

**Towards the Development of a Molecular Map in Switchgrass:  
I. Microsatellite Marker Development**

Annual Report 2000

L.E. Gunter, Zamin K. Yang, G.A. Tuskan, and S.D. Wullschleger

Environmental Sciences Division  
Oak Ridge National Laboratory  
Oak Ridge, TN 37830-6034

Date Published-February 2001

Research supported by  
Biofuels Systems Division  
Office of Transportation Technologies  
Activity No. EB 5203

Report prepared by  
Environmental Sciences Division  
Oak Ridge National Laboratory  
Oak Ridge, TN 37831-6034  
Under subcontract 000-0000000

For  
Herbaceous Energy Crops Program  
Environmental Sciences Division  
OAK RIDGE NATIONAL LABORATORY  
Oak Ridge, TN 37831-6422  
Managed by  
UT-BATTELLE, LLC  
For the  
U.S. DEPARTMENT OF ENERGY  
Under contract DE-AC05-00OR22725

## Project Summary

The long-term goal of the switchgrass breeding program is to improve regionally adapted varieties and increase biomass yield and feedstock quality. Although, to some extent, biomass yields are dependent on environmental constraints, increased yield can be achieved through the development of genotypes with improved seasonal adaptation, tolerance to unfavorable environmental conditions, and improved resistance to pest and disease. To date, improvement in switchgrass has relied on recurrent breeding strategies based on phenotypic or genotypic selection. Yield improvements have been modest by this method. If we expect to make significant increase in yields, we need tools that will allow us to map complex traits and uncover the genes that influence them. A genetic linkage map could be a powerful tool for accelerating switchgrass development through marker-assisted selection, breeding and recombination. This type of mapping requires the development of markers that can be associated with phenotypic traits in a population of known pedigree. The most commonly used markers for mapping include restriction fragment length polymorphisms (RFLP) and simple sequence repeats (SSR).

At ORNL, we have been concentrating on the development of SSR markers, while our colleagues at the University of Georgia are developing RFLP markers in order to select parents to produce a mapping population and from there to create a framework map from ~100 F1 progeny.

- **Primary Objective FY00** – To develop 25-50 SSR markers for use in switchgrass genetic analysis.
  
- **Research Accomplishments** –
  - Cloning and sequencing of 1152 switchgrass genomic DNA fragments.
  - Analysis and screening of 364 SSR-containing sequences.
  - Design and testing of 27 SSR primer-pairs.

## Introduction

The genetic improvement of switchgrass (*Panicum virgatum* L.) is an essential part of the Department of Energy's Bioenergy Feedstock Development Program's improvement of switchgrass as an herbaceous energy crop. The use of DNA-based markers has become a standard approach for modern plant breeders interested in enhancing the efficiency of identifying and introducing desirable traits into a specific genetic background. A genetic map based upon molecular markers provides a framework to locating key genes and proteins through estimates of linkage and recombination. In addition, molecular markers are valuable tools for analysis of polyploid species, and have resulted in an increased understanding of the evolution and genetics of polyploids.

Before we can begin to map quantitative traits, however, we must first develop a large suite of informative markers and establish a mapping population of sufficient size to enable us to detect linkage. This is due to the fact that linkage estimates are based on the detection of alternative alleles in a mapping population that are inherited from the parental genotypes. Therefore, numerous markers and progeny help increase accuracy of functional relationships between markers and genes controlling phenotypic traits.

## Project Tasks and Objectives

- **Task 1** - Development of Molecular Markers
- **Objective** – To obtain markers that may be useful in marker-assisted selection and breeding, in QTL analysis, and in understanding the evolution of polyploidy in switchgrass.

Most of the well-developed QTL maps in agronomically important plants have been produced using single or low copy RFLP. However, the use of high-copy repetitive sequences such as SSR is increasing due to the fact these markers can fill in gaps and create a high-density map more suitable for molecular marker analysis of QTL. High information content due to the presence of multiple alleles, genome-wide distribution, and the ability to integrate readily with other types of molecular markers make SSR particularly useful in map construction. The ability to identify homologous sequences across species is one of the primary advantages of using SSR and RFLP for comparative mapping

- **Task 2** - Development of a Mapping Population in Switchgrass
- **Objective** – To associate markers with QTL for traits of interest the HEC switchgrass breeding program.

Parents of a mapping population must have sufficient variation for a trait of interest at the level of both DNA sequence and phenotype. The magnitude of phenotypic and marker variation is of utmost importance, especially when the goal of genetic mapping is to find genes controlling specific traits. When the parents represent extreme phenotypes, it is likely that genetic variation exists between the two phenotypes, provided that these differences are not totally environmentally based.

- **Task 3** - Estimation of Linkage and Origin of Ploidy
- **Objective** – To better understand the genetic background of switchgrass and establish linkage relationships for mapping.

Since a large number of fragments corresponding to different genomes comprising a polyploid species can potentially be amplified by a single primer or hybridize to a single DNA probe, it is necessary to analyze each fragment as an independent event. When designing a genomic mapping experiment, genetic marker type and sample size are of primary consideration. For both allo- and autopolyploids, the detection of linkage in coupling of single dose restriction fragments (SDRF) which can be derived from RFLP, SSR and RAPD markers, is a function of the recombination fraction; basically, the same as for detecting dominant markers in a diploid. However, unlike in an inbred diploid, where homozygosity of loci is assumed, we may need to screen large numbers of markers to find SDRF that vary in only one of the four (for tetraploid switchgrass) or eight (for octoploid switchgrass) genomes.

## Research Results

### Task 1 - Development of Molecular Markers

#### *Cloning*

Ten µg of genomic DNA extracted from 'Alamo' leaf tissue using a modified CTAB protocol was digested with *Pst* I (New England Biolabs). Fragments from 0.5kb - 1kb, 1kb - 2kb, 2kb - 3kb and 3kb - 5kb were separated on a 1 % low melting agarose gel (Gibco-BRL) and purified using the Wizard PCR preps DNA purification system (Promega). These purified fragments were cloned into the *Pst*I site of pBluescript KSII+ plasmid (Stratagene), transformed into *E. coli* DH5α using a standard protocol (Life Technologies), and selected on LB plates containing 100 µg/ml ampicillin, 50 µg/ml X-Gal and 1 mM IPTG.

Bacterial colonies containing plasmids inserted with switchgrass genomic DNA were picked and grown in TB with 100 µg/ml ampicillin overnight at 37°C at 225 rpm. Plasmids were purified from bacterial cultures using Qiaprep 96 Turbo Miniprep Kit (Qiagen) on the Biomek 2000 (Beckman-Coulter).

#### *Sequencing*

BigDye Terminator (Applied Biosystems) cycle sequencing reactions were performed on a GeneAmp PCR System 9700 using modified T7 (5' AAT ACG ACT CAC TAT AGG GC 3') or T3 (5' AAT TAA CCC TCA CTA AAG GG 3') promoters as primers (Life Technologies). Reactions were prepared using 2.0 µl BigDye Terminator Ready Reaction Mix (Applied Biosystems), 4 µl 5x buffer (400mM Tris pH 9.0/ 10mM MgCl<sub>2</sub>), 20 pmol primer and 200-500 ng purified plasmid template in a 16 µl reaction volume. Cycle sequencing on the GeneAmp 9700 included 60 cycles of 95°C for 15 s, 50°C for 5 s, and 60°C for 4 m. Extension products were ethanol precipitated according to standard protocol and samples were sequenced on an ABI PRISM 3700 DNA Analyzer (Applied Biosystems).

#### *Sequence Analysis and Primer Design*

Sequences were analyzed using software designed to identify SSR (Finrep, a Fortran program for finding mono-, di-, tri- and tetranucleotide repeats written by Stefano Leonardi; available at <http://www.fsl.orst.edu/tgerc/protocol.htm>). Nearly 32% of the clones sequenced contained at least a small SSR motif (Table 1); however, less than five percent contained SSR motifs of sufficient length to merit further testing. The most common dinucleotide repeats were GA/CT motifs (Figure 1) while tri- and tetranucleotide SSR were GC rich (Figures 2 & 3). Many of the repeats were not perfect; that is, there may have been slight alterations in the motif or insertions of non-repetitive nucleotides within the repeat sequence (for examples see Table 2).

Thirty out of a potential 51 sequences of 600 or more base pairs (bp) were analyzed using a primer-picking program (Primer3, a web-based application written by Steve Rozen and Helen J. Skaletsky; available at [http://www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)). Three of these sequences were determined to be duplicates and were discarded. Priming sites for the remaining twenty-seven were selected using 5' and 3' sequence information that would amplify an SSR containing product of approximately 200 bp (Table 2). These sequences were

then submitted for 50nmol scale commercial oligonucleotide synthesis (Operon) and used in subsequent PCR analysis.

### *PCR analysis*

We tested 27 SSR by designing forward and reverse primers based on sequences flanking the repeat motif and amplifying products from Alamo genomic DNA from which the sequences were derived under the following conditions: Reactions contained 10mM Tris-HCl pH 9.0, 50mM KCl, 0.1% Triton X-100, 2.0 mM MgCl<sub>2</sub>, 0.1 mg/ml BSA, 200 μM each dGTP, dATP, DTTP and dCTP, 10 ng each primer, 10-25 ng template DNA and 0.5 units *Taq* Polymerase (Promega). Amplification conditions on the GeneAMP 9700 included an initial denaturation at 94°C for 45 s; followed by 30 cycles of 94°C for 15s, 50°C for 15s, and 72°C for 1 min; and concluded with a 5 min extension at 72°C. PCR fragments containing SSR sequences were separated in 2.0% agarose gels in 0.5X TBE containing 0.1 μg/ml ethidium bromide and acquired and digitized under UV using the GelDoc 1000 gel documentation system (BioRad).

Out of the 27 primer-pairs tested, 19 (70%) amplified at least one DNA fragment in the Alamo control. We then tested these primers using 52 additional switchgrass templates, including 7 Alamo, Kanlow and Summer genotypes from Ken Vogel which are parents of potential mapping populations; 35 Alamo and Kanlow genotypes from a UGA phosphorous study that are also being screened as potential parents for mapping; and 10 upland and lowland populations from our 1996 RAPD marker study. All of the primer pairs that amplified fragments in the control also amplified fragments in all tested templates. The average number of alleles per SSR locus was 3-4; the absolute number ranged from 1-9 (Table 1). There is apparent polymorphism among genotypes for at least one allele in all but one of the loci tested (Figure 4).

In order to determine if automation could be used to process large numbers of samples at a time, we labeled ten primers with a 5' fluorophore prior to PCR amplification. The reaction products were then diluted 1:20 with water, denatured in 10 μl formamide and processed on the 3700 DNA Analyzer capillary electrophoresis system using GeneScan software to capture observed fluorescence of specific fragments. We did observe an increased sensitivity in discriminating similarly sized alleles using this system (Figure 6); however, this method may have some weaknesses (see discussion).

**Task 2** - Development of a Mapping Population in Switchgrass

**Task 3** - Estimation of Linkage and Origin of Ploidy

These tasks are dependent upon the development of a robust set of useful molecular markers. Task 2 will be one of the goals of the UGA Switchgrass Breeding Program's research, and Task 3, the largest and most time-consuming task, will be dependent on the correlation of markers with linkage groups.

## **Discussion**

Through cloning and sequence analysis of over 1000 switchgrass genomic DNA fragments, we have discovered 51 potentially useful SSR-containing sequences in switchgrass. We designed primers to amplify these SSR based on flanking sequence information, and have tested 27 primer-pairs to date. Eighteen primer-pairs amplified polymorphic products in both the Alamo control and in other switchgrass templates, suggesting that they will be useful across families and populations. One primer pair produced a two-allele monomorphic locus (i.e., the same alleles appeared in all tested samples). Since all primers were tested under identical conditions, it may be possible to increase resolution and amplification of certain products through reaction optimization (increase/ decrease annealing temperature; increase/decrease magnesium concentration). A change in reaction conditions may also be sufficient to produce a product in the nine tested primer-pairs that did not show initial amplification in the Alamo control. We will continue to optimize conditions as we develop additional markers.

One concern is our ability to use automation techniques while genotyping without losing information. The advantage to automation is that samples can be processed and analyzed more quickly and accurately. Although we are observing 2-3:1 alleles using capillary vs. agarose gel electrophoresis, the capillary system software can accurately call alleles only within a 50-500 bp range. Some of our alleles are in the 800-1000 bp range (see Figure 4A). We do have an imaging system that can digitize and analyze alleles in this range (the BioRad GelDoc and MultiAnalyst System) if they are based on an agarose or acrylamide gel electrophoresis system. The disadvantages to this system are that 1) we may underestimate the number of polymorphic alleles, since a difference of two to three basepairs (the size of one repeat unit) is virtually undetectable, or at best questionable in most agarose/acrylamide systems, and 2) it takes more time for laboratory personnel to collect and analyze data from agarose gels than from the capillary system. Nevertheless, as we continue the development of SSR in switchgrass, we will determine which system is optimal in terms of effort and information content.

Results obtained in 2000, both from the initial development of SSR and screening of 121 RFLP probes (Ali Missaoui, pers. comm.) suggest that co-dominant markers are abundant and potentially informative in switchgrass as they can be used to discriminate among individual genotypes. The outlook on use of these markers in mapping and marker-assisted selection is favorable. Further development of this technology is warranted in order to provide valuable information and assistance to switchgrass breeders to meet the overall HEC program objective of readying switchgrass for commercial production.

## **Proposed Work for 2001**

Based on our success in adapting protocols for SSR development in switchgrass, we will continue to sequence additional clones and develop new markers for use in genetic analysis. We are continuing to sequence additional clones and screen the resulting sequences for SSR motifs and design primers to amplify these regions. We will optimize PCR reaction conditions by adjusting salt concentration (1.5-2.5 mM MgCl<sub>2</sub>) and annealing temperature (50-55°C) and explore data analysis options of the resulting alleles (agarose vs. capillary electrophoresis or a combination of the two). We anticipate having 100-200 polymorphic SSR loci identified by the end of 2001.

We will then survey the two parents chosen by the UGA breeding program as candidates from which to make a mapping population to determine which SSR are potentially informative. A cross between the chosen parents will be made sometime around June. Once the F1 population is available for genotyping, we can use the RFLP markers developed by UGA and the SSR developed here at ORNL to generate data from the F1 for establishing a framework map.

In addition to using SSR for mapping work, we intend to apply these markers to the analysis of spatial and temporal differences in Kanlow and Alamo variety trial field plots from which genomic DNA samples have been collected at three-year intervals over the past seven years. The analysis of hypothetical changes in gene frequency in these plots was originally proposed using RAPD markers. However, since only preliminary data has been collected and analyzed on gene frequency data in a subset of the total samples, we plan to complete the study using SSR markers since they provide a more species-specific, reproducible, and robust class of genetic marker.

## **Activities**

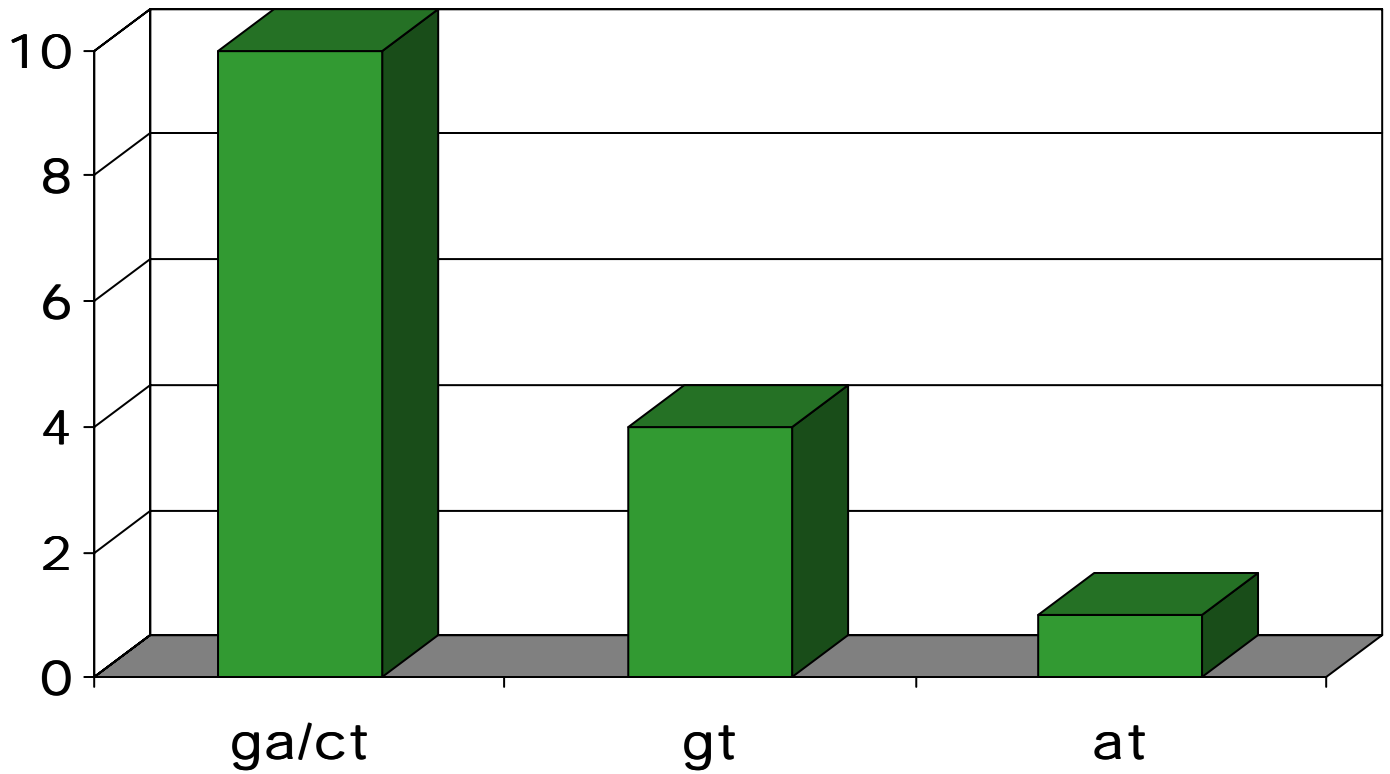
Gunter, LE, GA Tuskan and SD Wulschleger. 2000. Microsatellite Marker Development in Switchgrass. DOE/ORNL/Bioenergy Feedstock Development Program Herbaceous Research Contractor's Workshop. December 12-13, 2000, Atlanta, GA.

**Table 1.** Analysis of SSR-containing sequences from switchgrass sequencing project.

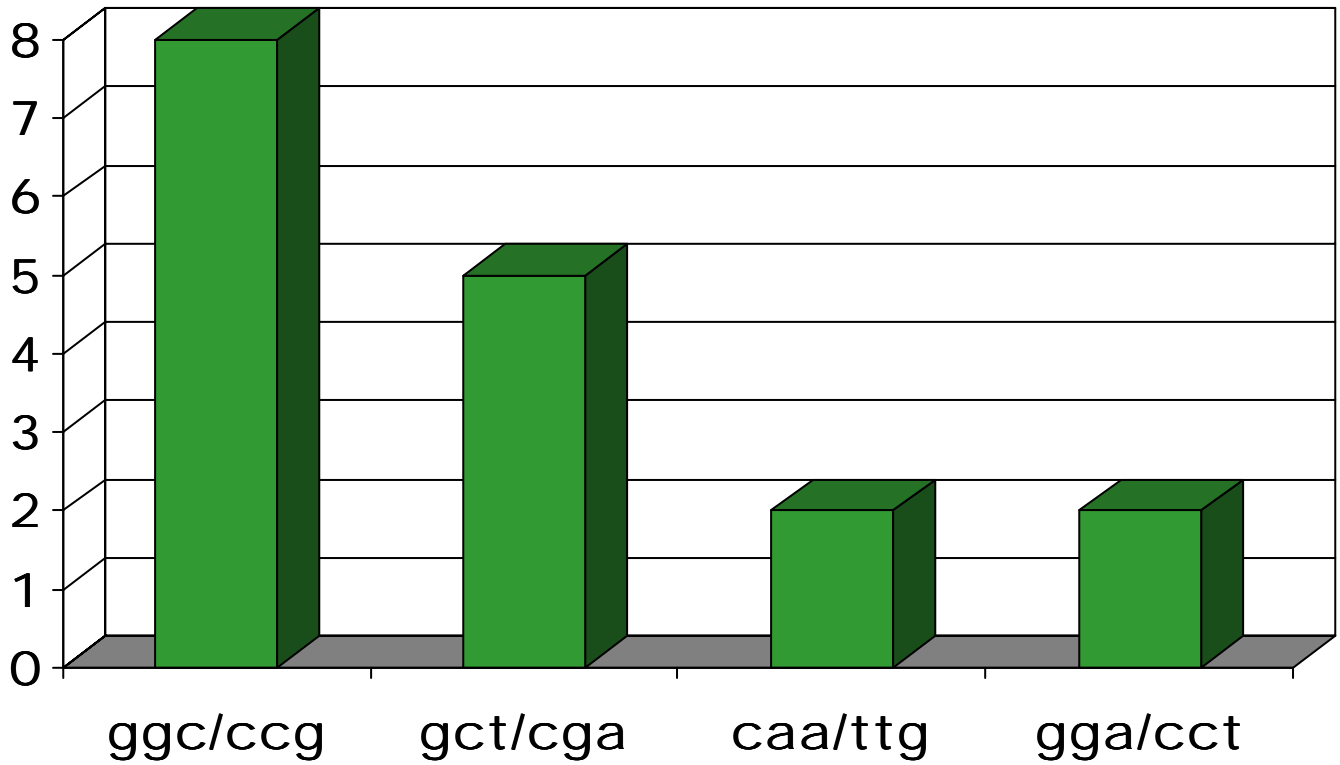
<i>Total number of sequences analyzed</i>	<b>1150 (575 clones)</b>
<i>SSR <math>\geq 4</math> repeat units</i>	<b>364 (31.6%)</b>
<i>Repeat type:</i>	
<i>Dinucleotide <math>\geq 7</math></i>	<b>16</b>
<i>Trinucleotide <math>\geq 5</math></i>	<b>25</b>
<i>Tetranucleotide <math>\geq 4</math></i>	<b>10</b>
<i>Total</i>	<b>51 (4.3%)</b>
<i>Primer pairs tested:</i>	
<i>Total</i>	<b>27</b>
<i>Product</i>	<b>19</b>
<i>Number of alleles/locus amplified</i>	<b>1-9</b>



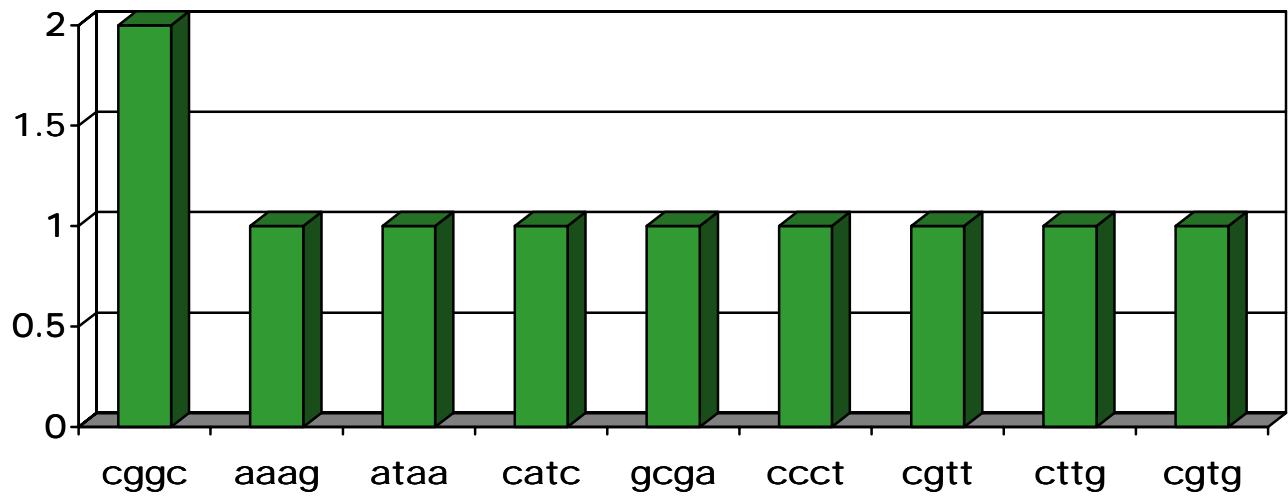
**Figure 1.** Frequency distribution of switchgrass dinucleotide SSR where the motif was repeated  $\geq 7$  times. Scale represents absolute numbers.



**Figure 2.** Frequency distribution of switchgrass trinucleotide SSR where the motif was repeated  $\geq 5$  times. Scale represents absolute numbers.



**Figure 3.** Frequency distribution of switchgrass tetranucleotide SSR where the motif was repeated  $\geq 4$  times. Scale represents absolute numbers.



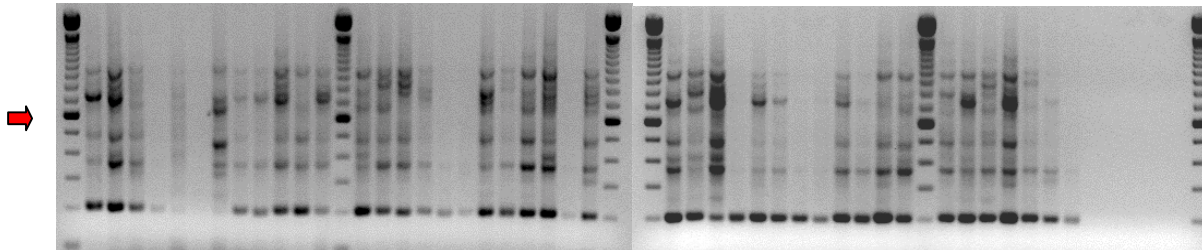
**Table 2.** Repeat type, primer sequence (F=forward; R=reverse) and average size of the most common allele at 19 SSR loci in *Panicum virgatum*.

Locus	Repeat	Primer Sequence (F)	Primer Sequence (R)	Size (bp)
PVSSR-001	(GA) <sub>11</sub>	CTGCTGCCGTTGCTATCC	GGCTCTGGAGGCTTGGAG	201
PVSSR-005	(GT) <sub>11</sub>	CGGCTCAAGCTTCCAAAAT	CGAGTGGTGGATGACTGGAT	210
PVSSR-007	(GA) <sub>26</sub>	TGCATGAGAAGCCAGAGAAA	TACTGCACTCGGAGGCTTTT	210
PVSSR-008	(CAA) <sub>8</sub> CA	GGCAAACAAAAAGGCAAACA	CAATGTGCCTGAACCTTGAA	197
PVSSR-009	(AG) <sub>2</sub> AT(AG) <sub>6</sub>	GATGACGGTTTTTGGGAGAA	GCCTAGCATCCTGCCTTACA	177
PVSSR-010	CT(CCT) <sub>4</sub>	ACATCATCAGCGCACAAACAT	TGGGGTGTGAGACTGTGAGA	209
PVSSR-011	A(AAAG) <sub>6</sub> AAA	CACGCAGTAGCATCTCGAAG	GGGCCTACTGAGGAGCAGAT	189
PVSSR-016	(CCG) <sub>2</sub> CTG(CCG) <sub>4</sub>	CGTGCTCACTTTGCTCACAT	TTGTGGCAGTGGATTAGCTG	193
PVSSR-017	GCG(CGG) <sub>5</sub> CCG	GGGCTGCAGCAGTAGTGG	AACTCCGGTATCCGCATCTT	173
PVSSR-018	(CATC) <sub>2</sub> (ATC) <sub>3</sub>	GTGAACTCGGTTGGACCCTA	GGCCCTTCTCCTTTGGTTT	189
PVSSR-019	(GCGA) <sub>4</sub>	TGGAAGGGTCCATTCTCGTA	ATGGCGCGCTTTTGTATT	194
PVSSR-020	A(AGG) <sub>4</sub> AG	GCTAGTGGGTTGTGCTCTCC	TCAGATTTGGCAAAGACGTG	224
PVSSR-023	(GT) <sub>3</sub> T(GT) <sub>7</sub>	TTTAGCAGTGTGCGGCTTAG	CTGCTTTTGACGAGGGTTTC	198
PVSSR-024	(CGG) <sub>8</sub>	GAACCTTCCACGAGCAAAC	CCTAGCAAGACCTCCCAACA	204
PVSSR-025	CGG(CCGG) <sub>3</sub> C	ATCAGATCAGCGCGATTAGG	GAACAAAACAAACCGCAAGC	217
PVSSR-026	(CCG) <sub>5</sub>	GCCAAAAGAATTGAACCAA	CGGGAGAACTACATGCAGAC	199
PVSSR-027	(GCT) <sub>6</sub>	TAATGTCGATGGTCGGCTTG	GCCGAGCAAGAGGTCGTC	205
PVSSR-028	CCT(CCCT) <sub>6</sub> (C) <sub>4</sub>	TCCTCCTCCCCACATCATTA	GGAGGAGTCAGGGCAAGAAT	210
PVSSR-029	(GCC) <sub>6</sub> (G) <sub>5</sub>	CGACAGCTCCGACGAGTT	CGGAGCTTCAATCGACGAC	158

**Figure 4.** Agarose gel fingerprint images of SSR alleles amplified from individual switchgrass genotypes. A. PVSSR-001 has a maximum of 9 different alleles in the analyzed templates approximately 75% of which are polymorphic. Both PVSSR-007 (B.) and PVSSR-023 (C.) have 4-5 alleles, but the variation in allele frequencies is much greater in PVSSR-023, suggesting that it may be a more informative locus.

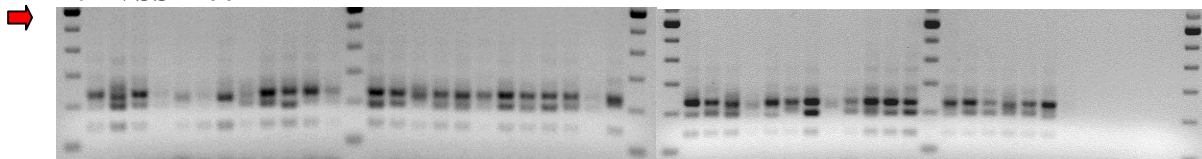
- ↑ Indicates a 100 bp increment ladder used to determine marker molecular sizes.
- ➡ Indicates the 600 bp standard within the ladder.

**A. PVSSR-001**

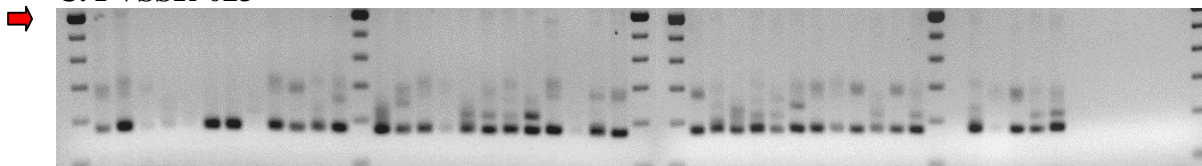


↑

**B. PVSSR-007**



**C. PVSSR-023**



**Figure 5.** This figure represents a comparison of an agarose gel image and capillary electrophoresis for discrimination of SSR alleles. The lower band of the gel image (left) in genotype P7 seems to co-migrate with bands in P6, P9 and A6, suggesting a size similarity. However, a size difference of three base pairs corresponding to a single copy of the CCT repeat motif comprising PVSSR-10 is clearly shown in the electropherogram on the right

