Genetic Diversity in Switchgrass Collections Assessed by EST-SSR Markers

B. Narasimhamoorthy · M. C. Saha · T. Swaller ·
J. H. Bouton

Abstract Switchgrass (*Panicum virgatum* L.), is a warm season C_4 perennial grass, native to North American tall grass prairies. The high biomass production potential of switchgrass with low inputs makes it an excellent choice as a sustainable bioenergy crop. The objective of this study was to determine the genetic variability within and among 31 switchgrass populations obtained from Germplasm Resources Information Network (GRIN). Six plants from each population (186 genotypes) were characterized with 24 conserved grass expressed sequence tag-simple sequence repeats (EST-SSR) and 39 switchgrass EST-SSR markers. The partitioning of variance components based on the analysis of molecular variance (AMOVA) revealed that the variability within population was significantly higher (80%) than among populations (20%). Pair-wise genetic distance estimates based on SSR data revealed dissimilarity coefficients for genotypes ranging from 0.45 to 0.81. The uplands and lowlands were generally grouped in distinct sub-clusters. The genotypes were grouped into the different adaptive zones based on the geographical locations of the collections. Ploidy estimation showed that 20 of the accessions were tetraploid, four of them were octoploids and the remaining seven had mixed ploidy levels. Flow cytometry analysis of genotypes within cultivars collected from commercial seed sources did not always support the ploidy mixtures that were found in GRIN collections. Three genotypes of two accessions that clustered differently from other genotypes of the same accessions also had different ploidy levels.

Keywords Switchgrass · Biofuel · Genetic diversity · Microsatellite markers · Ploidy

Abbreviations

EST-SSR	expressed sequence tag-simple sequence repeats
AMOVA	analysis of molecular variance
PIC	polymorphism information content
PI	plant introduction

B. Narasimhamoorthy · M. C. Saha (✉) · J. H. Bouton
Forage Improvement Division,
The Samuel Roberts Noble Foundation, Inc.,
2510 Sam Noble Parkway,
Ardmore, OK 73401, USA
e-mail: mcsaha@noble.org

T. Swaller
Ceres, Inc.,
1535 Rancho Conejo Blvd.,
Thousand Oaks, CA 91320, USA

Introduction

Switchgrass (*Panicum virgatum* L.) is a warm season C_4 perennial grass [24, 30], native to North American tall grass prairies. It has long been used for conservation plantings and as a forage crop in diverse areas of the USA. The potential for sequestering large amounts of atmospheric carbon in permanent switchgrass plantings has been well documented [36]. Switchgrass has the potential to produce cellulose for biofuels such as ethanol and butanol on lands incapable of supporting traditional food crops. The sustained productivity across a wide geographical range in addition to high biomass production with minimum inputs makes it an excellent choice for a sustainable bioenergy crop.

Switchgrass is a highly heterozygous, self-incompatible and out-crossing species [41] with multiple ploidy levels ranging from $2n=2x=18$ to $2n=12x=108$ [3, 7, 33].

However, only two primary ploidy levels, tetraploid ($2n=4x=36$) or octaploid ($2n=8x=72$), are often seen which have nuclear DNA contents of 3.1 and 6.1 pg 2C-1, respectively [48]. The natural population of switchgrass is classified into two distinct ecotypes; the upland and the lowland [19]. Upland ecotypes are adapted to the mid and northern latitudes of the USA, while lowland ecotypes are adapted to the southern USA [2, 4]. Lowland ecotypes are reported to be large and more robust than their upland counterparts. In addition, these ecotypes also differ with respect to photosynthesis [49], drought tolerance [31] and N-use efficiency [35]. Hybridization between the two ecotypes usually occurs only between plants of a similar ploidy level due to post fertilization endosperm incompatibility system that inhibits seed set or produces abnormal seeds across ploidy levels [26]. Ploidy level and DNA contents of several switchgrass populations have been estimated using microscopy and flow cytometry techniques. The flow cytometry results indicated that many remnant prairie sites in the Midwestern states contain both tetraploid and octoploid plants that might be separate breeding populations [20].

Interest in developing switchgrass as a potential and sustainable bioenergy crop has intensified recently. Switchgrass has so far been bred primarily as a forage crop with a target to enhance its nutritional value [46]. However, it is in its infancy as a bioenergy production crop since some of the feedstock target traits are different from forage traits. There is a need to genetically improve switchgrass for its quality attributes to suit the rising demand for energy feedstock. Knowledge on the amount of genetic diversity and polymorphism in switchgrass is necessary to enhance the effectiveness of breeding programs and germplasm conservation efforts.

In the past, the studies on diversity in switchgrass were focused on differentiating the uplands and the lowlands using chloroplast DNA [19] or a nuclear gene coding for a plastid protein [17]. Genetic assessment of switchgrass genotypes using restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) markers and phenotypic variability among switchgrass ecotypes collected from prairie remnants all revealed that the uplands and lowlands are genetically and phenotypically distinct clusters [5, 12, 29]. These studies provide valuable information for researchers and breeders working with *P. virgatum*, but are limited to a specific region and a small number of populations. However, information regarding the amount of genetic diversity present among several switchgrass collections will be useful to select the most diverse genotypes for initial crosses in breeding programs and for molecular mapping in addition to germplasm conservation efforts.

A large number of germplasm collections of *P. virgatum* accessions are available from USDA Germplasm Resources Information Network (GRIN) and are a valuable, publicly available resource for breeding, genetics, and genomics. Understanding the diversity within the GRIN collection is important to determine the geographic/genetic representation and to identify unique genotypes that would aid in developing improved cultivars.

In this study, simple sequence repeat (SSR) markers were used to examine the level of genetic diversity and patterns of relatedness within and among the switchgrass collections obtained from GRIN collections. SSRs, derived from expressed sequence tags (EST), are more transferable between different species within a genus [9, 34] and can be used for comparative mapping, notably among the grasses [13, 37, 44, 52]. EST-SSRs analyze polymorphism that are associated with the coding regions of the genome and are capable of detecting the genetic diversity available inside or adjacent to the genes [8, 44]. The specific objectives of our study were to analyze (i) the genetic diversity existing among the 31 switchgrass populations from GRIN using EST-SSR markers (ii) to determine the similarity or differences in the ploidy level of the GRIN switchgrass collections that would relate to the clustering pattern.

Materials and Methods

Plant Materials

Number of Accessions A total of 168 switchgrass Plant Introductions (PI) are available at GRIN online database (http://www.ars-grin.gov) representing collections from 20 states within the USA and three other countries. The majority of them (127) are collections from North Dakota. For this study, we chose a total of 33 PIs which comprised of one to two PIs randomly chosen from each of 20 US states depending upon the number of different PIs collected from those states and the three international collections. Two of the PIs were discarded since one of them (PI 421901) consisted of one live plant only, while the seeds from the other (PI 430240) variety 'Shelter' belonged to a different genus, possibly *Dichanthelium clandestinum*. Thus, the total 31 switchgrass populations used in this study represent three countries and 18 US states (Table 1).

Number of Plants Per Accession Seeds from each of the 31 GRIN accessions were germinated in petri dishes and at least 18 young seedlings were transplanted to pots in the greenhouse. A pilot experiment was carried out to decide on the number of plants per accession to be included in the study. For this purpose, five accessions were randomly chosen and 18 plants from each of these five accessions (total 90 genotypes) were analyzed separately with five SSR primers. Based on the preliminary analysis, six plants

Table 1 List of *P. virgatum* accessions used to identify among and within population variations

Population	Accession and origin	Abbreviation
1	PI 476292—T 2100—Arkansas, USA	AR1
2	PI 414065—BN-14668-65—Arkansas, USA	AR2
3	PI 337553—196—Argentina	Argentina
4	PI 442535—156—Belgium	Belgium
5	PI 476294—T 4613—Colorado, USA	CO
6	PI 469228—Cave-in-Rock—Illinois, USA	IL
7	Grif 16409—Blackwell—Kansas, USA	KS1
8	PI 421521—Kanlow—Kansas, USA	KS2
9	PI 315724—BN-10860-61—Kansas, USA	KS3
10	PI 431575—KY 1625—Kentucky, USA	KY
11	PI 315728—BN-13645-64—Maryland, USA	MD1
12	PI 476291—T 2099—Maryland, USA	MD2
13	Grif 16054—Central Iowa Germplasm—Missouri, USA	MO
14	PI 315725—BN-14669-92—Mississippi, USA	MI
15	PI 315723—BN-8358-62—North Carolina, USA	NC1
16	PI 315727—BN-11357-63—North Carolina, USA	NC2
17	PI 642193—70SG 001—North Dakota, USA	ND1
18	PI 642215—70SG 024—North Dakota, USA	ND2
19	PI 591824—Shawnee—Nebraska, USA	NE1
20	PI 549094—Trailblazer—Nebraska, USA	NE2
21	PI 476293—T 2101—New Jersey, USA	NJ
22	Grif 16410—Grenville—New Mexico, USA	NM1
23	PI 642190—Falcon—New Mexico, USA	NM2
24	PI 414069—BN-309-69—New York, USA	NY
25	PI 421520—Blackwell—Oklahoma, USA	OK1
26	PI 476297—Caddo—Oklahoma, USA	OK2
27	PI 537588—Dacotah—Oregon, USA	OR
28	PI 642191—Summer—South Dakota, USA	SD1
29	PI 598136—Sunburst—South Dakota, USA	SD2
30	PI 204907— Ankara, Turkey	Turkey
31	PI 422006—Alamo—Texas, USA	TX

randomly selected from each of the 31 populations (total 186 genotypes) were finally chosen for this study.

DNA Extraction Leaf tissues were collected from young seedlings and frozen in liquid nitrogen for DNA isolation. Approximately 200 mg of leaf tissue from each of the 186 genotypes was taken in a 2.0 ml micro-centrifuge tube and ground to fine powder using a mixer mill (Retsch, Hannover, Germany). DNA was extracted from ground tissue using the DNeasy Plant Mini kit (Qiagen Inc, Valencia, CA). DNA concentration was determined using a Nanodrop spectro-photometer ND1000 (Nanodrop Technologies, DE, USA).

Primer Selection and Prescreening A total of 101 conserved grass EST-SSRs (CNL) developed at Cornell University [21] and 192 switchgrass EST-SSRs (SG EST-SSR) developed by Ceres, Inc., CA, were prescreened on eight genotypes [one genotype each from a lowland cultivar 'Alamo' (AP13) and an upland cultivar 'Summer' (VS16) and six progeny from an AP13 × VS16 cross] to select primers with clean amplification for further analysis on the 31 switchgrass populations.

PCR Reaction and Genotyping The forward primers were synthesized with an additional 18 nucleotides from the M13 universal primer appended to the 5′ end [39]. PCR reactions were prepared in a reaction volume of 10 μl which contained 20 ng of template DNA, 10× colorless GoTaq® buffer (Promega, California, USA), 0.15 mM dNTPs, 1.0 pmol of each reverse and M13 universal primer, 0.25 pmol of the forward primer, and 0.5 U GoTaq® polymerase (Promega). The M13 universal primer was labeled either with blue (FAM), green (VIC), yellow (NED), or red (PET) fluorescent tags. PCR products (3 μl) with different fragment sizes and different fluorescent labels were pooled and combined with 10 μl deionized formamide and 0.5 μl GeneScan-250LIZ internal size standard and analyzed on an ABI PRISM® 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The reactions were visualized and scored with GeneMapper 3.7 software. For each marker, individual fragments were scored as present or absent for each of the 186 genotypes.

Data Analysis Polymorphism information content (PIC), a measure of allelic diversity at a given locus, was calculated

as follows: PIC $= 1 - \sum f_i^2$, where f_i is the frequency of the ith allele [20]. All cluster analyses were performed using NTSYSpc version 2.1 (NTSYS-PC 2.10, Applied Biostatistics, Setauket, NY, USA). A binary matrix was constructed and transformed to a genetic similarity matrix using the Dice similarity coefficient [10, 32]. For clustering, the similarity coefficients were used to construct UPGMA (unweighted pair group method with arithmetic means) dendogram using the SAHN (Sequential Agglomerative Hierarchical and Nested Clustering) module of the NTSYS-pc. The FIND module was used to identify all trees that could result from different choices of tied similarity or dissimilarity values. The TREE procedure was used to create the phenogram. The trees were retained if they appeared at least 80% of the time using a bootstrap analysis performed using the program PAUP* 4.0 beta [40] with 1,000 replications. Analysis of molecular variance (AMOVA) was done using *Arlequin* 2000 [38] to partition the genetic variance within populations and between populations.

Flow Cytometry Nuclear DNA contents of the 31 switchgrass populations were estimated using flow cytometry [11] following the procedure described by Hopkins et al. [16]. Briefly, young leaves from vegetative stage of the plants were collected and immediately placed on ice and stored at 4°C. The youngest portion of the leaf, which was often still inside the sheath, was used to isolate the nuclei. Approximately 25 mg of tissue was placed in a Petri dish with about 0.4 ml of cold buffer solution (4°C) CyStain®UV Precise P and chopped to a slurry with a razor blade prior to filtrations through 35-μm nylon mesh filters. Samples were stained with 1.6 ml of staining solution and were incubated for 30–60 min at 4°C. For measurement, PI fluorescence area signals (FL2-A) from 1,000 or more nuclei per sample were collected from the flow cytometer (Partec GmbH, Germany). Leaf samples of two randomly selected genotypes from each of the 31 accessions were obtained for flow-cytometry. In addition, the individuals from each accession that were grouped under different cluster and/or sub-clusters were also analyzed separately. Genotypes were reanalyzed for confirmation if there was a variation in the DNA content among the plants within an accession or if the plant had a DNA content intermediate between the different ploidy levels that could not be categorized. In addition leaf samples from 16 seedlings each of varieties Alamo, 'Cave-in-rock' and 'Blackwell' grown from commercial seeds (Warner seeds, Browning seeds, MBS seeds) were also analyzed separately to estimate their ploidy levels.

A live gate instrument configuration was used by employing the FL2-2 and FSC parameters which allowed the fluorescence measurement from nuclei to be used to generate a histogram of FL2-A. Mean position of the G0/G1 (nuclei) peak of sample and internal standard was determined by analyzing the data by FloMax software (Partec GmbH, Germany). The mean DNA content per plant was based on the more than 1,000 scanned nuclei. The standard used for comparison was 'Stark' spring diploid barley (*Hordeum vulgare* L), which has a known DNA content (10.68 pg). The formula used for converting flourescence values to DNA content was: Nuclear DNA content = (mean position of unknown peak)/(mean position of known peak) x DNA content of known standard.

Results and Discussion

Preliminary Analysis for Sub-sampling

The five randomly selected EST-SSR primer pairs generated a total of 51 fragments and the number of fragments per primer pair ranged from 5 to 14 in the 18 individuals selected from each accession. The minimum number of plants per accession required to identify all the fragments detected by each primer pair ranged from 4 to 5 (data not shown). Thus, six plants were randomly selected from each of the 31 populations to form the final genotype pool for this study.

SSR Selection and Amplification in *P. virgatum*

A total of 63 primer pairs were selected based on clean amplification from the primer prescreening which included 39 SG EST-SSRs and 24 CNL EST-SSRs. The size of the SSR amplification fragments ranged from 126 to 560 bp. The CNL primers developed from conserved sequences of five different cereal species were previously shown to amplify in hexaploid wheat (*Triticum aestivum* L.), ryegrass (*Lolium perenne* L.), and tall fescue (*Festuca arundinacea* Schreb) [22, 37, 52]. The number of fragments generated and PIC content of the 63 primer pairs are shown in Table 2. The CNL primer pairs generated a total of 135 different size fragments. The number of unique fragments per primer pair varied from 2 to 11 with an average of 5.63. The SG EST-SSRs generated 288 different fragments and the number of fragments per primer pair ranged from 3 to 17. The SG EST-SSRs exhibited higher levels of polymorphism compared to the CNL markers in this study with an average of 7.4 fragments per primer pair. The EST-SSR markers were derived from transcribed regions of the switchgrass genome. However, some of the repeat motif found in the 5′ untranslated region of the transcript may have generated a higher level of polymorphism. The conserved grass EST-SSRs in contrast were developed from the most conserved sequences of the EST database across five different cereals, which is likely the reason for a

Table 2 Number of bands and polymorphism information content for 24 conserved grass EST-SSRs, 39 switchgrass EST-SSRs

	Conserved grass EST-SSR			Switchgrass EST-SSR		
	Marker	Number of bands	PIC	Marker	Number of bands	PIC
1	CNL035	7	0.7299	SGES002	10	0.8357
2	CNL037	11	0.7980	SGES014	7	0.6697
3	CNL039	5	0.6704	SGES047	11	0.8190
4	CNL042	4	0.5867	SGES048	6	0.5110
5	CNL047	8	0.7077	SGES052	5	0.5934
6	CNL051	4	0.5324	SGES059	10	0.8712
7	CNL053	5	0.7547	SGES060	4	0.5638
8	CNL055	10	0.6611	SGES064	4	0.2134
9	CNL061	3	0.6316	SGES065	5	0.3634
10	CNL074	5	0.6129	SGES074	17	0.8988
11	CNL086	11	0.7218	SGES076	8	0.7224
12	CNL100	3	0.6200	SGES078	7	0.6701
13	CNL101	3	0.6385	SGES082	5	0.5615
14	CNL107	4	0.3806	SGES083	10	0.8094
15	CNL110	7	0.7648	SGES092	8	0.7105
16	CNL115	5	0.6486	SGES099	13	0.8905
17	CNL119	8	0.8107	SGES102	8	0.7038
18	CNL130	2	0.0979	SGES105	8	0.6925
19	CNL133	6	0.6154	SGES127	9	0.8160
20	CNL144	7	0.6616	SGES130	7	0.6559
21	CNL147	6	0.7179	SGES131	9	0.8098
22	CNL152	3	0.2661	SGES135	6	0.4303
23	CNL156	3	0.5924	SGES147	5	0.6218
24	CNL158	5	0.7557	SGES159	13	0.8993
25				SGES160	7	0.7165
26				SGES164	3	0.3646
27				SGES166	11	0.8696
28				SGES170	8	0.7494
29				SGES175	3	0.3396
30				SGES185	9	0.8424
31				SGES234	8	0.8146
32				SGES236	7	0.7601
33				SGES239	5	0.6822
34				SGES264	8	0.7896
35				SGES296	4	0.5956
36				SGES317	5	0.7106
37				SGES352	8	0.8074
38				SGES445	5	0.6723
39				SGES456	2	0.4200
	Mean	5.63	0.624	Mean	7.4	0.68
	SD	2.57	0.17	SD	3.07	0.17

lower polymorphism pattern compared to the SG EST-SSRs. A total of 423 fragments were generated from 63 primer pairs with a mean of 6.7 fragments per primer pair, demonstrating a high level of polymorphism in switchgrass as expected for an out-crossing species [18, 23].

The PIC is the ability of the marker to differentiate the genotypes based on the number of alleles generated and their frequencies. The PIC was calculated from all 186 genotypes for all the 63 primers (Table 2). PIC values ranged from 0.09 to 0.81 for the CNL markers and 0.21 to 0.89 for the SG EST-SSR markers. Amongst the two sets of markers, the SG EST-SSRs had the highest power to differentiate the genotypes with the highest mean PIC content of 0.68, while the CNL markers had a mean PIC content of 0.62. Although the number of alleles and PIC are dependent on the specific markers selected and the diversity of germplasm, the reported PIC values in this study not only indicate the usefulness of these markers for molecular

studies, but also that these markers are highly informative to differentiate genotypes and cluster them for genetic diversity analysis.

AMOVA to Partition the Genetic Variance Among Populations

The AMOVA using accession as the grouping criterion revealed that the variance within populations accounted for 79.6% of the total variance, while among population variance contributed only 20.4% (Table 3). The greater genetic variation within rather than among populations is usually seen in out-crossing species. Greater within population variation has been reported in perennial rye-grass, meadow fescue (*Festuca pratensis*), orchardgrass (*Dactylis glomerata* L.), rhodesgrass (*Chloris gayana* Kunth) and hardinggrass (*Phalaris aquatica* L.), [18, 23, 27, 45]. Genetic differentiation between populations based on Wright's inbreeding coefficient FST (0.273) estimated from AMOVA also indicates that the greater variation exists within populations. Greater within population variation suggests that the recurrent selection from within a defined population of switchgrass is possible as reported by Taliaferro et al. [43].

Genetic Diversity of *P. virgatum* Collections

The genetic relationships among and within the accessions were determined based on the DICE similarity coefficients for the 423 fragments (Fig. 1). The phenogram indicated that a similarity of approximately 45% exists among all entries in the GRIN collections. Overall, the 186 genotypes were divided into three major clusters I, II and III (Fig. 1, Table 4). Cluster I was divided into two distinct sub-clusters, IA and IB, at approximately 55% similarity. Three genotypes from Arkansas (PI 476292-AR1) and one genotype from South Dakota (variety 'Sunburst' PI 598136-SD2) were grouped under sub-cluster IA. The sub-cluster IB had the largest number of genotypes grouped under it including the three collections from Argentina, Belgium and Turkey. This group also had all the known upland types such as PI642193-IL variety 'Cave-in-rock', PI 315724-KS3 variety Blackwell, five genotypes of PI

591824-NE1 variety 'Shawnee', PI 549094-NE2 variety Trailblazer, PI 421520-OK1 variety Blackwell, PI 476297-OK2 variety 'Caddo', PI 642191-SD1 variety 'Summer', PI 537588-OR variety 'Dacotah' and five genotypes of Sunburst-SD2 grouped under them. In addition, genotypes from PI 431575-KY, Grif 16054-MO, PI 642193-ND1, PI 642215-ND2, PI 414069-NY and five genotypes from PI 315724-KS3 were clustered in sub-cluster IB. Shawnee was a direct selection from Cave-in-rock [47] and hence was expected to cluster together. Cave-in-rock and Shaw-nee are considered southern uplands while Dacotah, Trailblazer, Summer and Sunburst are considered as northern uplands [4]. Although this sub-cluster was comprised entirely of upland ecotypes, three of the six genotypes from a known lowland ecotype variety 'Alamo' (PI 422006-TX), variety 'Grenville' (Grif 16410-NM1) and variety 'Falcon' (PI 642190-NM2) were also grouped in IB. The three Alamo genotypes that were grouped under different clusters also had a very distinct morphology differentiating them from the rest of the Alamo genotypes suggesting a possible seed mixture. There was one unique genotype from AR1 collection in cluster I that did not group with any of the genotypes within the cluster. Five out of the six genotypes from New Jersey collections (PI 476293-NJ) formed a distinct group, cluster II.

Cluster III mostly comprised of lowland ecotypes was divided into two sub-clusters IIIA and IIIB at 55% similarity. PI 414065-AR2, PI 421521-KS2 variety Kanlow, PI 315725-MI and three genotypes from Alamo-TX were all grouped under sub-cluster IIIA. Kanlow and Alamo are both lowland tetraploids, but Kanlow is considered a northern lowland ecotype, while Alamo is a southern lowland ecotype [4]. Sub-cluster IIIB was comprised of genotypes from PI 315728-MD1, PI 476291-MD2, PI 315723-NC1, five genotypes from PI 315727-NC2 and one genotype from NJ-1. One genotype from NC2 did not group with any of the clusters. Thus this genotype seems to be a unique genotype that may be genetically different from all other genotypes studied. All the genotypes in cluster III were collected from similar geographic location indicating that they all may be the lowland ecotypes.

Switchgrass populations that have been studied show considerable variation across a geographic transect ranging

Table 3 Analysis of molecular variance (AMOVA) for 186 genotypes from 31 switchgrass populations based on 24 CNL and 39 switchgrass EST-SSRs markers

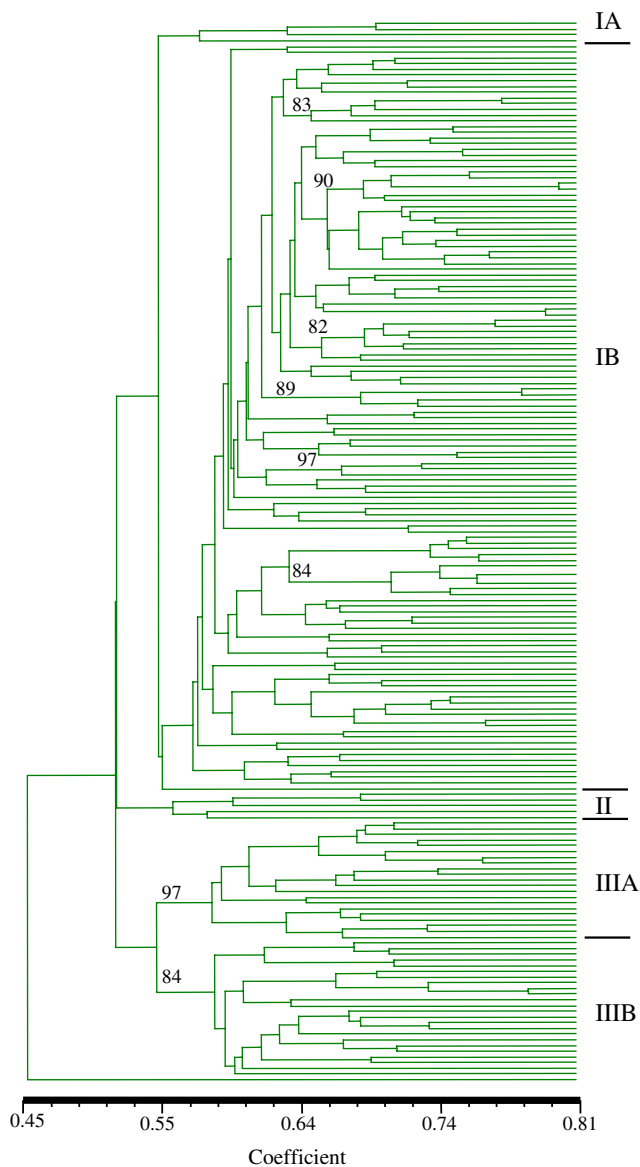| Source of variation	Sum of squares	Variance components	% variation	P values
Among populations	2,563.8	8.41	20.44	P<0.00001
Within populations	5,156.2	32.74	79.57	P<0.00001
Total	7,719.8	41.15		

Fixation index FST=0.274

Fig. 1 Cluster analysis using UPGMA of 31 *P. virgatum* accessions based on the DICE similarity coefficients estimates of 423 EST-SSR fragments. The major clusters and sub-clusters identified are indicated on the right

sub-cluster IB had most of the upland genotypes grouped together. Although, both the northern and the southern uplands were grouped together under IB, they still exhibited further differentiation within their sub-sub-cluster at 64% similarity (Fig. 1). This suggests geographical proximity and the ecotype played a significant role in determining the extent of genetic diversity. Sub-cluster IB also included the three collections from Belgium, Argentina, and Turkey for which ecotype is not known. However, three of the six genotypes from Alamo appear to be out-liers in sub-cluster IB considering the geographical proximity. Cluster analysis of the genotypes based on our marker data supports the notion that the uplands are differentiated from the lowlands in allelic content, and that clustering was mainly a function of different geographical locations of those collections.

Ploidy Estimation by Flow Cytometry

Samples from AP13 (a genotype from Alamo), VS16 (a genotype from Summer), and a genotype of Trailblazer were used as checks to estimate the ploidy levels of the accessions. AP13 and VS16 are the parents of the main public mapping population and are classified as tetraploids [28] and Trailblazer as an octoploid [16] through cytological studies. Cytogenetic analysis combined with flow cytometry has demonstrated that plants with average nuclear DNA contents 2.7 to 3.3 pg are tetraploid populations and 4.7 to 6.0 pg are octoploid populations [16, 25].

In this study the switchgrass accessions were grouped into three ploidy levels based on their DNA content (Table 5). There were 20 accessions that were classified as tetraploids, with average DNA content ranging from 2.4 to 3.4 pg, which included some of the varieties with known ploidy such as Summer, Kanlow, etc. There were four accessions classified as octoploids such as Trailblazer, Caddo, Blackwell and Argentina collections with average DNA content ranging from 4.5 to 5.8 pg. The ranges of DNA content observed for the different ploidy levels in our study were very similar to the previous reports [16]. However, we found 7 out of 31 accessions with mixed ploidies including well characterized varieties such as Cave-in-rock, Shawnee, and Alamo (Table 5). Two out of the six plants from Cave-in-rock had an average nuclear DNA content of 3.7 pg, while one plant from Shawnee had DNA content of 3.48 pg and one plant from Alamo had DNA content of 4.54 pg. Previous studies found Cave-in-rock to have lower DNA content from flow cytometry that led to categorizing it as a hexaploid [16, 51]. However, cytological evidence has revealed that Cave-in-rock and Shawnee are octoploids while Alamo is a tetraploid [16]. The original source of Alamo seeds received by GRIN was

from 36 to 46° N latitude [4]. The upland and the lowland ecotypes tend to be phenotypically and genetically distinct from each other as seen in our study. Genotypes from upland synthetic cultivar Summer grouped into a cluster distinct from two lowland synthetic cultivars, Kanlow and Alamo [29]. Similar distinct clustering among the uplands and lowlands was seen among 14 populations of switchgrass ecotypes [12]. The clustering pattern in most of the sub-clusters could be attributed to the ecotypes and the geographic locations of the accessions (Fig. 1, Table 4). For instance, sub-cluster IIIB had genotypes from eastern states, while IIIA had all the collections from southern states including the classified lowland tetraploids. Similarly the

Table 4 Genotype and accession clusters generated by similarity co-efficient method

I			II	III		
IA	IB			IIIA	IIIB	
AR1(3/6)[a]	AR1(2/6)	AR1(1/6)	NJ(5/6)	AR2	MD1	NC2(1/6)
Sunburst-SD2(1/6)	Argentina			Kanlow-KS2	MD2	
	Begium			MI	NC1	
	CO			Alamo-TX(3/6)	NC2(5/6)	
	Cave-in-rock-IL				NJ(1/6)	
	Blackwell-KS1					
	KS3					
	KY					
	MO					
	ND1					
	ND2					
	Shawnee-NE1					
	Trailblazer-NE2					
	Greenville-NM1					
	Falcon-NM2					
	NY					
	Blackwell-OK1					
	Caddo-OK2					
	Dacotah-OR					
	Summer-SD1					
	Sunbrust-SD2(5/6)					
	Turkey					
	Alamo-TX(3/6)					

[a] Mixed ploidy levels

collected in Texas in 1977 and was not certified [1]. This suggests that the mixed ploidy levels we observed in Alamo may have been due to a seed mixture as Alamo had never been a certified cultivar. Cave-in-rock and Shawnee were received by GRIN in 1967 and 1997 [1]. The co-existence of hexaploid and octoploid in these accessions may be due to outcrossing with another accession of different ploidy level. However, very low crossability of 0.06% [42] and postfertilization incompatibility [26] in interploidy crosses suggest that gene flow could not happen between ploidy levels.

The other four PIs with mixed ploidies were PI476292 and PI642193 each with one octoploid genotype, and PI476294 and PI315724 each with one hexaploid genotype. All the plants that were classified as hexaploids in this study through flow cytometry revealed similar levels of DNA content with biological replications. Hultquist et al. [20] reported that switchgrass populations found within Midwestern prairies had a mixture of ploidy levels, suggesting the need to check these germplasm for DNA content before using it in the breeding program. In our study three out of six plants from variety Blackwell collected from Oklahoma were classified as tetraploid while the same variety Blackwell collected from Kansas was classified as octoploid. Blackwell is reported to be an octoploid southern upland variety [4, 16]. The Blackwell collection from Oklahoma was not certified

seed and was received by GRIN during 1978, again suggesting the possibility of seed mixture.

In order to examine the concerns of mixed ploidy levels among some of the cultivars, 16 seedlings, each developed from commercial seed lots of three varieties Alamo, Cave-in-rock and Blackwell were analyzed separately for the DNA content to estimate the ploidy level. All the seedlings from each of the commercial seed lots of Alamo and Blackwell were found to be tetraploids (mean DNA content of 2.97 pg) and octoploids (mean DNA content of 4.76 pg), respectively. However, in Cave-in-rock, one seedling each was designated as tetraploid and hexaploid with a mean DNA content of 2.96 and 3.68 pg, respectively and the rest 14 of them were octoploids with a mean DNA content of 4.75 pg. Out-crossing between the Cave-in-rock and other accessions of different ploidy levels may have led to polyploid levels within the population. However, the occurrence of a tetraploid plant in Cave-in-rock can best be explained if it existed from its inception as a population of mixed ploidy.

Switchgrass Diversity and Relationship to Ploidy Level

The different clustering pattern seen among the accessions and genotypes within accession (Fig. 1 Table 4) may not

Table 5 DNA content and estimated ploidy level of 31 switchgrass collections

	Accession	Ploidy	No. of plants/accession	Avg. DNA content in pg
1	T 2100—Arkansas	$4x/8x$[a]	$4/2$[b]	$3.18/4.92$[c]
2	BN-14668-65—Arkansas	$4x$	2	2.97
3	198—Argentina	$8x$	2	4.42
4	157—Belgium	$4x$	2	2.51
5	T 4613—Colorado	$4x/6x$	5/1	3.42/4.10
6	Cave-in-Rock—Illinois	$8x/6x$	4/2	5.77/3.72
7	Blackwell—Kansas	$8x$	2	4.87
8	Kanlow—Kansas	$4x$	2	2.74
9	BN-10860-61—Kansas	$4x/6x$	4/1	3.14/3.82
10	KY 1625—Kentucky	$4x$	2	3.04
11	BN-13645-64—Maryland	$4x$	2	2.84
12	T 2099—Maryland	$4x$	3	2.61
13	Central Iowa Germplasm—Missouri	$4x$	2	2.74
14	BN-14669-92—Mississippi	$4x$	2	3.12
15	BN-8358-62—North Carolina	$4x$	2	2.87
16	BN-11357-63—North Carolina	$4x$	2	2.68
17	70SG 001—North Dakota	$4x/8x$	5/1	2.73/4.68
18	70SG 024—North Dakota	$4x$	2	3.20
19	Shawnee—Nebraska	$4x/8x$	1/5	3.42/4.70
20	Trailblazer—Nebraska	$8x$	2	5.10
21	T 2101—New Jersey	$4x$	2	2.46
22	Grenville—New Mexico	$4x$	2	2.80
23	Falcon—New Mexico	$4x$	2	3.42
24	BN-309-69—New York	$4x$	2	3.26
25	Blackwell—Oklahoma	$4x$	3	2.84
26	Caddo—Oklahoma	$8x$	2	4.96
27	Dacotah—Oregon	$4x$	2	3.14
28	Summer—South Dakota	$4x$	2	3.28
29	Sunburst—South Dakota	$4x$	2	3.36
30	Ankara, Turkey	$4x$	2	3.09
31	Alamo—Texas	$4x/8x$	5/1	2.62/4.54
	AP13	$4x$	1	2.54
	VS16	$4x$	1	2.90
	Alamo commercial seeds	$4x$	16	3.02
	Cave-in-rock commercial seeds	$4x/6x/8x$	1/1/14	2.46/3.78/4.68
	Blackwell commercial seeds	$8x$	16	4.76

[a] Mixed ploidy levels
[b] Number of plants with corresponding ploidy levels
[c] Corresponding DNA content

entirely be attributed to the differences in ploidy level (Table 5). All the octoploids were grouped under the sub-cluster IB while all the genotypes grouped under clusters II and III were tetraploids. Genotypes from the same accession that fell into different sub-clusters were analyzed to see if they would relate to the differences in ploidy level. Among the five plants of AR1 grouped in IA and IB, only two of them clustered in IB had a different ploidy status. The Alamo plants have two distinct phenotypes and grouped under two major clusters, I and III (Table 4). Only one of the three Alamo plants in cluster IB had been estimated to be an octoploid with 4.54 pg DNA content (Table 5). One individual genotype from Shawnee-NE1 had different ploidy status compared with other individuals within the

population. The unique genotype from AR1 in Cluster I that did not group with any of the sub-clusters and the unique genotype from NC2 that did not group with any of the major clusters did not differ from the other plants within the accession for ploidy level (Fig. 1; Tables 4 and 5). The genotype from NJ that was grouped under a different cluster or sub-cluster did not differ from the other plants within the accession for its ploidy level. On the other hand, the six plants from Cave-in-rock-IL grouped in IB were estimated to have mixed ploidy levels (Fig. 1, Tables 4 and 5). Only three genotypes (two from AR1 and one from Alamo) were found to differ from other genotypes within the same accession for both clustering pattern and ploidy levels.

Switchgrass is a highly adaptable species where day length and tolerance to cold and heat controls the adaptation zone of the switchgrass population. Most of the populations cannot be moved north or south more than one hardiness zone without adversely affecting flowering, vigor and survival rate [4]. Similarly there is little or no gene flow across ploidy levels due to reproductive isolation [26]. Despite the adaptation and incompatibility across ploidy levels, pollen or seed migration could create a shift in the population resulting in mixed ploidy levels amongst certain collections. In this study there is no clear separation of upland accessions based on their ploidy levels. Missaoui et al. [29] found similar results and suggested that these accessions may have descended from the same maternal origin since chloroplast inheritance in switchgrass was shown to be maternal [26].

Conclusions

In conclusion, molecular marker analysis suggested a great amount of genetic variability exists among these switchgrass collections. Variation within population was much higher compared to the variation among populations. Cluster analysis distinctly grouped the uplands and lowlands into separate clusters. The genotypes were generally sub-clustered into the different adaptive zones based on the geographical locations of the collections. Ploidy estimation on the GRIN collections and confirmation with commercial seeds revealed the possibility of seed mixture in GRIN collections. Only in a few instances differences in ploidy level within accessions attributed to the differences in clustering. Therefore, the GRIN switchgrass collection is a very valuable resource for use by scientists wishing to study species diversity or for breeding for improved cultivars.

References

1. Alderson J, Sharp WC (1995) Grass varieties in the United States. CRC Press Inc, Florida, pp 194–199
2. Brunken JN, Estes JR (1975) Cytological and morphological variation in *Panicum virgatum* L. Southwest Nat 19:379–385
3. Burton GW (1942) A cytological study of some species in the tribe Paniceae. Am J Bot 29:355–359
4. Casler MD, Vogel KP, Taliaferro CM, Wynia RL (2004) Latitudinal adaptation of switchgrass populations. Crop Sci 44:293–303
5. Casler MD (2005) Ecotypic Variation among Switchgrass Populations from the Northern USA. Crop Sci 45:388–398
6. Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722
7. Church GL (1940) Cytotaxonomic studies in the gramineae *Spartina, Andropogon*, and *Panicum*. Am J Bot 27:263–271
8. Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping A and B genomes of wheat. Theor Appl Genet 104:399–407
9. Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for Medicago spp. Theor Appl Genet 108:414–422
10. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302
11. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E (1983) Rapid flow cytophotometric analysis of the cell cycle in intact plant tissues. Science 220:1049–1051
12. Gunter LE, Tuskan GA, Wullschleger SD (1996) Diversity among populations of switchgrass based on RAPD markers. Crop Sci 36:1017–1022
13. Gupta P, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323
14. Gustafsson S, Lonn M (2003) Genetic differentiation and habitat preference of flowering-time variants within *Gymnadenia conopsea*. Heredity 91:284–292
15. Hammerli A, Reusch TBH (2003) Inbreeding depression influences genet size distribution in a marine angiosperm. Mol Ecol 12:619–629
16. Hopkins AA, Taliaferro CM, Murphy CD, Christian D (1996) Chromosome number and nuclear DNA content of several switchgrass populations. Crop Sci 36:1192–1195
17. 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
18. Huff DR (1997) RAPD characterization of heterogeneous perennial ryegrass cultivars. Crop Sci 37:557–564
19. 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
20. Hultquist SJ, Vogel KP, Lee DJ, Arumuganathan K, Kaeppler S (1997) DNA content and chloroplast DNA polymorphisms among accessions of switchgrass from remnant mid-western prairies. Crop Sci 37:595–598
21. Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510
22. Kirigwi FM, Zwonitzer JC, Mian MAR, Wang Z-Y, Saha MC (2007) Microsatellite markers and genetic diversity assessment in *Lolium temulentum*. Genet Resources Crop Evol 55:105–114
23. Kolliker R, Stadelmann FJ, Reidy B, Nosberger J (1998) Genetic variability of forage grass cultivars: A comparison of *Festuca pratensis* Huds, *Lolium perenne* L. and *Dactylis glomerata* L. Euphytica 106:261–270
24. Koshi PT, Stubbendieck J, Eck HV, McCully WG (1982) Switchgrass: forage yield, forage quality, and water use efficiency. J Range Mgt 35:623–627

25. Lu K, Kaeppler SM, Vogel KP, Arumuganathan K, Lee DJ (1998) Nuclear DNA content and chromosome numbers in switchgrass. Great Plains Res 8:269–280

26. Martinez-Reyna JM, Vogel KP, Caha C, Lee DJ (2001) Meiotic stability, chloroplast DNA polymorphisms, and morphological traits of Upland Lowland switchgrass reciprocal hybrids. Crop Sci 41:1579–1583

27. Mian MAR, Zwonitzer JC, Chen Y, Saha MC, Hopkins AA (2005) AFLP diversity within and among Hardinggrass populations. Crop Sci 45:2591–2597

28. Missaoui AM, Paterson AH, Bouton JH (2005) Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. Theor Appl Genet 110:1372–1383

29. 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

30. Moss DN, Krenzer EG, Brun WA (1969) Carbon dioxide compensation points in related plant species. Science 164:187–188

31. Nickell GL (1972) The physiological ecology of upland and lowland *Panicum virgatum*. Ph.D. dissertation (Diss. Abstr. No. 73-04957). Univ. of Oklahoma, Norman

32. Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269–5273

33. Nielsen EL (1944) Analysis of variation in *Panicum virgatum.* J Agric Res 69:327–353

34. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

35. Porter CL (1966) An analysis of variation between upland and lowland switchgrass *Panicum virgatum* L. in central Oklahoma. Ecology 47:980–992

36. Sanderson MA, Reed RL, McLaughlin SB, Wullschleger SD, Conger BV, Parrish DJ, Wolf DD, Taliaferro C, Hopkins AA, Ocumpaugh WR, Hussey MA, Read JC, Tischler CR (1996) Switchgrass as a sustainable bioenergy crop. Bioresour Technol 56:83–93

37. Saha MC, Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

38. Schneider S, Kueffer JM, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: A software for population genetic data analysis. Genetics and Biometry Laboratory. University of Geneva, Switzerland

39. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragment. A poor man's approach to genotyping for research and high throughput diagnostics. Nat Biotechnol 18:223–33

40. Swofford DL (2000) PAUP* 4.0 Beta Version: Phylogenetic analysis using Parsimony and other methods (software). Sinauer, Sunderland, MA

41. Talbert LE, Timothy DH, Burns JC, Rawlings JO, Moll RH (1983) Estimates of genetic parameters in switchgrass. Crop Sci 23:725–728

42. Taliaferro CM, Hopkins AA (1996) Breeding characteristics and improvement potential of switchgrass. In: Cundiff JS et al (ed) Proceedings of the Third Liquid Fuel Conference, Nashville, TN. 15–17 Sept. 1996. ASAE, St. Joseph, MI, pp 2–9

43. Taliaferro CM, Vogel KP, Bouton JH, McLaughlin SB, Tuskan GA (1999) Reproductive characteristics and breeding improvement potential of switchgrass. In: Overend RP, Chornet E (eds) Biomass—a growth opportunity in green energy and value added products. Elsevier, pp147–153

44. Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). Theor Appl Genet 106:411–422

45. Ubi EB, Kolliker R, Fujimori M, Komatsu T (2003) Genetic diversity in diploid cultivars of rhodesgrass determined on the basis of amplified fragment length polymorphism markers. Crop Sci 43:1516–1522

46. Vogel KP, Gorz HJ, Haskins FA (1989) Breeding grasses for the future. Crop Science Society of America. Contributions from Breeding Forage and Turf Grasses, CSSA Special Pub No. 15

47. Vogel KP, Hopkins AA, Moore KJ, Johnson KD, Carlson IT (1996) Registration of 'Shawnee' switchgrass. Crop Sci 36:1713

48. Vogel KP (2006) Genetic improvement of switchgrass as a biomass energy crop. Abstracts Third Annual World Congress on Industrial Biotechnology and Bioprocessing, July 11–14, 2006, Toronto, Canada, p 62

49. Warner DA, Ku MSB, Edwards GE (1987) Photosynthesis, Leaf Anatomy, and Cellular Constituents in the Polyploid C4 Grass *Panicum virgatum*. Plant Physiol 84:461–466

50. Weir B (1990) Genetic data analysis: methods for discrete population genetic data. Sinauer, Sunderland, MA

51. Wullschleger SD, Sanderson MA, McLaughlin SB, Biradar DP, Rayburn AL (1996) Photosynthetic rates and ploidy levels among populations of switchgrass. Crop Sci 36:306–312

52. Yu J-K, La Rota M, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751