Theor Appl Genet (2011) 122:677–686
DOI 10.1007/s00122-010-1477-4

ORIGINAL PAPER

Development of 1,030 genomic SSR markers in switchgrass

Y. W. Wang · T. D. Samuels · Y. Q. Wu

Received: 24 March 2010 / Accepted: 30 September 2010 / Published online: 27 October 2010

Abstract Switchgrass, *Panicum virgatum* L., a native to the tall grass prairies in North America, has been grown for soil conservation and herbage production in the USA and recently widely recognized as a promising dedicated cellulosic bioenergy crop. A large amount of codominant molecular markers including simple sequence repeats (SSRs) are required for the construction of linkage maps and implementation of molecular breeding strategies to develop superior switchgrass cultivars. The objectives of this study were (1) to identify SSR-containing clones and to design PCR primer pairs (PPs) in SSR-enriched genomic libraries, and (2) to validate and characterize the designed SSR PPs. Five genomic SSR enriched libraries were constructed using genomic DNA of 'SL93 7 × 15', a switchgrass genotype selected in an Oklahoma State University (OSU) southern lowland breeding population. A total of 3,046 clones from four libraries enriched in (CA/TG)n, (GA/TC)n, (CAG/CTG)n and (AAG/CTT)n SSR repeats were sequenced at the OSU Core Facility. From the sequences, we isolated 1,300 unique SSR-containing clones, from which we designed 1,398 PPs using SSR Locator V.1 software. Among the designed PPs, 1,030 (73.7%) amplified reproducible and strong bands with expected fragment size, and 802 detected polymorphic alleles, in SL93 7 × 15 and 'NL94 16 × 13', two parents of one mapping population. All of the four libraries contained a high rate of perfect SSR repeat types, ranging from 62.7 to 76.2%. Polymorphism of the effective SSR markers was also tested in two lowland and two upland switchgrass cultivars, encompassing 'Alamo' and 'Kanlow', and 'Blackwell' and 'Dacotah', respectively. The developed SSR markers should be useful in genetic and breeding research in switchgrass.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1477-4) contains supplementary material, which is available to authorized users.

Y. W. Wang · T. D. Samuels · Y. Q. Wu (✉)
Department of Plant and Soil Sciences,
Oklahoma State University, Stillwater, OK 74078, USA
e-mail: yanqi.wu@okstate.edu

Y. W. Wang
Department of Grassland Science,
China Agricultural University, Beijing 100193, China

Introduction

Switchgrass (*Panicum virgatum* L.) is an economically important member of the tribe Paniceae in the subfamily Panicoideae of the family Poaceae. It is a warm season, C_4 perennial grass native to the tall grass prairies of North America, and naturally distributed in non-forested areas from 15°N latitude in central Mexico across a large portion of the USA to 55°N latitude in Canada except for the areas west of the Rocky Mountains (Hitchcock 1951; Stubbendieck et al. 1982). The grass is cross-pollinated, self-incompatible, and highly polymorphic (Vogel 2004). Two switchgrass ecotypes, lowland and upland, are identified on the basis of morphological characteristics and habitat preferences (Porter 1966). Lowland switchgrass plants are taller and coarser than upland plants while upland switchgrass plants have longer rhizomes resulting in a more spreading growth habit. Four subecotypes, northern upland, southern upland, northern lowland and southern lowland were further recognized based on latitudinal adaptation (Casler et al. 2004). Switchgrass has a basic chromosome number of $x = 9$ (Gould 1975). Early cytological studies reported the

species comprises a polyploidy series with reported somatic chromosome number being of 18, 36, 54, 72, 90, and 108 (Nielsen 1944; Barnett and Carver 1967). Recent flow cytometry investigations indicated all lowland switchgrass is tetraploid ($2n = 4x = 36$), whereas majority of upland switchgrass is octaploid ($2n = 4x = 72$) and the remaining includes tetraploid (Hopkins et al. 1996; Lu et al. 1998).

Switchgrass has been widely used for soil conservation and as a warm-season pasture grass in the USA (Vogel 2004). More recently, the grass species has been identified as a promising dedicated cellulosic bioenergy crop due to its high yielding potential, adaptability to marginal land, perennial nature, ease of stand establishment using seed, desirability to soil conservation and wildlife preservation, flexibility for multiple uses, and its compatibility to existing farming systems (McLaughlin 1992; McLaughlin and Kszos 2005). Current breeding objectives of developing switchgrass as a biofuel crop are to increase genetic potential of biomass yield and enhance traits related to conversion efficiency to ethanol or butanol (Bouton 2008). It takes about 15 years to develop a new switchgrass cultivar by conventional breeding procedures. The development of genomic resources including molecular markers should enable molecular breeding for improving switchgrass as a forage and bioenergy crop (Tobias et al. 2008).

Molecular markers have great potential to help breeders by linking genotypic and phenotypic variations, and speed up the process of developing improved cultivars (Gupta and Varshney 2000). DNA simple sequence repeats (SSRs), also known as microsatellites, are tandem repeats of 2–6 bp DNA core sequences, which are highly variable in the number of repeats at a specific locus. SSRs are widely distributed throughout the eukaryotic genomes in both protein-coding and non-coding regions (Morgante et al. 2002). They have the advantages of being PCR-based, reliable, co-dominant, multi-allelic, chromosome specific, and highly informative genetic markers. They are amenable to high-throughput genotyping, thus suitable for construction of high-density linkage maps, QTL mapping, gene cloning, germplasm diversity analysis, cultivar identification, and marker-assisted selection. SSR markers can be developed from either genomic or expressed sequence tag (EST) libraries. The development of EST-based SSRs needs less time and effort, and is cost effective compared to genomic SSR development if EST sequences are available. Rapidly growing EST databases are an important source for developing EST-derived SSR markers for switchgrass. Tobias and colleagues reported the primer sequences of 32 effective SSR markers developed from a switchgrass EST project (Tobias et al. 2005, 2006). Later, Tobias et al. (2008) expanded the effort with developing additional 830 EST-derived SSR markers. Recently, both conserved grass EST SSRs and switchgrass EST SSRs demonstrated as useful markers in genetic diversity analysis among switchgrass populations (Narasimhamoorthy et al. 2008). It appears the EST SSRs transferable among the grass species are more conservative in generating polymorphic alleles than those derived from switchgrass EST sequences (Narasimhamoorthy et al. 2008). Tobias et al. (2008) observed the mean number of reliable fragments per individual plant amplified using 830 switchgrass EST-SSR primer pairs (PPs) was 2.18 although a few PPs produced 5–8 amplicons.

Although EST-SSRs are useful for genetic analysis in switchgrass, their disadvantages of relatively low polymorphism and high concentration in gene-rich regions of the genome may limit their usage, especially for the construction of linkage maps (Cho et al. 2000; La Rota et al. 2005; Chabane et al. 2005; Pinto et al. 2006). In contrast to EST-SSRs, prior investigations indicate genomic SSRs are highly polymorphic and tend to be widely distributed throughout the genome resulting in better map coverage (Taramino et al. 1997; Warnke et al. 2004; La Rota et al. 2005; Saha et al. 2005). Advances in techniques for the construction of microsatellite-enriched libraries and DNA sequencing in recent years have resulted in the accelerated development of large numbers of genomic SSR markers in annual ryegrass (*Lolium multiflorum*) (Hirata et al. 2006), perennial ryegrass (*L. perenne*) (Gill et al. 2006), tall fescue (*Festuca arundinacea*) (Saha et al. 2006), timothy (*Phleum pratense*) (Cai et al. 2003), and zoysiagrass (*Zoysia japonica*) (Cai et al. 2005). To date, few switchgrass genomic SSRs are available in public domain. The objectives of this study were (1) to identify SSR-containing clones and design PCR primer pairs (PPs) from sequencing SSR-enriched genomic libraries, and (2) to validate and evaluate the designed SSR PPs.

Materials and methods

Plant materials and genomic DNA isolation

Two switchgrass genotypes, 'SL93 7 × 15' and 'NL94 16 × 13', and four cultivars, 'Alamo', 'Blackwell', 'Dacotah' and 'Kanlow' (Alderson and Sharp 1994), were used in the study. SL93 7 × 15 was selected from a southern lowland breeding population initiated in 1993 at OSU, while NL94 16 × 13 was a clone selected from a northern lowland breeding population initiated in 1994. SL93 7 × 15 and NL94 16 × 13 were parents of a mapping population. Alamo represented the southern lowland ecotype and Kanlow the northern lowland ecotype while Blackwell and Dacotah were selected to stand for the southern upland and northern upland ecotypes, respectively. Alamo, Kanlow, SL93 7 × 15, NL94 16 × 13 and Dacotah are tetraploids, while Blackwell is octaploid.

Individual plants of the two genotypes and four cultivars were maintained in separate 4-inch pots filled with SUN-GRO Metro-Mix 200 series soil (Sun Gro Horticulture, WA) in an OSU greenhouse. Genomic DNA was extracted from the healthy leaf tissues using Genomic DNA Maxi Kit (Plant) (IBI Scientific, IA, USA) following the manufacturer's instructions. Since switchgrass is largely cross-pollinated and self-incompatible, plants in switchgrass cultivars are presumably highly heterozygous. Therefore, genomic DNA of the individual cultivars was extracted using one sample made by bulking equal amount of leaf tissues from 20 randomly selected plants. DNA quantity and quality were measured by a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and 1.0% agarose gel electrophoresis, respectively.

Construction and sequence analysis of SSR-enriched genomic libraries

DNA of SL93 7 × 15 was used to construct all SSR-enriched genomic libraries. Five SSR-enriched genomic libraries (CA, GA, ACC, CAG and AAG) were constructed by Genetic Identification Services (GIS, Chatsworth, CA) (Jones et al. 2002). In brief, 1 μg of genomic DNA was partially restricted with a mix of seven blunt-end cutting enzymes (*Rsa*I, *Hae*III, *Bsr*B1, *Pvu*II, *Stu*I, *Sca*I and *Eco*RV). Fragments in the size range of 300–750 bp were adapted and subjected to magnetic bead capture using biotinylated capture molecules in a protocol provided by the manufacturer (CPG Inc., NJ, USA). Captured molecules were amplified and restricted with *Hin*dIII to remove the adapters. The resulting fragments were ligated into the *Hin*dIII site of pUC 19 plasmid. Recombinant molecules were electroporated into competent *Escherichia coli* DH 5α cells.

Clones were recovered by mixing 10–20 μL from an individual library with 20–30 μL of Genomic Grade Culture Media SOC Sterilized Broth (Growcells, Irvine, CA) and spread out onto one LB/AMP (100 mg/mL)/X-GAL (80 mg/mL)/IPTG (50 mM) plate (Teknova Inc., CA, USA), using a sterilized bacterial spreader to achieve a uniform distribution of the mixture. Eight plates were made for each library. Plates were grown overnight (ca 18 h) at 37°C in an incubator without light, and then immediately placed in 4°C for at least two additional hours to enhance the blue color. White (clone containing) colonies were picked at random with sterilized toothpicks and placed into 96-well blocks, each well containing 1 mL of LB broth plus Ampicillin (100 μg/mL). Blocks were covered with breathable film and grown overnight at 37°C with shaking at 250 rpm. Glycerol (15%) stocks were made of each 96-well block of overnight culture and stored at −70°C. The remaining solution in the blocks was spun down in a centrifuge at 2,500×*g* for 7 min. The solution was decanted and the culture blocks were covered with aluminum tape. Plasmid preparation was performed by alkaline lysis and by the OSU Recombinant DNA/Protein Resource Facility. The sequences for the recombinant clones were obtained by automated DNA sequencing via ABI BigDye® Terminator v1.1 Sequencing Kit (Applied Biosystems, CA, USA) analyzed on an ABI Model 3730 DNA Analyzer (Applied Biosystems, CA, USA) at the OSU Facility. The sequence data and quality information about the base calls of the clone sequences were analyzed by Sequence Scanner v1.0 software (Applied Biosystems Inc., Foster City, CA, USA). Only these SSR-containing sequences were selected that the base calls corresponding to the SSR flanking sites have a quality value (QV) greater than 20. The sequence parts from the vector were trimmed by the OSU Bioinformatics Core Facility. CAP3 program of Huang and Madan (1999) was used to form contiguous sequences (contigs) and to identify singletons (singlets). The CAP3 software parameters encompassing 'base quality cutoff for clipping', 'overlap length cutoff', 'overlap similarity score cutoff ', and 'minimum number of good reads at clip position' were set to default values while 'overlap identity cutoff = 95' and 'clipping range = 6' were selected to show the best alignment. Within individual aligned contigs, sequences containing redundant SSR sequences were visually removed one by one and were not included in further work.

Sequence checking and primer design

For SSR identification, search parameters selected for SSR Locator V.1 software (da Maia et al. 2008) included: 5-di-, tri-, tetra-; 4-penta-, hexa-; 3-hepta-, octa-, nona-, and decanucleotide repeats. All other parameters for SSR searching were set to program defaults. The SSR structure was simply divided into two broad descriptive categories: perfect and composite repeats. In a perfect SSR, the repeat sequence is not interrupted by any base not belonging to the motif, e.g., $(N_1N_2)_x$, $(N_1N_2N_3)_x$. All other SSR repeats deviating from perfect repeats were classified as composite repeats, which included imperfect, interrupted and compound SSR types according to Jones et al. (2001). The parameters of the primer design with SSR Locator V.1 software were as following: the amplicon size was set from 140 to 350 bp; GC clamp was set 0; the primer size was set from 18 to 22 bp with optimum size of 20 bp; the annealing temperature was set from 55 to 61°C with optimum of 59°C; the content of G/C was set from 45 to 50; both start and end point of a region scanned were set as automatic; and the end stability was set as 250. All primers were checked to identify redundancy. Unique SSR primers were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA). An M13-tagged sequence (CACGACGTTGTAAAACGAC) was added to the 5′ end of each forward primer.

In order to examine possible redundancy with published EST SSRs, comparison was performed using the forward primers from the 830 EST-SSR markers of Tobias et al. (2008) against the unique SSR-containing sequences from the genomic libraries. Each set of sequences was compared by specialized NCBI blast program for aligning two (or more) sequences together called bl2seq (Tatusova and Madden 1999). This specialized blast program was optimized for highly similar sequences per NCBI default parameters with the exception that the word size algorithmic parameter was changed from 28 to 16 due to the size of the primers (19–22 bp). Reverse primers were checked for those sequences with which the EST-SSR forward primers showed a successful alignment to the genomic SSR sequences.

Polymerase chain reaction and gel electrophoresis

SSRs were amplified using designed primer pairs on Biosystems 2720 thermal cyclers (Applied Biosystems, CA, USA) as described by Wu and Huang (2008) under the following PCR reaction condition in 10.5 µl volumes containing the following components: template DNA, 10 ng/µl × 1.5 µl; forward and reverse primers, 1.0 µM × 1.34 µl each; dNTPs, 10 mM × 0.2 µl each; $MgCl_2$, 25 mM × 0.6 µl; 10× PCR Buffer ($MgCl_2$ free), 1.0 µl; IR-M13 Forward primer, 1.0 µM × 0.2 µl; Taq enzyme 1.0×, 0.025 µl; Nuclease-free water, 4.295 µl. Cycling parameters were 95°C, 5 min followed by 14 cycles of 94°C, 20 s; 58°C, 1 min; 72°C, 30 s; additional 28 cycles of 94°C, 20 s; 55°C, 1 min; 72°C, 30 s; and a final extension of 72°C, 10 min. Added 5.0 µl Blue Stop Solution (95% formamide, 25 mM EDTA, and 2% bromophenol blue) to each PCR reaction tube (total volume being 15.5 µl), mixed thoroughly, spun down, and denatured for 3 min at 94°C in the thermo cycler. To detect amplified fragments of SSR markers, PCR products were separated using 6.5% KB plus polyacrylamide gel solution with a LI-COR 4300 DNA Analyzer (LI-COR Biosciences, Lincoln, NE, USA).

Primer screening, evaluation and data collection

Designed primer pairs were screened using SL93 7 × 15 and NL94 16 × 13 for their effectiveness to amplify SSR fragments of the expected size and to detect allele polymorphism. The effective primer pairs from the screening were confirmed and evaluated further on the four switchgrass cultivars. Every PCR reaction in both screening and evaluation of the two genotypes and four cultivars were replicated two times and replicated PCR products were run side by side in gel electrophoresis. The number and size of clean and reproducible DNA fragments were visually scored using Saga^{GT} software version 3.3 (LI-COR Inc., Lincoln, NE).

Results

SSR isolation

Initially, diagonal clones of two or four 96-well blocks for each library were sequenced to examine whether the enrichment procedure was successful. In the library constructed for enrichment of (ACC/GGT)n repeats, 25 clones out of 29 clones sequenced contained SSR sequences, but none of them was the expected repeats (data not shown), thus this library was discarded. In the other four libraries, predominant motifs were the expected, designated as Lib A (CA/TG), B (GA/TC), C (CAG/CTG) and D (AAG/CTT), and selected for the SSR development.

A total of 3,046 clones were sequenced from the four selected SSR-enriched genomic libraries (Table 1). Of the sequenced clones, 2,543 (83.5%) contained SSR loci, and 1,300 were identified as unique SSR clones. Redundancy was most found within the same libraries while fewer duplicates were also identified across libraries. Both Lib A and B of dinucleotide repeat motifs contained much more unique SSR clones than the other two libraries, C and D of trinucleotide repeat motifs (Table 1).

Characterization of SSR loci

With the unique SSR clones, 1,479 SSR loci were identified (Table 2). Based on the repeat type and structure, the SSRs of 71.0% unique clones in the four libraries belonged to perfect repeat types, ranging from 62.7 to 76.7%. Of the perfect repeat types, the average number of motif repeat units was the highest in Lib A while the lowest in Lib C. Totally, there

Table 1 Efficiency of SSR isolation from four SSR-enriched genomic DNA libraries

Lib	Motif	Clones sequenced	Clones containing SSRs	Unique SSR clones (%)
A	CA/TG	764	708	443 (62.6)
B	GA/TC	754	621	437 (70.4)
C	CAG/CTG	764	621	154 (24.8)
D	AAG/CTT	764	593	266 (44.9)
Total		3,046	2,543	1,300 (51.1)

Table 2 Frequency of repeat types isolated from four switchgrass genomic SSR-enriched libraries

Repeat type	Lib A	Lib B	Lib C	Lib D	Total
Perfect					
SSRs	311 (62.7%)	378 (76.7%)	140 (71.8%)	221 (74.9%)	1,050 (71.0%)
MRN	20.6	17.8	7.4	13.9	14.9
Composite					
SSRs	185 (37.3%)	115 (23.3%)	55 (28.2%)	74 (25.1%)	429 (29.0%)
MLN	3.2	2.5	3.4	3.2	3.1
MRN	10.9	11.3	7.5	11.1	10.2
Total SSRs	496	493	195	295	1,479

MRN mean repeat number, *MLN* mean loci number per clone

were 429 clones containing composite SSRs in the four libraries. Lib A had a higher percentage of composite repeats than the other three libraries. The mean number of SSR loci per composite clone was 3.1, ranging from 2.5 to 3.4 among the libraries. The average number of motif repeat units of composite type was 7.5 in Lib C to 11.3 in Lib B (Table 2).

The SSR frequency of each motif in the four libraries is presented in Fig. 1a–e. Lib A consisted of 78.1% CA/TG- and AC/GT-repeat sequences and Lib B had 81.6% GA/TC- and AG/CT-repeat motif sequences. Respective expected repeat motifs of the Lib C and D were 41.8% (GCA/TGC, CAG/CTG and GCT/AGC) and 49.2% (GAA/TTC, AAG/CTT and AGA/TCT). Over the four libraries, the predominant motifs of SSR loci were dinucleotide repeats (AC/GT, CA/TG, GA/TC and AG/CT), representing 78.1%. The expected trinucleotide motifs (GCA/TGC, CAG/CTG, GCT/AGC, GAA/TTC, AAG/CTT and AGA/TCT) accounted for 15.7%.

Primer design and evaluation

From the 1,300 unique SSR clones, 1,398 primer pairs were designed since some sequences contained more than one SSR locus (Table 3). Using these primer pairs, we found 1,030 (73.7%) primer pairs amplified products in expected size in SL 93 7 × 15 and NL94 16 × 13. Among the effective primer pairs, 832 (80.8%) generated clean and reproducible amplicons in the four switchgrass cultivars Alamo, Blackwell, Kanlow, and Dacotah. Examples of PCR products amplified by SSR primer pairs in SL93 7 × 15 and NL94 16 × 13 and in the four cultivars were shown in Fig. 2a, b. Marker names for the 1,030 primer pairs, along with SSR motif, primer sequences, T_m (melting temperature), expected size and amplified products are given in the Electronic Supplementary Material (ESM).

Polymorphism of SSR markers

The number of alleles per PP amplified on each of the two switchgrass genotypes and four cultivars were counted to evaluate the polymorphism of the SSR marker. The allele number was first examined in SL 93 7 × 15 and NL94 16 × 13. The number of alleles per genotype amplified by all of the SSR primer pairs in the two clones was 2.82, with a range of 1–11 alleles (Table 4, ESM). Among the 1,030 effective SSR primer pairs, 802 (77.9%) were polymorphic between the two genotypes (Table 4). Among the 832 consistent effective SSR primer pairs in the four cultivars, 655 (78.6%) were polymorphic. Average numbers of alleles per individual DNA sample were 2.88 for lowland cultivars and 2.85 for upland cultivars, respectively. Detailed information is given in Table 4. It is evident that SSR markers with dinucleotide repeat motifs (i.e., CA/TG and GA/TC) were more polymorphic than the SSRs of trinucleotide repeat motifs (CAG/CTG and AAG/CTT) (Table 4).

Discussion

Efficiency of genomic SSR marker development

Switchgrass has been recently recognized and intensively developed as a dedicated cellulosic crop for biofuel feedstock production in the United States. The use of SSR-enriched genomic libraries is an important means of developing SSR markers. Our results showed that 83.5% of the sequenced clones in enriched libraries were SSR-containing, and 51.1% of SSR-contained clones were unique. Compared with the other reported plant SSR-enriched libraries, the rate of SSRs containing clones in switchgrass was lower than the results obtained in sunflower (*Helianthus annus*, 89.0%, Tang et al. 2002), similar to the results obtained in zoysiagrass (81.7%, Cai et al. 2005), and eggplant (*Solanum melongena*, 81.7%, Nunome et al. 2009), but significantly higher than timothy (62.3%, Cai et al. 2003), Italian ryegrass (67.6%, Hirata et al. 2006), bunching onion (*Allium fistulosum*, 34.4%, Tsukazaki et al. 2007), melon (*Cucumis melo*, 60.8%, Fukino et al. 2007), and jute (*Corchorus olitorius*, 51.2%, Mir et al. 2009). The percentage of unique clones in this study was 51.1%, lower than those in jute (67.3%), sunflower (60.4%) and timothy (53.4%), higher than all of the other species mentioned

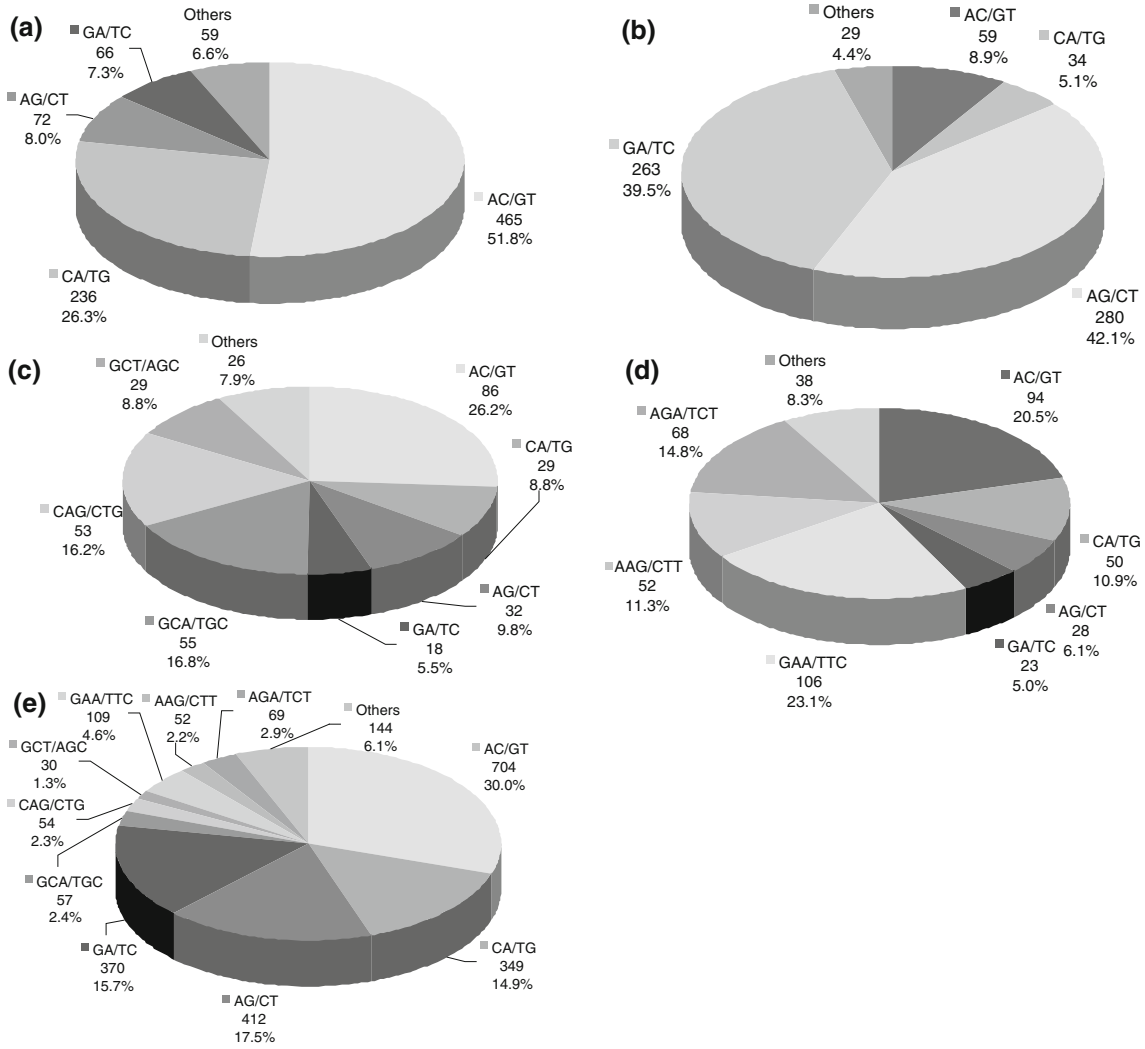

Fig. 1 Sequence frequencies of motif types in the genomic SSRs isolated from four libraries in switchgrass. **a** Sequence number and frequency of different SSR motifs in library A (CA/TG). **b** Sequence number and frequency of different SSR motifs in library B (GA/TC). **c** Sequence number and frequency of different SSR motifs in library C (CAG/CTG). **d** Sequence number and frequency of different SSR motifs in library D (AAG/CTT). **e** Collective sequence number and frequency of different SSR motifs over the four libraries

Table 3 Design and evaluation of SSR primer pairs in four switchgrass genomic libraries

Lib	Designed PPs	Strong amplification PPs	
		In two genotypes	In four cultivars
A	474	315	294
B	474	362	284
C	171	138	114
D	279	215	140
Total	1,398	1,030	832

above. We designed 1,398 SSR primer pairs from 3,046 enriched-SSR clones, the effective primers in two parents and in four switchgrass cultivars were 1,030 and 832, respectively, 33.8 and 27.3% of the total clones. These results indicated the enrichment procedure was efficient to generate a large number of SSR markers in switchgrass.

The most common dinucleotide repeats in genomic microsatellite sequence of *Arabidopsis thaliana*, are AT, GA, and then CA, and the most common trinucleotide motifs are AAG/TTC, ATC/TAG, AAC/TTG, and then AAT/TAA (Cardle et al. 2000). Due to its self-complementary nature, usually AT/TA is not used in construction for SSR-enriched libraries. According to the efficiency of genomic SSR marker development in this study, the motifs occurring at rates from high to low were CA (45.0%), GA (33.2%), AAG (9.7%), and CAG (6.0%). The higher rates of frequency and polymorphism seen in CA- and/or GA-containing microsatellites are in agreement with the previous results obtained in other grasses, such as tall fescue (Hirata et al. 2006; Saha et al. 2006), timothy (Cai et al.

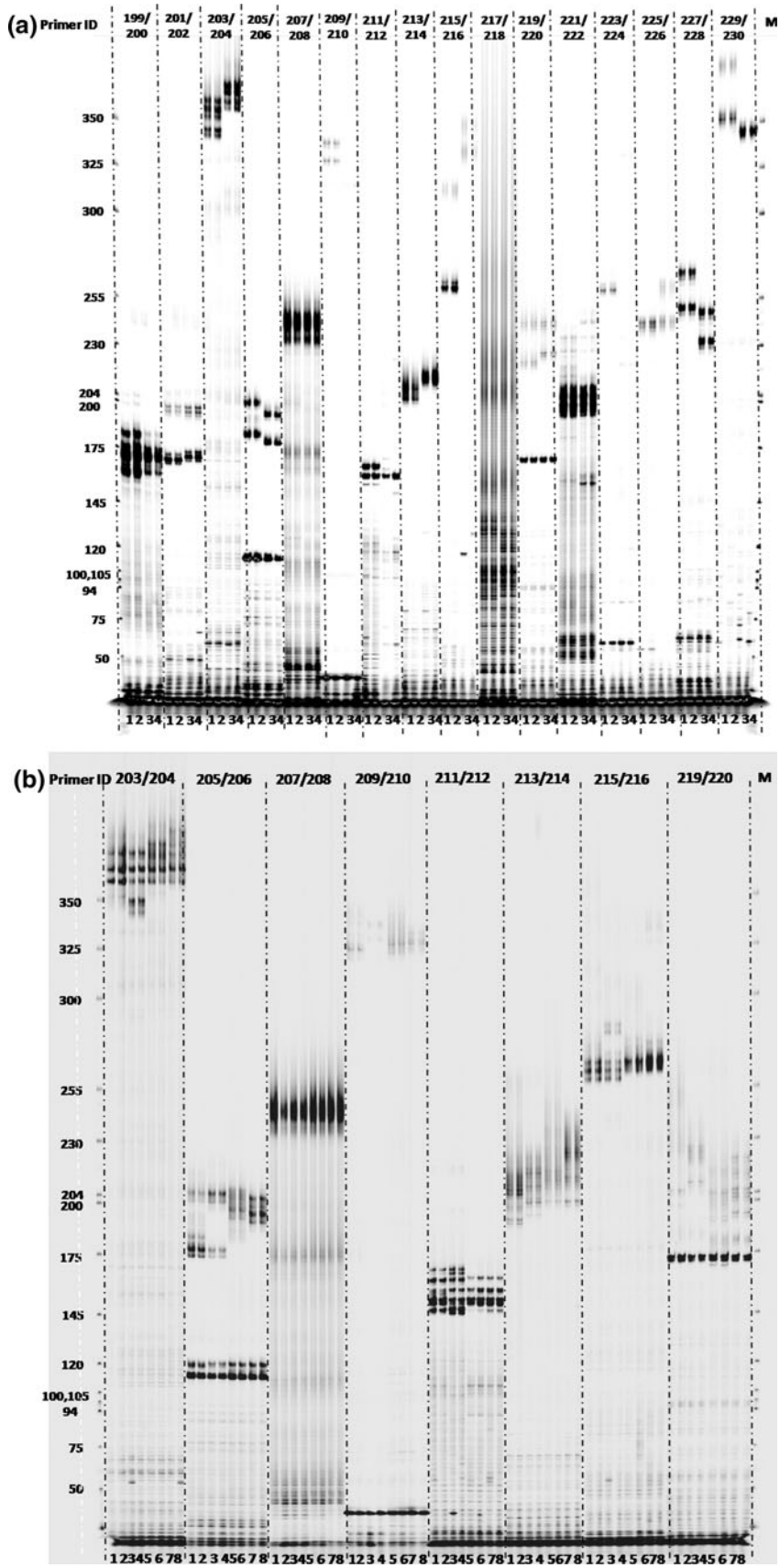

Fig. 2 **a** PCR products amplified by 16 primer pairs (PVCA 199–230 listed on the *top* of the gel image) in switchgrass genotypes 'SL93 7 × 15' (*gel lanes 1* and *2* labeled in each SSR primer pair panel below the bottom of the gel image) and 'NL94 16 × 13' (*lanes 3* and *4*) with two replications. Standard size markers are given on both sides. **b** PCR products in four switchgrass cultivars with two replications amplified by eight effective SSR primer pairs selected in Fig. 2a. Within each primer pair image panel, the order of DNA samples from *left* to *right* is Alamo (*lanes 1* and *2*), Kanlow (*lanes 3* and *4*), Blackwell (*lanes 5* and *6*), and Dacotah (*lanes 7* and *8*)

Table 4 Polymorphism of effective SSR primer pairs in four switchgrass genomic libraries

Lib	Mean number of alleles per genotype or cultivar				Polymorphic SSR PPs			
	SL93 and NL94[a]	Alamo and Kanlow	Blackwell and Dacotah	Among four cultivars	SL93 versus NL94	Alamo versus Kanlow	Blackwell versus Dacotah	Among four cultivars
A	2.51	2.61	2.46	2.54	252	205	238	256
B	2.27	2.48	2.47	2.48	305	176	181	251
C	3.38	3.17	3.18	3.18	90	46	42	75
D	3.13	3.25	3.27	3.26	155	37	39	73
Mean/total	2.82	2.88	2.85	2.87	802	464	500	655

[a] *SL93* = SL93 7 × 15, *NL94* = NL94 16 × 13

Table 5 Comparison between published cDNA primers (Tobias et al. 2008) and sequences of the genomic DNA libraries in this study using bl2seq basic local alignment search tool (BLAST) (NCBI, Bethesda, MD)

Marker ID of Tobias et al. (2008)	Genomic SSR sequence ID in this study	Associated SSR PP ID	% Read (similarity)	Read length (bp)	*E* value	Bit score*
sww875	PVCA7-A2	PVCA853/854	100	20	4.00E-08	38.1
sww2318	PVCA7-A12	PVCA 865/866	100	21	6.00E-09	39.9
sww1627	PVCAG4-E10	PVCA2477/2478	100	21	5.00E-09	39.9
sww2379	PVCAG6-C12	PVCA2567/2568	100	21	1.00E-08	39.9

* *Note* the bit score represents the normalized value that can be used to compare alignment scores from different searches

2003), perennial ryegrass (Jones et al. 2001), zoysiagrass (Cai et al. 2005), and other cereal crops. It appears that CA- and/or GA-enriched genomic SSR libraries are the most efficient in the development of SSR markers in the grass.

SSR marker polymorphism

SSR markers may have broad applications in switchgrass genetics and breeding, including germplasm diversity and relatedness analysis, cultivar identification, genetic mapping, QTL mapping, and eventually maker-assisted selection. SSR primer polymorphism was first examined in the two parents of our mapping population. A high polymorphism rate (77.9%) was detected in the two parents. This result indicated that SSR loci are highly variable between the two parents, and most markers would be useful for construction of an SSR-based linkage map. Furthermore, among these working primer pairs, 80.8% (832) of them amplified well among four switchgrass cultivars, which were representative of the diversity of lowland and upland ecotypes. Except in Lib A, both the intra- and inter-ecotype polymorphism rates were lower than the rate in two parents in other three libraries. It probably because the DNA samples of the four cultivars were isolated from leaf tissues pooled with 20 individual plants per each cultivar.

Of 96 EST-derived SSR primer pairs, 32 (33.3%) were effective in the production of SSR alleles in individuals of both Kanlow and Alamo (Tobias et al. 2006). In a later and much larger study, out of 1,780 EST-SSR primer pairs designed, 830 (46%) were effective, 298 (35.9% of the effective set) produced polymorphism bands between two parents of a mapping population (Tobias et al. 2008). Tobias et al. (2008) reported the mean number of amplicons per individual was 2.18. Our results indicated that the percentage of working primer pairs over the 1,398 designed was 73.7% in two parents, and 59.4% between Kanlow and Alamo. The mean numbers of amplicons per individual were 2.73 for Kanlow and 3.03 for Alamo. The high polymorphisms obtained with the genomic SSRs in our study were similar to the results in cereal crops, e.g., rice, wheat, and barley (Cho et al. 2000; La Rota et al. 2005; Chabane et al. 2005).

Sequence comparison of SSR loci between EST-SSRs and genomic SSRs

Comparison between the 1,300 trimmed sequences of unique SSR clones in this study and the sequences of 830 forward and reverse primer pairs from Tobias et al. (2008) *P. virgatum* cDNA library and 32 EST-SSR primer pairs of Tobias et al. (2006), showed 4 EST-SSR forward and reverse primers matched to 4 sequences of this investigation (Table 5). The four genomic sequences were used to develop four effective SSR primer pairs. This indicates that the four genomic SSR primer pairs amplify the same SSR loci as the four EST-SSR markers by Tobias et al. (2008).

Expressed sequence tags (ESTs) are a rich source for developing EST-derived SSR markers in switchgrass. The advantage of the EST-SSR approach is the low cost if the sequences are already available, although the polymorphism

rate for EST-SSR markers is lower. Tobias et al. (2005) reported 334 potential SSR loci, of which 79% of their SSR motifs were trinucleotide repeats, predominant motif was (CCG)n. Furthermore, Tobias et al. (2008) identified 2,817 EST-SSRs from 61,585 ESTs, most SSR loci were GC-rich trinucleotide, in particular, out of 830 reliable amplified primer pairs, 713 (85.9%) SSR loci were trinucleotide repeats. In contrast to EST SSRs, the predominant motifs of genomic SSR loci were dinucleotide repeats, representing 78.2%, and trinucleotide repeats only representing 15.7%. The proportion of dinucleotide repeat motifs was greater among genomic SSRs than EST-SSRs. This is comparable to those observed in earlier studies in rice, maize, poplar, tomato, cotton, and soybean (Cardle et al. 2000; Cho et al. 2000; La Rota et al. 2005). Sufficient published data from other plant and animal species have proved that most trinucleotide SSR loci possess low variability than dinucleotide-containing SSR loci (Chakraborty et al. 1997; Schug et al. 1998; Chabane et al. 2005). In our results, compared with trinucleotide motif repeats libraries, primers of dinucleotide motif repeats exhibited a higher level of polymorphism. In this respect, the results of our study are in agreement with the conclusions of previous research works. Redundancy between the four genomic libraries in this study and EST-derived SSR sequences by Tobias et al. (2008) is low (approximately 0.4%). This clearly indicates that we developed 1,026 novel SSR markers for switchgrass.

In conclusion, we reported the development of 1,030 genomic SSR markers from four enriched genomic SSR libraries in switchgrass. A high percentage of SSR polymorphisms in two parents of our mapping population and four switchgrass cultivars representing diversity of ecotypes were discovered. As the SSR markers are highly polymorphic they should be useful in genetic and breeding research in switchgrass.

Acknowledgments We gratefully thank Sharon Williams, Pu Feng, and Gary Williams for maintenance of plant materials used in this study in greenhouse, Dr. Ming Su from Oklahoma State University for helping format the Electronic Supplementary Material, and Ms. Yan Song and Dr. Peter Hoyt of the OSU Microarray and Bioinformatics Core Facility for trimming the sequences. This work has been supported, in part, by the NSF EPSCoR award EPS 0814361. We are also grateful to the China Scholarship Council (No. [2006] 3142) and Quality Forage Products Promotion Program (2006-G38) funded by the Ministry of Agriculture of China for supporting a visiting research stay of Yunwen Wang at Oklahoma State University.

References

Alderson J, Sharp WC (1994) Grass varieties in the United States. USDA Soil Conservation Service, Agriculture Handbook No. 170, Washington, DC, pp 194–199

Barnett FL, Carver RF (1967) Meiosis and pollen stainability in switchgrass *Panicum virgatum* L. Crop Sci 7:301–304

Bouton J (2008) Improvement of switchgrass as a bioenergy crop. In: Vermerris W (ed) Genetic improvement of bioenergy crops. Springer, New York, pp 309–345

Cai HW, Yuyama N, Tamaki H, Yoshizawa A (2003) Isolation and characterization of simple sequence repeat markers in the hexaploid forage grass timothy (*Phleum pretense* L.). Theor Appl Genet 107:1337–1349

Cai HW, Inoue M, Yuyama N, Takahashi W, Hirata M, Sasaki T (2005) Isolation, characterization and mapping of simple sequence repeat markers in zoysiagrass (*Zoysia* spp.). Theor Appl Genet 112:158–166

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Casler MD, Vogel KP, Taliaferro CM, Wynia RL (2004) Latitudinal adaptation of switchgrass populations. Crop Sci 44:293–303

Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. Genet Res Crop Evol 52:903–909

Chakraborty R, Kimmel M, Strivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

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

da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FI, Costa de Oliveira A (2008) SSR locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. Int J Plant Genomics. doi:10.1155/2008/412696

Fukino N, Sakata Y, Kunihisa M, Matsumoto S (2007) Characterization of novel simple sequence repeat (SSR) markers for melon (*Cucumis melo* L.) and their use for genotype identification. J Hort Sci Biotechnol 82(2):330–334

Gill GP, Wilcox PL, Whittaker DJ, Winz RA, Bickerstaff P, Echt CE, Kent J, Humphreys MO, Elborough KM, Gardner RC (2006) A framework linkage map of perennial ryegrass based on SSR markers. Genome 49:354–364

Gould FW (1975) The grasses of Texas. Texas A & M University Press, College Station

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Hirata M, Cai HW, Inoue M, Yuyama N, Miura Y, Komatsu T, Takamizo T, Fujimori M (2006) Development of simple sequence repeat (SSR) markers and construction of an SSR-based linkage map in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 113:270–279

Hitchcock AS (1951) Manual of the grasses of the United States, 2nd edn (revised by Chase A). United States Department of Agriculture Miscellaneous Publication 200, USA, pp 1–1051

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 X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Jones ES, Dupal MP, Kolliker R, Drayton MC, Forster JW (2001) Development and characterization of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Jones KC, Levine KF, Banks JD (2002) Characterization of 11 polymorphic tetranucleotide microsatellites for forensic applications in California elk (*Cervus elaphus canadensis*). Mol Ecol Notes 2:425–427

La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23

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

McLaughlin SB (1992) New switchgrass biofuels research program for the Southeast. Proceedings of the Annual Automatic Technology Development Contractor Coordinating Meeting, Dearborn, pp 111–115

McLaughlin SB, Kszos LA (2005) Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass Bioenergy 28:515–535

Mir RR, Banerjee S, Das M, Gupta V, Tyagi AK, Sinha MK, Balyan HS, Gupta PK (2009) Development and characterization of large-scale simple sequence repeats in jute. Crop Sci 49:1687–1694

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH (2008) Genetic diversity in switchgrass collections assessed by EST-SSR markers. Bioenergy Res 1:136–146

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

Nunome T, Negoro S, Kono I, Kanamori H, Miyatake K, Yamaguchi H, Ohyama A, Fukuoka H (2009) Development of SSR markers derived from SSR-enriched genomic library of eggplant (*Solanum melongena* L.). Theor Appl Genet 119:1143–1153

Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, De Souza AP (2006) Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed 125:378–384

Porter C (1966) An analysis of variation between upland and lowland switchgrass, *Panicum virgatum* L. Ecology 47:980–992

Saha MC, Mian R, Zwonitzer JC, Chekhovskiy K, Hopkins AA (2005) An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). Theor Appl Genet 110:323–336

Saha MC, Cooper JD, Rouf Mian MA, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449–1458

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol Biol Evol 5:1751–1760

Stubbendieck J, Hatch SL, Butterfield CH (1982) North American range plants. University of Nebraska Press, Lincoln

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1130

Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247–250

Tobias CM, Twigg P, Hayden DM, Vogel KP, Mitchell RM, Lazo GR, 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 BW, Barry K, McCann MC, Carpita NC, Lazo GR (2008) Comparative genomics in switchgrass using 61, 585 high-quality expressed sequence tags. The Plant Genome 1:111–124

Tsukazaki H, Nunome T, Fukuoka H, Kanamori H, Kono I, Yamashita K, Wako T, Kojima A (2007) Isolation of 1, 796 SSR clones from SSR-enriched DNA libraries of bunching onion (*Allium fistulosum*). Euphytica 157:83–94

Vogel KP (2004) Switchgrass. In: Moser LE, Burson BL, Sollenberger LE (eds) Warm-season (C4) grasses. ASA-CSSA-SSSA Monograph, Madison, pp 561–588

Warnke SE, Barker RE, Jung G, Sim SC, Mian MAR, Saha MC, Brilman LA, Dupal MP, Forster JW (2004) Genetic linkage mapping of an annual x perennial ryegrass population. Theor Appl Genet 109:294–304

Wu YQ, Huang Y (2008) QTL mapping of sorghum resistance to greenbugs by molecular markers. Theor Appl Genet 117:117–124