexachlorofluorescein (HEX) fluorescent tags. Forward primers were appended at the 5'-end with the M13 sequence (i.e., CACGACGTTGTAAAACGAC) to allow indirectly labeling of reactions (Tobias et al. 2008). Reverse primers were appended with the sequence GTTCTT (PIGa) or GTTT (PIGb) at the 5'-end. These sequences were found to promote nontemplated (A) addition and facilitated subsequent genotyping (Tobias et al. 2008). Polymerase chain reactions (PCR) were performed in 8 µl total volume using 3.5 µl 1× JumpStart REDTaq ReadyMix (Sigma, St. Louis, MO, USA), 2 µl 5 ng/µl genomic DNA, 1.25 µl of H₂O, 0.5 µl 5 µM M13-FAM/HEX primer, 0.5 µl 5 µM reverse/0.5 µM forward primer, 0.125 µl 5 M betaine (Sigma, St. Louis, MO, USA), and 0.125 µl 50 mg/ml BSA (CHIMERx, Milwaukee, WI, USA). Thermocycling conditions consisted of an initial melting step (94°C for 3 min), followed by 30 cycles of 94°C for 15 s, 55°C for 90 s, and 72°C for 2 min, and a final elongation step (72°C for 20 min), followed by an indefinite soak at 4°C. PCR products (2 µl) using different fluorescent labels (i.e., FAM and HEX) were pooled and combined with 15 µl Hi-Di formamide (Applied Biosystems, Foster City, CA, USA) and 0.5 µl of carboxy-X-rhodamine (ROX) standard (GeneFlo-625 ROX; CHIMERx, Milwaukee, WI, USA). SSR allele genotyping was performed using an ABI 3730 fluorescent sequencer (POP-6 and a 50-cm array; Applied Biosystems, Foster City, CA, USA). Alleles were scored using GeneMarker Software version 1.5 (SoftGenetics, State College, PA, USA). PCRs were repeated on approximately 10% of the samples, and we obtained 99% repeatability.

Estimates of genetic parameters

All amplicon products from each SSR primer pair were considered single-locus alleles for the purpose of estimat-

ing the following genetic parameters. Genetic diversity measures, sample size (*n*), observed number of alleles (No alleles), number of private alleles (No. private), average number of alleles (Na), total number of alleles per individual (Na/individual), and polymorphism information content (PIC), were estimated for each locus and population (Table 2). PIC was calculated as follows: $PIC = 1 - \sum f_i^2$, where f_i is the frequency of the *i*th allele. We compared the genetic diversity by ecotypes, ploidy levels, and cultivars.

Population structure

We used several methods to investigate the genetic structure in our sample of cultivars. For this purpose, data obtained with the SSR loci were scored in a binary format as presence (1) or absence (0) of bands. A pairwise, individual-by-individual Euclidean distance matrix generated in GeneAEx 6.0 (Peakall and Smouse 2006) for the binary data was used to perform subsequent analyses. Principal coordinate analysis (PCoA) was performed. The degree of genetic differentiation among cultivars was estimated using Nei's genetic distance, calculated between all pairs of populations (GeneAEx). Finally, an analysis of molecular variance (AMOVA) based on 9,999 permutations was performed on all individuals, to account for marker diversity associated with ecotypes, ploidy, cultivars, and plants (GeneAEx). Additionally, the percentage of variation associated with apparent cluster groups was determined by an a posteriori AMOVA.

Bayesian clustering algorithms available in the program STRUCTURE (v. 2.2) (Pritchard et al. 2000) were used to infer whether there were genetic discontinuities in multilocus genotype data independent of the populations from which the individuals were sampled. We ran STRUCTURE using 50,000 Markov chain Monte Carlo iterations with 50,000 burn-in iterations and 10 replicates per run. We used the "admixture model" in which each individual draws a fraction of its genome from each of *K* subpopulations, and the case of "no prior population information". The most likely true value of *K* was estimated using Bayes' rule as specified in Pritchard et al. (2000) and the DK method proposed by Evanno et al. (2005).

cpDNA sequence analysis

Chloroplast DNA (cpDNA) was amplified from the following six intergenic regions: trnL(UAA), trnT(UGU)-trnL(UAA) 5', and trnL(UAA) 3'-trnF (GAA) (Taberlet et al. 1991); trnH(GUG)-psbA (Hamilton 1999); psbJ-petA and atpI-atpH ndhA (Shaw et al. 2007) (Online Resource 2). PCR were performed in 6-µl volumes containing 1× JumpStart REDTaq ReadyMix (Sigma, St. Louis, MO, USA), 0.2 µM each primer, 1 M betaine (Sigma, St. Louis, MO, USA),

Table 2 Number of alleles and polymorphism information content (PIC) for 55 SSRs used to differentiate switchgrass cultivars

Locus	No. alleles (501)	Lowland private (78)	Upland private (186)	U4× private (9)	U8× private (119)	PIC
sww112	6	2	1	0	1	0.54
sww129	3	0	1	0	1	0.50
sww150	11	0	5	2	3	0.80
sww151	16	2	8	0	8	0.82
sww153	13	2	5	0	3	0.81
sww157	6	3	1	0	1	0.57
sww162	6	1	1	0	0	0.56
sww174	9	2	4	0	3	0.67
sww175	6	0	1	0	1	0.64
sww181	5	0	2	0	1	0.54
sww185	9	0	3	0	2	0.76
sww208	6	0	2	0	1	0.68
sww210	10	0	9	1	5	0.61
sww215	4	1	1	0	1	0.28
sww274	4	1	1	0	1	0.55
sww387	3	1	0	0	0	0.40
sww391	17	3	6	0	3	0.87
sww432	12	2	2	0	2	0.86
sww438	28	11	9	1	7	0.85
sww439	11	0	2	0	1	0.87
sww457	6	0	5	0	4	0.44
sww554	8	3	1	0	1	0.43
sww583	9	0	3	0	1	0.84
sww593	7	2	3	0	1	0.59
sww607	7	1	2	0	1	0.66
sww615	12	0	7	0	5	0.82
sww645	2	0	0	0	0	0.49
sww651	8	0	3	0	1	0.71
sww664	5	1	1	0	1	0.68
sww686	9	1	6	0	2	0.73
sww687	8	3	2	0	1	0.61
sww845	5	0	2	0	1	0.45
sww857	18	2	9	0	5	0.89
sww2309	13	4	5	0	2	0.81
sww2312	13	1	3	0	2	0.78
sww2320	5	1	1	0	0	0.68
sww2341	9	3	3	0	1	0.52
sww2376	11	2	0	0	0	0.77
sww2377	9	2	3	0	2	0.55
sww2385	11	3	0	0	0	0.85
sww2387	7	2	2	0	1	0.68
sww2393	18	4	8	1	6	0.79
sww2394	9	3	3	0	1	0.72
sww2403	14	0	8	0	6	0.84
sww2415	16	0	10	1	7	0.81
sww2416	5	1	1	0	1	0.64
sww2420	6	1	3	0	3	0.59
sww2422	9	2	4	0	1	0.76

Table 2 continued

Locus	No. alleles (501)	Lowland private (78)	Upland private (186)	U4× private (9)	U8× private (119)	PIC
sww2425	9	1	4	1	3	0.64
sww2426	7	2	4	0	3	0.35
sww2431	8	1	5	0	2	0.70
sww2443	6	0	2	0	2	0.45
sww2455	6	1	0	0	0	0.78
sww2463	9	0	5	1	4	0.61
pvssr132	12	0	4	1	3	0.70
Mean	9.1	1.4	3.4	0.2	2.2	0.66

Bold indicates a subset of primers allowing the identification of ecotype and ploidy

and 10 ng template DNA. Thermocycling conditions were as follows: 80°C for 5 min; 35 cycles of 95°C for 1 min, 50°C for 1 min with a ramp of 0.3°C s⁻¹, 65°C for 5 min. PCR products were purified by adding 2 µl of 0.1 U µl⁻¹ Exonuclease I (USB Corp., Cleveland, OH, USA) and 0.1 U µl⁻¹ Shrimp Alkaline Phosphatase (USB Corp., Cleveland, OH, USA) and incubating 30 min at 37°C followed by 20 min at 80°C and 30 s at 4°C. Purified PCR products were sequenced in both directions using separate sequencing reactions. Sequencing reactions were performed using a BigDye Terminator v3.1 Cycle Sequencing Kit Applied Biosystems, Foster City, CA, USA) according to the method of Platt et al. (2007) and resolved on an ABI 3730 Genetic Analyzer Applied Biosystems, Foster City, CA, USA). Sequences were aligned and contigs were compared using Codon Code Aligner version 3.5 (CodonCode Corp., Dedham, MA, USA) using the MUSCLE algorithm (Edgar 2004). Haplotypes among sequences were identified using GeneAEx 6.0 (Peakall and Smouse 2006) and submitted to GenBank (accessions HQ110705–HQ110715).

Results

Ploidy assessment

Flow cytometry of our switchgrass samples confirmed the expected ploidy of the cultivars used in this study (Table 1). All of the lowland cultivars were classified as tetraploids, with average DNA content ranging from 2.8 to 3.2 pg. Similarly, two upland cultivars, Summer and Dacotah were classified as tetraploids, with average DNA content ranging from 2.7 to 2.8 pg. The remaining nine upland cultivars were all classified as octaploids with average DNA content ranging from 5.5 to 5.9 pg. Six of the plants of unknown origin were tetraploid with DNA content of 2.9 pg and two were octoploid with DNA content of 5.7 pg, which led to a combined of 3.6 pg (Table 1).

Genetic diversity

The 55 primer pairs used in this study produced amplification fragments corresponding to the expected lengths (Tobias et al. 2008). Overall, 501 alleles were detected with an average of 9.1 alleles and a range from 2 to 28 alleles per locus (Table 2). Out of the 55 loci, 18 produced 10–28 alleles per locus, 32 amplified 5–9 alleles, while the rest had 4 or fewer alleles. The PIC at each locus ranged from 0.28 to 0.89 with an average of 0.66 PIC overall loci.

We observed 315 alleles in the seven lowland cultivars ($n = 60$) with 78 of them being specific for the lowland ecotype. Similarly, the 11 upland cultivars plus the eight unknown samples, which were all classified as upland ($n = 132$) possessed 423 alleles with 186 being specific for the upland ecotype (Table 3). Grouping the cultivars by ploidy, we observed 226 alleles (9 private) in the 4× uplands ($n = 32$) and 410 alleles (119 private) in the 8× uplands ($n = 95$), including the eight plants of unknown origin (six 4× and two 8×). The number of alleles per individual was 84 and 85 in the 4× lowland and 4× upland cultivars, respectively, and 120 in the 8× upland. The PIC was 0.64, 0.58, 0.70 for the 4× lowland, 4× upland, and 8× upland, respectively. When cultivars were analyzed separately, PIC values ranged between 0.40 and 0.67 (Table 3). Overall, 109 cultivar-specific alleles (~22% of the total) were detected with the highest number of cultivar-specific alleles observed in Alamo and Forestburg (19 and 13, respectively) and the lowest observed in Timber, Summer, and Shawnee (1, 1, 0, respectively).

Genetic differentiation

Nei's genetic distance coefficients between switchgrass cultivars ranged from 0.02 to 0.16 with a mean genetic distance between cultivars of 0.10 (Online Resource 3). The mean Nei genetic distance between cultivars among ecotypes was 0.13 while the mean genetic distance of cultivars

Table 3 Genetic diversity characteristics of switchgrass cultivars based on 55 SSR loci

Genotype	Ecotype/ploidy	<i>N</i>	No. alleles	No. private	<i>Na</i>	<i>Na</i> /individual	PIC
All samples	–	192	501	–	9.1	100	0.70
Lowland	(L4×)	60	315	78	5.7	84	0.64
Upland	(U4×, U8×)	132	423	186	7.7	110	0.70
Upland 4×	(U4×)	37	226	9	4.1	85	0.58
Upland 8×	(U8×)	95	410	119	7.5	120	0.70
Alamo	(L4×)	16	199	19	3.6	87	0.55
Kanlow	(L4×)	16	155	8	2.8	83	0.49
Miami	(L4×)	2	111	4	2.0	82	0.45
SG5	(L4×)	9	153	3	2.8	85	0.51
Stuart	(L4×)	2	101	3	1.8	78	0.40
Timber	(L4×)	9	163	0	3.0	83	0.52
Wabasso	(L4×)	6	127	9	2.3	80	0.47
Dacotah	(U4×)	16	194	4	3.5	85	0.56
Summer	(U4×)	16	181	1	3.3	85	0.54
Blackwell	(U8×)	16	258	10	4.7	117	0.67
Carthage	(U8×)	8	234	10	4.3	117	0.67
Cave-in-Rock	(U8×)	16	244	5	4.4	116	0.66
Forestburg	(U8×)	17	224	13	4.1	118	0.65
Pathfinder	(U8×)	9	215	3	3.9	110	0.64
Shawnee	(U8×)	4	196	1	3.6	119	0.65
Shelter	(U8×)	5	171	2	3.1	105	0.60
Sunburst	(U8×)	9	219	3	4.0	119	0.64
Trailblazer	(U8×)	8	231	6	4.2	116	0.66
Unknown	(U4×, U8×)	8	185	5	3.4	98	0.61

N sample size, *No alleles* observed number of alleles, *No. private* number of private alleles, *Na* average number of alleles, *Na/individual* total number of alleles per individual, *PIC* polymorphism information content

within ecotypes was 0.07 (0.02–0.10 within the lowland ecotypes and 0.03–0.09 within the upland ecotypes). The Nei distance revealed most genetic differentiation between upland and lowland ecotypes, while cultivars within ecotypes were comparatively more similar. Likewise, the PCoA successfully discriminated all individual accessions from the 18 cultivars analyzed (Fig. 1a). Moreover, ecotypes and ploidy levels were completely separated from each other in two dimensions with 62.6% of the total genetic variation explained, (47.3 and 15.4%, respectively; Fig. 1b). In the AMOVA, SSR variation was significant ($P < 0.0001$) among ecotypes and ploidy accounting for 19% of the total variation (Table 4) and 51% (19/37) of the marker variation among cultivars. Structure analyses confirmed our findings by PCoA and AMOVA. When only two populations ($K = 2$) were assumed, we identified two genetic clusters, which corresponded to the samples divided by ecotype (Fig. 2a). Similarly, when $K = 3$ were assumed, our samples were divided into the three main groups: 4× lowland, 4× upland, and 8× upland (Fig. 2b).

The most likely true value of K identified by STRUCTURE was $K = 6$ (Fig. 2c). Clustering using different meth-

ods denoted groupings mostly consistent with the six groups. For example, both STRUCTURE ($K = 6$; Fig. 2c) and PCoA (Fig. 3a, b) analyses revealed similar patterns separating cultivars according to geographic origin: Eastern Gulf Coast and Southern Great Plains origins for lowland cultivars and Northern Great Plains, Central Great Plains, or ancient Eastern Savanna (east of the Mississippi River) for octaploid upland cultivars. When analyzed by AMOVA, the variation ($P < 0.0001$) among these six lineages accounted for 21% of the total variation (Table 4) and 60% (21/35) of the marker variation among cultivars.

cpDNA sequence results

At least two plants from each cultivar were sequenced for the previously referenced intergenic regions of the chloroplast genome. Of the six sequenced regions four were informative with at least one polymorphic locus (*trnT-trnL*, *trnL(UAA)*, *atpH-atpI*, and *psbJ-petA*). A total of 13 polymorphic sites were identified within these four intergenic regions (Table 5). This includes the previously identified 49-bp insertion in the *trnL(UAA)* intron, which has been

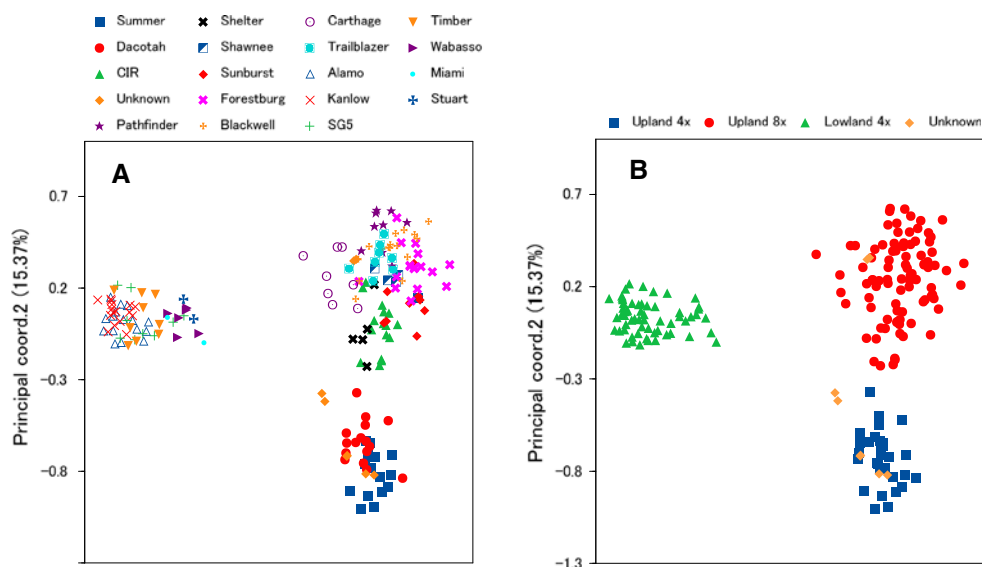


Fig. 1 Principal coordinates analyses (P. coord. 1 and 2) of switchgrass cultivars (a) based on 55 SSR loci depicting an ecotype and ploidy level differentiation (b)

Table 4 Analysis of molecular variance (AMOVA) for 192 individuals from 18 switchgrass cultivars based on 55 switchgrass SSR loci

Source of variation	df	MS	Variance component	Percentage of variation	P-value
Ecotype and ploidy	2	776.73	10.31	19	<0.0001
Cultivars/ecotype-ploidy	19	114.01	9.93	18	<0.0001
Plants/cultivar/ecotype-ploidy	170	33.46	33.46	62	<0.0001
Lineage ^a	5	453.62	10.93	21	<0.0001
Cultivar/lineage	16	90.27	7.33	14	<0.0001
Plants/cultivar/lineage	170	33.50	33.50	65	<0.0001

^a Includes ecotype, ploidy, and inferred geographic lineage based on STRUCTURE analysis with $K = 6$

suggested as diagnostic of the upland versus lowland ecotypes (Missaoui et al. 2006). From our data, the 49-bp indel was not strictly diagnostic for the switchgrass ecotypes, as it was present in two eastern Gulf Coast lowland accessions, Miami and Wabasso. Combining data from all polymorphisms identified six unique haplotypes; two of these were strictly associated with the lowland phenotype and four were associated with the upland phenotype.

Discussion

The first goal of this study was to genetically differentiate switchgrass ecotypes using molecular markers. We first described the genetic profiles of seven lowland cultivars and 11 upland cultivars using 55 SSR primer pairs and found a significant number of ecotype-specific alleles (78 lowland and 186 upland; Table 2). The data set consistently discriminated individuals from the different cultivars into their expected ecotype membership by all methods used: Nei's genetic distance (Online Resource 3), AMOVA (Table 4), PCoA (Fig. 1a), and STRUCTURE (Fig. 2a).

Our SSR data indicated that the eight samples of unknown origin were upland ecotypes (Fig. 1b). The high frequency of private alleles and separate clustering of lowland and upland populations in our study suggest they are largely reproductively isolated. Known causes of this reproductive isolation include both pre- and postfertilization incompatibilities (Martínez-Reyna and Vogel 2002) and differences in the timing of flowering due to the strong photoperiodism of the species (Martínez-Reyna and Vogel 2008).

We also investigated the utility of chloroplast markers to differentiate ecotypes (Table 5). We found that not all lowland cultivars (i.e., Wabasso and Miami; Table 5) were missing the 49 nucleotides in the *trnL* (UAA) intron which was reported by Missaoui et al. 2006 to unambiguously differentiate upland and lowland ecotypes. The cultivars Wabasso and Miami are examples of lowland germplasm from unique ecological regions not previously investigated using cpDNA sequence variation that might be genetically highly differentiated. Our results indicate that no single chloroplast polymorphism unambiguously separates upland and lowland ecotypes, but haplotypes of polymorphisms can differentiate the plants within the scope of this study.

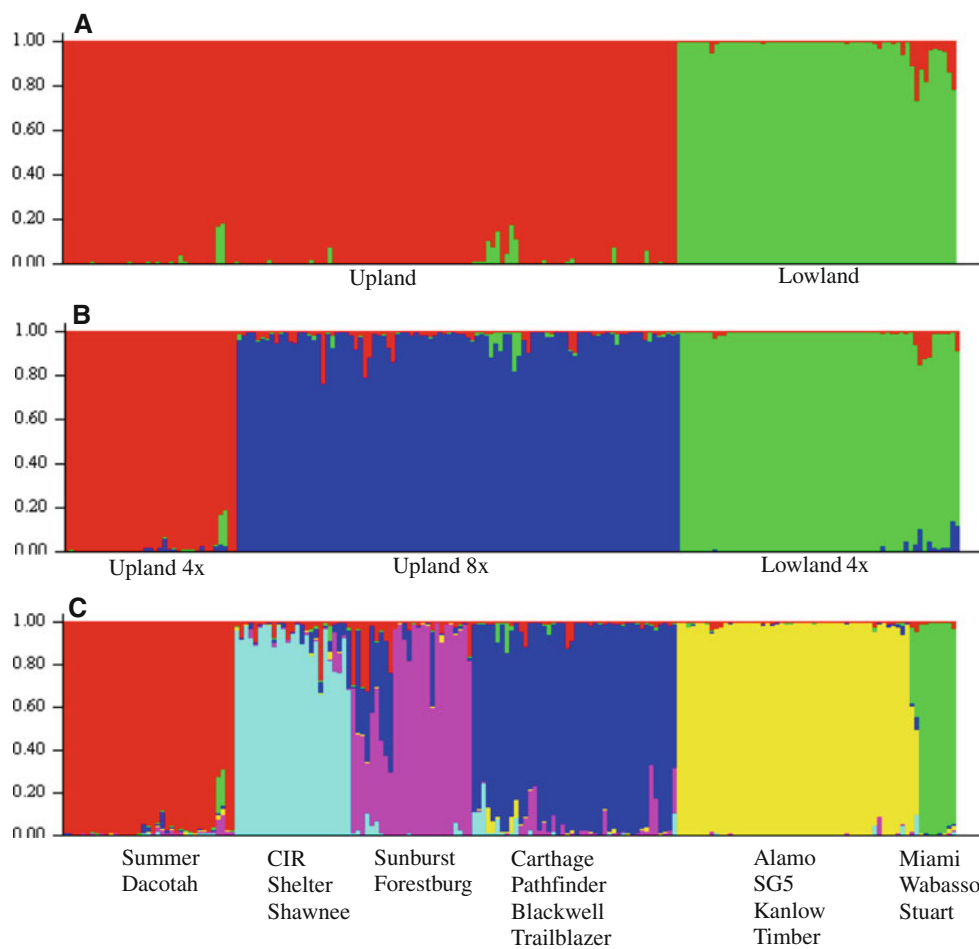


Fig. 2 STRUCTURE analyses of 18 switchgrass cultivars using 55 SSR loci. In each plot individuals are represented by a thin vertical line divided into K-colored segments that represent the individual's estimated membership fractions in each of the K clusters/populations

(labeled at the bottom). Results are presented differentiating ecotypes (a, $K = 2$), ploidy (b, $K = 3$), and regions (c, $K = 6$). The most likely true value of K identified by STRUCTURE was $K = 6$

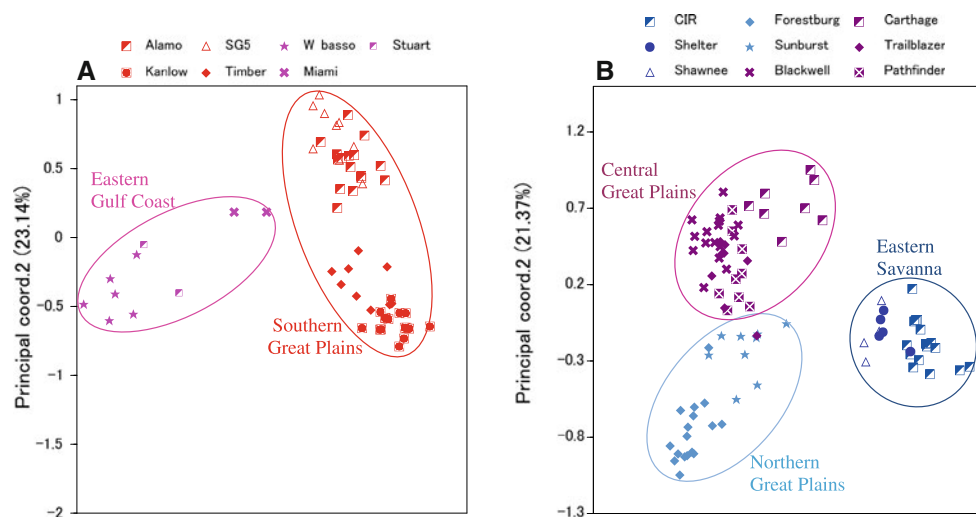


Fig. 3 Principal coordinates analyses (P. coord. 1 and 2) of switchgrass cultivars based on 55 SSR loci depicting differentiation within 4× lowland ecotypes (a) and within 8× upland ecotypes (b)

Table 5 Chloroplast DNA (cpDNA) polymorphisms from sequence analysis of 42 switchgrass plants representing 18 cultivars, representing a total of six haplotypes

Cultivar	Ecotype	<i>n</i>	cpDNA region, number of base pairs, and base-pair position number													
			trnT-trnL 854 bp			trnL (UAA) 579bp ^a	atpH-atpI 836 bp	psbJ-petA 832 bp								
			306	368	787	338–386	452	176	177	178	179	180	181	285	464	
SG5	Lowland	2	C	–	–	Absent	G	T	T	T	T	T	T	T	A	
Alamo	Lowland	2	C	–	–	Absent	G	T	T	T	T	T	T	T	A	
Kanlow	Lowland	2	C	–	–	Absent	G	T	T	T	T	T	T	T	A	
Timber	Lowland	2	C	–	–	Absent	G	T	T	T	T	T	T	T	A	
Stuart	Lowland	2	C	–	–	Absent	G	T	T	T	T	T	T	T	A	
Miami	Lowland	2	A	A	A	Present	G	T	T	T	T	T	T	G	G	
Wabasso	Lowland	5	A	A	A	Present	G	T	T	T	T	T	T	G	G	
Blackwell	Upland	2	C	A	G	Present	T	T	T	T	T	T	T	G	G	
Pathfinder	Upland	2	C	A	G	Present	T	T	T	T	T	T	T	G	G	
Trailblazer	Upland	2	C	A	G	Present	T	T	T	T	T	T	T	G	G	
Shawnee (T)	Upland	1	C	A	G	Present	T	T	T	T	T	T	T	G	G	
Shelter	Upland	2	C	A	G	Present	G	A	A	A	A	A	A	G	G	
Cave-in-Rock	Upland	2	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Shawnee (A)	Upland	1	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Dacotah	Upland	2	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Summer	Upland	2	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Forestburg	Upland	2	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Carthage	Upland	2	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Sunburst (C)	Upland	1	C	A	A	Present	G	A	A	A	A	A	A	G	G	
Sunburst (A)	Upland	4	A	A	A	Present	G	A	A	A	A	A	A	G	G	

^a 49-bp indel reported by Missaoui et al. (2006)

However, switchgrass has a low cpDNA sequence variation and its strict maternal inheritance can easily lead to discrepancies with patterns of nuclear DNA diversity when there has been recent hybridization between genetic pools. Our study has confirmed that the lowland and upland ecotypes are genetically well differentiated (Missaoui et al. 2006; Narasimhamoorthy et al. 2008; Cortese et al. 2010), but we recommend that nuclear markers must be used in conjunction with chloroplast sequences to correctly classify upland and lowland germplasm, especially when the origin is novel, unknown, or the phenotype is ambiguous.

Our second goal was to differentiate switchgrass ploidy levels using molecular markers. The ability to differentiate ploidy levels in switchgrass is extremely important to avoid incompatibilities during breeding (Martínez-Reyna and Vogel 2002) and since gene dosage effects may explain midparent heterosis observed in switchgrass (Martínez-Reyna and Vogel 2002). While we could not clearly distinguish allele dosage (i.e., duplex, triplex, etc.) due to variations in relative SSR allele peaks, we show the first case of clear genetic differentiation among switchgrass ploidy levels using molecular markers. Our aim was to provide a high-throughput method for ploidy determination as an

alternative to chromosome counts and flow cytometry since these techniques are time consuming and might be difficult to implement. Switchgrass plants were tested for ploidy by flow cytometry where octoploid samples possessed close to double value of DNA content in than tetraploid cultivars, and the results were compared to our SSR loci analysis. Plants classified by flow cytometry as tetraploid averaged ~84 alleles (range 74–95) versus ~116 alleles (range 102–129) for octaploid plants (Table 3; range data not shown). Thus, differences in allele numbers were diagnostic and consequently ploidy levels were significantly differentiated by AMOVA (Table 4). Additionally, clusters related to the number of alleles present in each individual (i.e., ploidy levels) were revealed by PCoA (Fig. 1b) and STRUCTURE (Fig. 2b). The markers also proved efficient in classifying samples of uncertain origin into ploidy levels. Six samples from our breeding program were classified as 4× upland (~89 alleles) and two were classified as 8× upland (~125 alleles), which exactly matched our results based on flow cytometry (Fig. 2b).

In switchgrass, distinct inheritance patterns occur in tetraploids (disomic) and octoploids (polysomic) reflecting differences in mode and/or timing of origin (Okada et al.

2010). However, in our study markers with apparent diploid segregation in tetraploid cultivars (i.e., single-locus disomic markers; maximum of two peaks observed per individual) were rare or likely due to the lack of representation of low frequency alleles. Our molecular marker data, though, is consistent with the reported limited genetic exchange between switchgrass ploidy levels due to pre- and postfertilization incompatibility systems that inhibit intermatings between octaploids and tetraploids (Martínez-Reyna and Vogel 2002). In fact, we detected nine private alleles for the 4× uplands, and 119 private alleles for 8× uplands (Table 2) suggesting significant genetic isolation between ploidy levels. The overabundance of private alleles related to the 8× uplands suggests a higher diversity level, independent from ploidy in the 8×, but the fact that only nine alleles are unique to the 4× upland genetic pool suggests more recent events. According to McMillian (1959), modern grasses represent more recent developments occurring since the recession of the Pleistocene Glaciation. Moreover, switchgrass colonizing prairie ecosystems during the past 11,000 years may have originated from southern refugia east of the Rocky Mountains (Casler et al. 2007a, b). Surviving switchgrass populations probably experienced strong selection pressures during both the preglacial cooling periods and post-glacial warming periods, and afterward migration north could have resulted in additional selection and drift.

Since 8× is the predominant upland type and the frequency of upland 4× populations increases with latitude, one possible explanation for the large number of 8× private alleles over 4× upland is that 8× has shown increased survivorship during past glaciation and migration events. This explanation is feasible since higher level polyploids often possess high genetic diversity and levels of heterozygosity, which may increase their ability to survive, colonize, and buffer against founder effects (Parisod and Besnard 2007). Since, the most recent switchgrass polyploidization event occurred 1 million year ago (Huang et al. 2003) and 4× may be the ancestral ploidy, the 8× genetic pool may have retained more of the ancestral alleles because of reduced genetic drift, resulting in a large number of private alleles. On the other hand, species linked to glaciations often show decreases of genetic diversity with distance from the refugia due to founder effects (Parisod and Besnard 2007). According to McMillian (1959) a likely refugium for upland switchgrass may have been western Texas; thus the large distances from this refugium of current 4× upland range would explain their low number of private alleles.

Our third goal was to identify patterns of DNA polymorphisms related to geographic origin and ecological regions. Differentiation at this level might be possible since switchgrass populations have been found to possess considerable variation in adaptation across their geographic range,

mainly related to photoperiodism and latitudinal adaptation (Casler et al. 2007b). Furthermore, several studies have reported that switchgrass collected from relatively narrow geographic regions may be genetically similar (Missaoui et al. 2006; Narasimhamoorthy et al. 2008; Cortese et al. 2010). Our SSR and chloroplast sequence data suggest that the 18 cultivars used herein can be divided into groups largely related to geographic zones. In fact, both PCoA (Fig. 1b) and STRUCTURE (Fig. 2b) grouped the cultivars into similar groups based on nuclear marker data. Such clustering revealed six groups (best *K* identified by STRUCTURE; two lowland and four upland groups) that when analyzed by AMOVA accounted for 21% of the total variation and 60% (21/35) of the marker variation among cultivars, which were higher variation percentages than the observed for ecotype and ploidy (Table 4).

The lowland cultivars from the Southern Great Plains (Alamo and Kanlow) were genetically very similar, and Timber and SG5 (unknown origin) grouped closely with these cultivars. Although, Timber reportedly originated in North Carolina, outside of the typical lowland range, it has been previously associated genetically with Kanlow (Cortese et al. 2010), indicating that Timber likely originates from the Central Great Plains or may be a selection out of Kanlow. On the other hand, a distinct cluster was formed by the Eastern Gulf Coast cultivars (Miami, Stuart, and Wabasso). Although only 10 individual plants from this region were included in our study, we detected 58 private alleles (data not shown) in this sample from this geographic region. Additionally, the chloroplast sequences yielded a unique haplotype, specific to two out of the three cultivars in this group (Table 5), supporting the idea of genetic differentiation between Gulf Coast and Great Plains lowland switchgrass genetic pools.

Similarly, we found three 8× upland switchgrass groups, mainly consistent with Northern Great Plains, Central Great Plains, and Eastern Savanna lineages, based on nuclear DNA analyses, while the 4× Northern Great Plains cultivars formed a separate cluster (Figs. 1b, 2b). For the chloroplast sequences of upland individuals, we observed four unique haplotypes, although two of the haplotypes were unique at only one base-pair and included only plants from the cultivar Sunburst and Shelter (Table 5). Thus, only two main haplotypes were observed in the upland switchgrasses, with three out of four cultivars from the Central Great Plains possessing the same haplotype (Blackwell, Trailblazer, and Pathfinder). Similarly, the Northern Great Plains (8×, Forestburg and Sunburst and 4×, Summer and Dacotah) and Eastern Savanna (Cave-in-Rock, Shelter, Shawnee) cultivars shared the same haplotype.

Results from recent studies showing switchgrass genetic groupings related to geographic origin vary considerably (Missaoui et al. 2006; Narasimhamoorthy et al. 2008;

Cortese et al. 2010). Population differentiation may occur over vast geographic distances, caused by differential climate, daylength, or habitat (Casler et al. 2007a). Switchgrass differentiation by ecological region may have been created by differential selection pressures and drift over hundreds or even thousands of years. More recently, however, agricultural activities (e.g., transport of hay containing seeds across regions) and other human disturbances involving switchgrass may have confounded some of the genetic attributes of switchgrass remnant populations making it difficult to detect genetic differences. Also, gene migration occurring via pollen or seed could have potentially diluted genetic differences related to large-scale latitude, climatic, or habitat factors. The results presented herein represent the most comprehensive study to date using both highly informative markers (i.e., 501 alleles from SSR) along with four informative chloroplast sequences providing evidence of genetic variation associated with ecological regions in switchgrass.

According to McMillian (1959), switchgrass may have survived during the Pleistocene Glaciations in three southern refugia: Upland Plains, Lowland Plains, and Eastern Gulf Coast (Online Resource 4). Our results, based on both SSR and chloroplast sequencing, are consistent with a three-refugia hypothesis. In this hypothesis, the two lowland genetic pools identified in our study, the Southern Great Plains and Eastern Gulf Coast, would be associated with McMillian's Lowland Plains and Eastern Gulf Coast refugia, respectively. Similarly, at least two upland genetic pools may have survived in a larger Upland Plains refugium. Based on chloroplast sequence data and given the strong latitudinal adaptation of switchgrass populations, we hypothesize that the colonization events since the last glacial recession occurred northbound first and then eastbound with differing switchgrass adaptations, possibly created by drift and selection during and after the glaciations events, giving rise to the observed genetic pools based on SSR data. The lack of east–west cpDNA differentiation points to more recent eastward migrations from the Great Plains to the Eastern Savanna.

The last goal of this study was to identify the potential of SSR markers as a tool to fingerprint specific switchgrass cultivars. Our selected SSR primers proved to be a powerful tool for the detection of genetic differences in switchgrass. All 192 plants in this study were fully differentiated from each other with as few as two SSR loci (sww438 and sww857). A minimum of 13 microsatellite loci were sufficient to classify 100% of the plants into their correct ecotype with all 4× upland and 92% of the 8× upland correctly classified (data not shown). However, at least 25 loci were required for classification of switchgrass lineages and to distinguish cultivars (data not shown). Genetic diversity was fairly high and consistent across cultivars

regardless of sample size, indicating that the genetic variability among or within cultivars has not been diminished by breeding and that small prairie remnants can be valuable sources of genetic diversity. Cultivars varied in marker frequencies and in the presence or absence of individual alleles (Table 3). In fact several cultivars possessed high numbers of cultivar-specific alleles, Alamo and Forestburg (19 and 13, respectively) which could readily be used for identification purposes.

A hierarchical classification approach by PCoA proved an effective method for cultivar classification. For example, when lowland and upland cultivars were analyzed together, no clear distinctions were observed at the cultivar level, but analyzing the ecotypes separately revealed differentiation of switchgrass cultivars from highly disparate origins (Fig. 3a, b). Further hierarchical classification analyses of upland cultivars allowed the clear separation of specific cultivars (e.g., Forestburg and Carthage) and revealed expected and known relationships (e.g., clustering of Shawnee and Cave-in-Rock; Vogel et al. 1996) (Online Resource 3). Similarly, samples of unknown origin could be classified such as in the case of six samples from our breeding program that grouped with Summer and Dacotah and two samples grouped with Carthage (Online Resource 3).

In conclusion, we show molecular markers could be valuable for switchgrass breeding and genetic programs as they allowed the classification of switchgrass germplasm by ecotype, ploidy, genetic pools, and cultivar. Further studies involving remnant native populations will be necessary to confirm and understand the evolutionary forces that created the switchgrass genetic pools described herein as well as the current distribution of genetic diversity over the natural range.

Acknowledgments The authors thank Nick Baker and Jered Giombi for their assistance with various aspects of this experiment, and Josh Hyman, Marie Adams, and Pam Prince for their invaluable help during SSR genotyping. This research was sponsored and funded by the US Department of Energy Great Lakes Bioenergy Research Center Grant DE-FC02-07ER64494.

References

- Casler MD (2010) Changes in mean and genetic variance during two cycles of within-family selection in switchgrass. *Bioenerg Res* 3:47–54
- Casler MD, Boe AR (2003) Cultivar × environment interactions in switchgrass. *Crop Sci* 43:2226–2233
- Casler MD, Stendal CA, Kapich L, Vogel KP (2007a) Genetic diversity, plant adaptation regions, and gene pools for switchgrass. *Crop Sci* 47:2261–2273
- Casler MD, Vogel KP, Taliaferro CM, Ehlke NJ, Berdahl JD, Brummer EC, Kallenbach RL, West CP, Mitchell RB (2007b) Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Sci* 47:2249–2260

- Casler MD, Mitchell RB, Vogel KP (2010) Switchgrass. In: Joshi S et al (eds) Handbook of bioenergy crops, vol 2. Taylor & Francis, New York
- Cortese LM, Honig J, Miller C, Bonos SA (2010) Genetic diversity of twelve switchgrass populations using molecular and morphological markers. *Bioenerg Res*. doi:10.1007/s12155-010-9078-2
- Costich DE, Friebe B, Sheehan MJ, Casler MD, Buckler ES (2010) Genome-size variation in switchgrass (*Panicum virgatum*): flow cytometry and cytology reveal rampant aneuploidy. *Plant Genome* (in press)
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:74–75
- Gunter LE, Tuskan GA, Wulfschleger SD (1996) Diversity among populations of switchgrass based on RAPD markers. *Crop Sci* 36:1017–1022
- Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol Ecol* 8:513–525
- Hopkins AA, Taliaferro CM, Murphy CD, Christian D (1996) Chromosome number and nuclear DNA content of several switchgrass populations. *Crop Sci* 36:1192–1195
- Huang S, Su X, Haselkorn R, Gornicki P (2003) Evolution of switchgrass (*Panicum virgatum* L.) based on sequences of the nuclear gene encoding plastid acetyl-CoA carboxylase. *Plant Sci* 164:43–49
- Hultquist SJ, Vogel KP, Lee DJ, Arumuganathan K, Kaeppler S (1996) Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, *Panicum virgatum* L. *Crop Sci* 36:1049–1052
- Hultquist SJ, Vogel KP, Lee DJ, Arumuganathan K, Kaeppler S (1997) DNA content and chloroplast DNA polymorphisms among switchgrasses from remnant midwestern prairies. *Crop Sci* 37:595–598
- Lu K, Kaeppler SW, Vogel K, Arumuganathan K, Lee DJ (1998) Nuclear DNA content and chromosome numbers in switchgrass. *Gt Plains Res* 8:269–280
- Martínez-Reyna JM, Vogel KP (2008) Heterosis in switchgrass: spaced plants. *Crop Sci* 48:1312–1320
- Martínez-Reyna JM, Vogel KP (2002) Incompatibility systems in switchgrass. *Crop Sci* 42:1800–1805
- McMillian C (1959) The role of ecotypic variation in the distribution of the central grassland of North America. *Ecol Monogr* 29:285–308
- Missaoui AM, Paterson AH, Bouton JH (2006) Molecular markers for the classification of switchgrass (*Panicum virgatum* L.) germplasm and to assess genetic diversity in three synthetic switchgrass populations. *Genet Resour Crop Evol* 53:1291–1302
- Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH (2008) Genetic diversity in switchgrass collections assessed by EST-SSR markers. *Bioenerg Res* 1:136–146
- Nielson EL (1944) Analysis of variation in *Panicum virgatum*. *J Agric Res* 69:327–353
- 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
- Parisod C, Besnard G (2007) Glacial in situ survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Mol Ecol* 16:2755–2767
- Peakall R, Smouse PE (2006) GenA1Ex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Platt AR, Woodhall RW, George AL (2007) Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *Biotechniques* 43:58–62
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Sanderson MA, Adler PR, Boateng AA, Casler MD, Sarath G (2007) Switchgrass as a biofuels feedstock in the USA. *Can J Plant Sci* 86:1315–1325
- Shaw J, Lickey EB, Schilling EE, Small R (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare: III. *Am J Bot* 94:275–288
- Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Mol Biol* 17:1105–1109
- Tobias C, Twigg P, Hayden D, Vogel K, Mitchell R, Lazo G, Chow EK, Sarath G (2005) Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass. *Theor Appl Genet* 111:956–964
- Tobias CM, Hayden DM, Twigg P, Sarath G (2006) Genic microsatellite markers derived from EST sequences of switchgrass (*Panicum virgatum* L.). *Mol Ecol Notes* 6:185–187
- Tobias CM, Sarath G, Twigg P, Lindquist E, Pangilinan J, Penning B, Barry K, Carpita N, Lazo GR (2008) Comparative genomics in switchgrass using 61,585 high-quality EST. *Plant Genome* 1:111–124
- Vogel KP, Mitchell RB (2008) Heterosis in switchgrass: biomass yield in swards. *Crop Sci* 48:2159–2164
- Vogel KP, Hopkins AA, Moore KJ, Johnson KD, Carlson IT (1996) Registration of ‘Shawnee’ switchgrass. *Crop Sci* 36:1713

Online Resource 1. Primer sequences for SSRs used to differentiate switchgrass cultivars.

Locus ¹	Forward primer	Reverse primer	Repeat	Accession/Sbi gene index ²
sww112*	M13-TGTTGCTCAACTTCCACTCG	PIGa-GTGTACCCCGCCATCTGC	CAG	EF057213.1
sww129*	M13-ATGCTTGTGGTTGCCTTAGC	PIGa-GAAAAGGCTTAGCGCACAGA	TTA	EF057241
sww150*	M13-TGCAGCTGAATCCCTTCTTT	PIGa-GAACTCCAGCCACCAGTTG	AAAC	EF057289.1
sww151*	M13-GAGGAGGGGAGGAACGAG	PIGa-TCGTGCTATACACCCAAGACC	TG	EF057189.1
sww153*	M13-GGGAGCCACTGGACTACTGA	PIGa-AGTCGCACTGCACCTATCGT	AG	EF057197
sww157*	M13-GACTTCAATGTCGCGCTGT	PIGa-AACTTTGATCAACCGCATCC	GCG	EF057163.1
sww162*	M13-ATGAGGCTTCCTCACCGTTC	PIGa-GGTGGACGATGAGGTCGTAA	ACT	EF057239.1
sww174*	M13-CCTTCCGTGTGCGACTCTAT	PIGa-AAATAGCAACAGGCCGGATA	GCTA	EF057281.1
sww175*	M13-AGGGCTTCATCAACAACCTGG	PIGa-GGCAGCAGCAGTGTCACTAA	ACG	EF057286.1
sww181*	M13-TGCAGCTGAATCCCTTCTTT	PIGa-GAACTCCAGCCACCAGTTG	GGC	EF057233
sww185*	M13-ATAGCGCGTTTTGCTCCTTA	PIGa-CTTCCTGCCGAGTTTAC	GCC	EF057285.1
sww208*	M13-ACCACCTACTTGCTCCGATG	PIGa-GAGGCCTCAAGAAGATCCAG	GCG	EF057255
sww210*	M13-CCAGCAAGGCCAAGTACTACC	PIGa-CGCCTACTTGAGCTGACCTG	CCA	EF057295.1
sww215*	M13-GGATGAAGATTGGCGATTCC	PIGa-GTTGCCGTTGAGGAGGATT	GT	EF057215.1
sww274*	M13-ACCTCAAGCTCGTCACTGCT	PIGa-CAGCGCATGCTCTTACAAAA	AGA	EF057288.1
sww387	M13-CGCCAAGAACTGCATCTACA	PIGa-TTTTGCAAGGAAACCAAACC	CT	FE655854/Sb01g017860.1
sww391	M13-CTGCACTGAAAGACGAAGCA	PIGa-TAATTCCTTGAAGCAACCA	ACT	FL872841/Sb01g045180.1
sww432	M13-GTTCCGTGATGTTGGTAGGG	PIGa-TTTTTGCCGAGAATTAGGTG	GCA	FE623727/Sb07g026500.1
sww438	M13-AGAAGAAGCACAAATCCGCAC	PIGa-ATCCATCCAAATTCCAAACG	GAA	DN147403
sww439	M13-TTTCCTTTCTCCTTCGGTT	PIGa-TGCTTAAATTTACATGCCGC	GT	DN147444
sww457	M13-TACAGTCTCTTCTGGCGGCT	PIGa-AGCGAGTCCCTTCATCTTCA	GCA	FL739114/Sb06g021350.1
sww554	M13-GATAGCCAGCACACCTGGTC	PIGa-CCTAGCAGCCACTACCTTGC	GGA	FE608866/Sb01g018430.1
sww583	M13-TGCTAAATAACTGGAGCGCA	PIGa-CTGATGAGCTGAAACGACCA	GC	DN151504
sww593	M13-GCTCTGTAGGCCCAATTGTC	PIGa-AGAATACCTGGACCCAACCC	TTC	FE611002
sww607	M13-CGGTTCTTGTATGGCAGGAT	PIGa-GAACCGCCTCGACCTCAC	CTG	FL731084
sww615	M13-GTTGAGTCGCAGTCCAGTGA	PIGa-ATTAGTGAGGCCACCCACAA	AG	FE645319
sww645	M13-TGCTCTGGATGTCGTCGTAG	PIGa-GATGGTGCCCAAGAAGGAG	GTT	FL874618/Sb01g003710.1
sww651	M13-TCAGCACTAGGGTTTTTGCC	PIGa-GAGCTTCTGCTGCTCGATCT	GCC	FL954766/Sb07g026160.1
sww664	M13-GGAATGAGCTCGCTCTCTTG	PIGa-TCCTCCACACACCAACCC	CG	DN147453
sww686	M13-CGACTTGGAAGTATGGTCC	PIGa-AACAATAGCAGAACAGGGCG	TA	FL895386/Sb08g013620.1
sww687	M13-CAACAGCTGATCTGTGCGAG	PIGa-CACTGCGAGACAAGGTTGC	CGC	DN150347.1
sww845	M13-GCTCCCAGTTCTTCATCTGC	PIGa-TCAGAGGGAACAGAGCGAAT	CGT	FL947519/Sb04g021380.1
sww857	M13-ATCTATCTGCATTGCCGGAG	PIGa-GGTGGCAGATCAGCTACACA	GA	FL923216/Sb04g026510.1
sww2309	M13-ATCGCTGCACCACTACTTCAC	PIGb-AGCCGTGTTTCAGGGTGAG	GGT	FL822213/Sb01g009480.1
sww2312	M13-TGGCATCTGGATGGATACAAG	PIGb-GAACACGGTCCCATATGAAT	TGC	FE636188/Sb06g018220.2
sww2320	M13-AACACTAACCCCCGATGAAAA	PIGb-GCATGTTGCTTCATCAGTA	TCA	FE618813

sww2341	M13-AAAACTCGCCTCGAACTGC	PIGb-CGTCCCGCTCTATCTGGTC	CGG	FE622269/Sb02g040560.1
sww2376	M13-CCCTCTCTTCCCTGTCAAATC	PIGb-ACCATCAATCACCATGCAACT	CGG	FE642428
sww2377	M13-AGGAGACCAGTGATGAGAGCA	PIGb-AGGCTTGTTTGTAAAGGCGAGT	GCA	FE642646/Sb01g019710.1
sww2385	M13-CTCCTCACCAAGTACCATCCA	PIGb-CATTGTAGGCTGCACATCAGA	CAG	FL883690
sww2387	M13-GAGCCCTCACAAAACCAGAG	PIGb-GACGTTCTTCAGGTGCTTC	GCC	FE645856/Sb03g040170.1
sww2393	M13-CTTCCTCATCGTCCACCAGT	PIGb-CAGAGGGGGTACTCAAGCTC	CGC	FE648168/Sb02g009470.1
sww2394	M13-CTCACCTGCTAATCCATCTGC	PIGb-CACCTTCACTGGGTAGTCCA	CTC	FL956852/Sb01g005800.1
sww2403	M13-AAGGAAATCATCGTCTGATGC	PIGb-AAGACAGCCCGTGGTTTATTC	CTC	FL889019/Sb01g035780.1
sww2415	M13-CTGAAATTAAGTCCGGGGTTC	PIGb-CGTTCAAGAGCAGGTCCTAAG	CAG	FL735165/Sb06g025150.1
sww2416	M13-GCAGTCGTTATGGATCTGACC	PIGb-CACAAAGTTCCAGACACACGA	TGG	FL743977/Sb01g002150.2
sww2420	M13-CTTGTACCCGATCCTTTTCGT	PIGb-AGGAACCGAAAGGAAACTGAA	TTG	DN149966
sww2422	M13-GGTGAAGACTGAAGAGAGCA	PIGb-ATCGCCTATGCACACAATTC	GGC	DN149594
sww2425	M13-CCTGTAGGGAGGGTAGTCTCG	PIGb-CTATCTTGGCCACCGGTATG	GGC	FL934782/Sb10g030460.1
sww2426	M13-TGTCCCAAGTACCAATGTTCC	PIGb-CAGGTGCTGACTGCTGTAGG	CAG	FL913573/Sb05g024510.1
sww2431	M13-TTTGATCAATCCTCCTCTTCG	PIGb-GGTTGGTCGACGCGATAC	GGC	DN146907
sww2443	M13-ATGAAAAACGGGCATGTGTAG	PIGb-AGAACAGGTGATAGCCCAACA	GTA	FE608191/Sb05g022230.3
sww2455	M13-CTCCGCAGTAGCAGCTTACAT	PIGb-AGTTGTAGGAGGCGTTGGTCT	AGC	FE609854/Sb02g037890.1
sww2463	M13-TGACGACGAGGACTTCGAG	PIGb-GTTGTTAGCGTTCCCTTGAT	GGC	FE617577/Sb08g002810.1
pvssr132*	M13-TGCGGCTGCTTATCTAGCTT	PIGb-CTTCTAGCTCGCCCCGTATCTAT	NA ³	

¹Genomic SSR (Lee Gunter, Personal Communication).

²The GenBank accession number used to reference the corresponding SSR sequences are listed along with the gene index number of the best significant hit to the sorghum genome.

³Information not available.

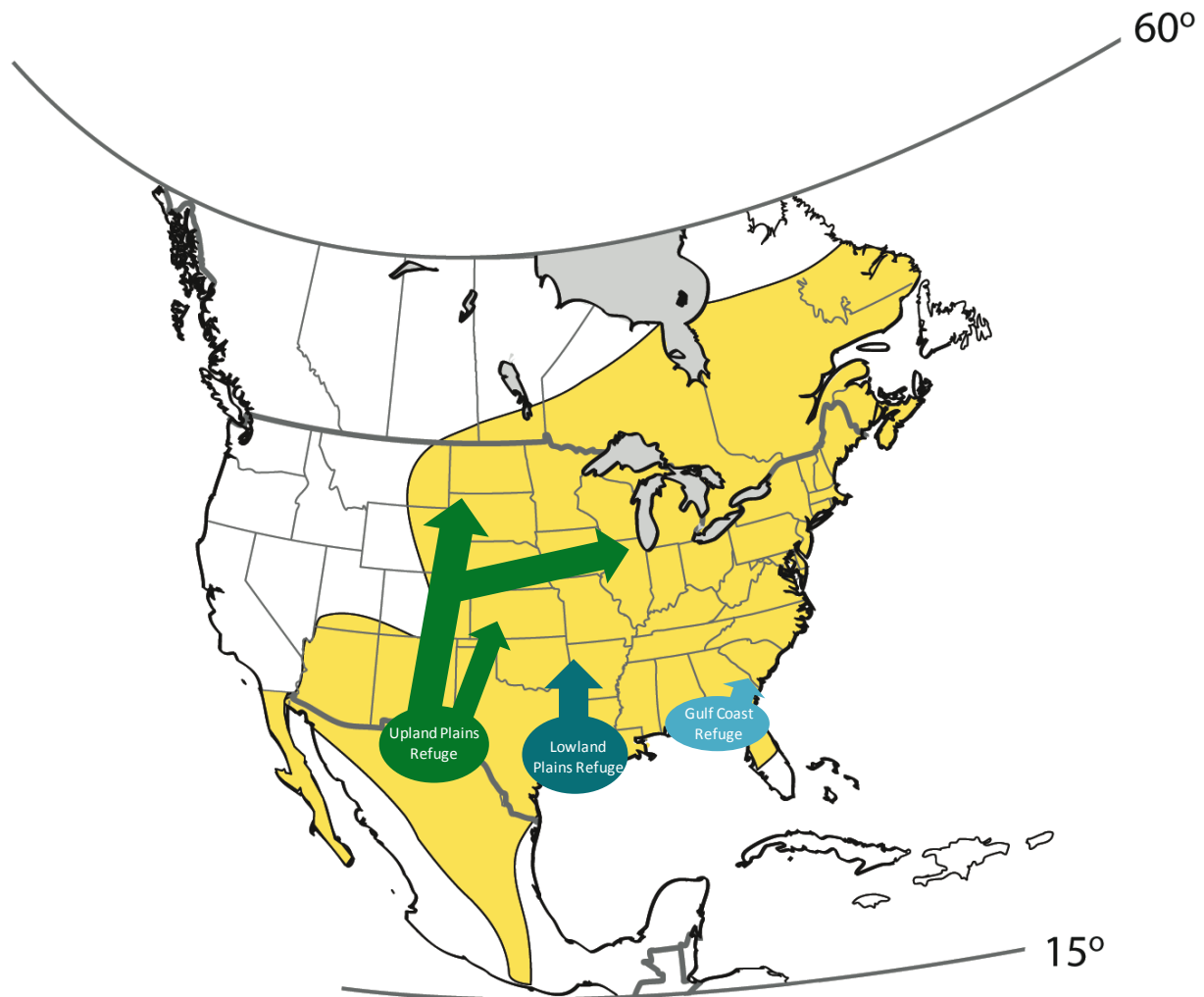
Online Resource 2. Primer sequences for five chloroplast DNA regions used to differentiate switchgrass cultivars.

Region	Forward primer	Reverse primer
trnL (UAA)	CGAAATCGGTAGACGCTACG	GGGGATAGAGGGACTTGAAC
trnT(UGU)-trnL(UAA) 5'	CATTACAAATGCGATGCTCT	TCTACCGATTTCGCCATATC
trnH(GUG)-psbA	ACTGCCTTGATCCACTTGGC	CGTAATGCTCACAACCTCCC
psbJ-petA	ATAGGTACTGTARCTGGTATT	AACAGTTTGAAAAGGTTTCARTT
atpI-atpH ndhA	TATTTACAAGCGGTATTCAAGCT	CCAATCCAGCAGCAATAAC

Online Resource 3. Pairwise Nei's genetic distances among 18 switchgrass cultivars based on 55 SSR loci.

Population	Alamo	Kanlow	Miami	SG5	Stuart	Timber	Wabasso	Dacotah	Summer	Blackwell	Carthage	Cave-in-Rock	Forestburg	Pathfinder	Shawnee	Shelter	Sunburst
Kanlow	0.03																
Miami	0.08	0.09															
SG5	0.03	0.05	0.09														
Stuart	0.08	0.10	0.10	0.10													
Timber	0.03	0.02	0.08	0.04	0.09												
Wabasso	0.09	0.09	0.10	0.09	0.05	0.08											
Dacotah	0.10	0.11	0.12	0.10	0.13	0.10	0.12										
Summer	0.11	0.12	0.13	0.11	0.14	0.10	0.13	0.04									
Blackwell	0.11	0.12	0.14	0.11	0.14	0.10	0.13	0.06	0.06								
Carthage	0.11	0.11	0.13	0.11	0.13	0.10	0.13	0.08	0.09	0.06							
Cave-in-Rock	0.12	0.12	0.14	0.13	0.14	0.11	0.15	0.07	0.07	0.07	0.08						
Forestburg	0.13	0.13	0.15	0.13	0.16	0.12	0.15	0.07	0.07	0.05	0.08	0.08					
Pathfinder	0.11	0.11	0.13	0.11	0.12	0.10	0.12	0.06	0.07	0.04	0.06	0.07	0.05				
Shawnee	0.13	0.13	0.15	0.14	0.15	0.12	0.16	0.08	0.09	0.07	0.08	0.03	0.07	0.06			
Shelter	0.13	0.13	0.15	0.13	0.13	0.12	0.14	0.08	0.08	0.08	0.09	0.05	0.09	0.08	0.07		
Sunburst	0.13	0.13	0.16	0.13	0.15	0.12	0.14	0.06	0.06	0.05	0.08	0.07	0.05	0.05	0.07	0.08	
Trailblazer	0.11	0.11	0.13	0.11	0.13	0.10	0.13	0.06	0.07	0.03	0.07	0.07	0.05	0.04	0.06	0.08	0.05

Online Resource 4. Switchgrass three-refugia hypothesis and spread after the Pleistocene glaciations: Upland Plains, Lowland Plains, and Eastern Gulf Coast. The yellow background represents the maximum distribution of switchgrass prior to European settlement of North America (Casler et al. 2010).



Online Resource 5. Hierarchical principal coordinates (PCoA 1 and 2) differentiation of upland switchgrass cultivars.

