2-14-2006

Genic microsatellite markers derived from EST sequences of switchgrass (*Panicum virgatum* L.)

Christian M. Tobias
USDA, ARS, Western Regional Research Center, Genomics and Gene Discovery Unit, 800 Buchanan Street, Albany, CA

Daniel M. Hayden
USDA, ARS, Western Regional Research Center, Genomics and Gene Discovery Unit, 800 Buchanan Street, Albany, CA

Paul Twigg
University of Nebraska at Kearney, Kearney, Nebraska

Gautam Sarath
University of Nebraska - Lincoln, Gautam.sarath@ars.usda.gov

Molecular Ecology Notes (2006) **6**, 185–187 doi: 10.1111/j.1471-8286.2006.01187.x

PRIMER NOTE

Genic microsatellite markers derived from EST sequences of switchgrass (*Panicum virgatum* L.)

CHRISTIAN M. TOBIAS,* DANIEL M. HAYDEN,* PAUL TWIGG† and GAUTAM SARATH‡
*USDA, ARS, Western Regional Research Center, Genomics and Gene Discovery Unit, 800 Buchanan Street, Albany, CA 94710, USA, †Biology Department, Bruner Hall of Science, University of Nebraska at Kearney, Kearney, Nebraska 68849, USA, ‡USDA, ARS, Wheat, Sorghum and Forage Research Unit, Keim Hall, E.C. University of Nebraska, Lincoln, Nebraska 68583, USA

Abstract

Switchgrass is a large, North American, perennial grass that is being evaluated as a potential energy crop. There is a need to assess genetic diversity in stored accessions and in remaining native stands to assist breeding and conservation efforts. Marker development will also be necessary for genetic linkage mapping. Toward this end, 32 switchgrass genic di-, tri- and tetranucleotide repeat microsatellites were identified from expressed sequence tags (ESTs). These microsatellites were used to screen individuals from two different named cultivars. The markers displayed a high level of polymorphism consistent with the tetraploid, allogamous behaviour of the cultivars tested.

Keywords: energy crop, genetic diversity, polyploid, simple sequence repeat

Received 25 August 2005; revision accepted 12 September 2005

Switchgrass is a member of the subfamily Panicoideae closely related to proso millet, maize, sorghum and sugarcane. It is utilized as a summer forage, and also has promise as an herbaceous energy crop due to its large size and perennial nature. Improved genetics through breeding is the most economical route to improved value in the low input systems of cultivation employed for forage grasses. Inexpensive genetic marker systems for switchgrass are therefore required for marker-assisted selection to be successful. Such markers will also be useful for guiding conservation strategies to preserve remaining North American native prairie sites. Based on nuclear DNA content and analysis of segregation ratios, switchgrass has been shown to include both tetraploid and octoploid ecotypes. Restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) markers have been used for linkage studies and to discriminate ecotypes (Missaoui *et al*. 2005), but no microsatellite markers have been developed for this species. Expressed sequence tag (EST)-derived microsatellites have the added advantages of assaying functional diversity in transcripts, allowing greater transferability across species, and in some cases having additional utility as anchor markers for comparative mapping (Varshney *et al*. 2005). In this study, ESTs from cv. Kanlow (Tobias *et al*. 2005) were used as a source of sequence information for the development of microsatellites. Available sequence data from switchgrass were screened using the program SSRIT to identify microsatellite sequences larger than 20 bp (Temnykh *et al*. 2001). Primers were then designed using PRIMER 3.0 (Rozen & Skaletsky 2000) with an optimal primer T_m across the gene-specific regions of 60 °C and a target product size of 200 bases.

We isolated genomic DNA using a modification of a urea prep method (Shure *et al*. 1983). Forward primers were tailed with M13 (5′-CACGACGTTGTAAAACGAC-3′) to allow fluorescent labelling via incorporation of either a FAM- or a VIC-labelled M13 primer into the amplification mixture (Boutin-Ganache *et al*. 2001). The polymerase chain reactions (PCRs) were carried out in a final volume of 15 μL on an MJ Research PTC-225 thermocycler with a thermal profile consisting of a 2-min initial denaturation step at 94 °C followed by 35 cycles of 20 s at 94 °C, 20 s at the primer-specific annealing temperature (Table 1) and 1 min at 72 °C. A final 72 °C extension step of 30 min was included to promote nontemplated nucleotide addition at the 3′ end of the PCR product (Clark 1988). Reactions were carried out in 10 mM Tris-HCl (pH 9.0 at 25 °C), 50 mM KCl, 0.1% (v/v) Triton X-100, 1% (v/v) DMSO, 0.01% (w/v) gelatin, 0.75–2.5 mM

Correspondence: Christian M. Tobias, Fax: (510) 559–5818; E-mail: ctobias@pw.usda.gov

Table 1 Characterization of 32 microsatellites isolated from switchgrass cv. 'Kanlow' ESTs and tested on 10 individuals

Marker name	Repeat motif in library	Primer sequences (5′–3′)	Size range (bp)	T_a (°C) [$MgCl_2$ (mM)]	Total no. of bands	No. of bands per individual	Diversity index *D*	GenBank Accession no.
4821_A04	$(CAG)_7$	F: M13-GTCCGGGGTTCAGATAAACA R: CCGTTCAAGAGCAGGTCCTA	203–212	56 [1.5]	3	1–2	0.76	DN144448
4806_H12	$(GGC)_7$	F: M13-CCTCAATCCCTTTGAAATCC R: GTCCGTATCCTCCACGAGAG	503–519	58 [1.5]	5	1–2	0.6	DN143384
4808_F04	$(GCAT)_5$	F: M13-GGTCACCGTCAACTGATGC R: GGCAATGCCGATGATATAAA	221–233	55 [1.5]	4	1–2	0.74	DN143517
4816_C03	$(AG)_{16}$	F: M13-AGTTGAAGAGTGGCCTTGGA R: TCACCTTCCAACATTCACGA	215–279	51 [1.5]	8	1–2	0.88	DN144060
4827_D09	$(AG)_4CC$ $(AG)_6$	F: M13-CTTGCCTCGATTTGGAACTG R: GCGAGCTTGGAGGAGAGC	230–261	56 [1.5]	11	1–2	0.9	DN144982
4828_E03	$(TA)_{18}$	F: M13-CCTTACCTCCGGCATTTCTT R: ACAGGTACCAGTTCAATCACA	198–215	54 [2.5]	3	1–3	0.68	DN145072
4842_D11	$(GAG)_9$	F: M13-CGATCTGACCTGACCTCTCC R: GTACATGAAGCGGGTGATGA	181–193	54 [1.25]	5	1–3	0.74	DN146202
4848_B04	$(AT)_{13}$	F: M13-TCCAGATGACTCCCAGGAAC R: TCATCACTCGATTCCTCAAGC	214–281	54 [1.25]	14	1–3	0.88	DN146641
4853_D04	$(TA)_{11}$	F: M13-CCGGGAATGTTCGTCAAG R: TCCGATAGCCATGCATCATA	232–238	55 [1.5]	3	1–2	0.48	DN147088
4928_E12	$(TC)_{16}$	F: M13-GCTGAAACCAGGAAACGAAA R: CACCACACATCTGGCTTCTG	249–273	54 [2.5]	6	1–2	0.82	DN148200
4933_G06	$(GCC)_7$	F: M13-CACATCGATCACTTCATCTGG R: GCACTTCCATGACCTCATCA	210–219	53 [2.0]	2	2	0.18	DN148479
4934_H04	$(CGG)_7$	F: M13-CAGGATGCAGGAAGGGATTA R: CACGACGATGATGAGGACAA	167–204	54 [2.5]	4	2–3	0.74	DN148576
4937_G11	$(ACA)_7$	F: M13-TAGTCAGCAACCCCAAGGAG R: CAATGTGCCTGAACCTTGAA	194–207	56 [1.5]	6	1–2	0.88	DN148799
4941_F01	$(GA)_{17}$	F: M13-CTAGACGAGCAGCGTCAGTG R: CGGGGAGAGAGGGTAGAACT	229–231	54 [1.25]	3	1–2	0.46	DN149096
4942_H03	$(ATCT)_7$	F: M13-GTCCACGTCCTGGTAGAAGC R: CTGGTGCTGGGATCTCCAAT	202–210	54 [1.5]	3	1–2	0.34	DN149194
4943_D09	$(CCG)_8$	F: M13-ACAAGACCGAGCAGAGGATG R: TACACTTCCCCTTCGCAAAA	217–221	52 [1.5]	4	1–2	0.78	DN149247
5048_B06	$(GGC)_9$	F: M13-GCCCTCTTGTTCATGTCGAT R: ACCACCCGAGCCTTCC	328–354	54 [1.25]	5	1–3	0.78	DN142309
5005_B08	$(AT)_{29}$	F: M13-AGGAACATCATCACATCAGCA R: CAGCATCGTCGGCTTCTG	217–225	54 [2.5]	4	1–2	0.76	DN140825
5008_B05	$(GCA)_9$	F: M13-GCTGATTGCTCAATCCTGCT R: ACCTCCATGGTCACAACACA	206–218	52 [1.5]	7	1–3	0.88	DN140953
5013_F08	$(CCG)_9$	F: M13-CTCCAGTGCTCCACACTCCT R: AGACGAAGATGGCGAAGATG	204–255	54 [2.0]	9	1–2	0.86	DN141358
5023_A02	$(CAG)_9$	F: M13-CTCCAGTCTCAACCACAGCA R: CCATCCACCTTCATCAGCTC	294–315	56 [1.25]	8	1–3	0.88	DN141376
5028_C11	$(AG)_{33}$	F: M13-CTCCGTCTCTCCCTCTCTCC R: CACCATGCTCAAGAAGCTCA	136–191	56 [1.25]	10	2–3	0.90	DN141676
5036_D10	$(AG)_{30}$	F: M13-GCACAACTCGGTCGCTCTA R: TCCGCCTGCTTCATAATCTC	140–173	54 [2.5]	8	1–2	0.88	DN141907
5052_C08	$(CT)_{12}$	F: M13-AAGGTGAATATGGCCAACGA R: AAGGGAAAGATTCATGCGTTT	248–270	47 [1.25]	13	1–4	0.88	DN142646
5054_F06	$(TC)_{12}$	F: M13-CGAGCACAAATCGAAACTCC R: TCCTATGATGGCATGGATGA	157–191	52 [2.5]	11	0–4	0.88	DN142838
5057_A04	$(AG)_{15}$	F: M13-TCCACCTTCCTTCCATTCAC R: ACGTGGAACCAGGTGCAG	285–339	52 [0.75]	8	1–2	0.86	DN143022
5205_E06	$(CGG)_8$	F: M13-ACAGTCCCACCCTCACCTCT R: ATGACTACCACCGTCGGAGA	175–186	56 [1.5]	5	1–2	0.76	DN150540
5211_B07	$(AGC)_8$	F: M13-CCATTCCGATCCATCAAGAG R: CACGGCTCGGTCTTGTTAAG	244–255	55 [1.5]	4	1–2	0.78	DN150631

Table 1 *Continued*

Marker name	Repeat motif in library	Primer sequences (5′–3′)	Size range (bp)	T_a (°C) [$MgCl_2$ (mM)]	Total no. of bands	No. of bands per individual	Diversity index *D*	GenBank Accession no.
5214_B11	$(GGC)_7$	F: M13-ATTTGGTCGAGGGGAGGAG R: AGCTCGTACAGCTGCGAGTC	207–210	56 [1.5]	2	1–2	0.46	DN152503
5215_F12	$(GCA)_{10}$	F: M13-TTTTGCCCGTAACTTTGGTC R: CACCAGCATCCTCCTGGTAT	248–271	55 [1.5]	7	2–3	0.82	DN152378
5222_C09	$(GCA)_7$	F: M13-GCCGTCACGCTATAAGATCC R: GCCAGCATCCTTGCTTGTAG	182–188	56 [1.5]	3	1–2	0.6	DN150272
5232_A03	$(AG)_{35}$	F: M13-GGAGAGAAAACCAGAGCGAGT R: GCCGTACGTCTAGCTGGTG	153–208	56 [1.5]	12	1–2	0.9	DN150784

Repeat motif is listed 5′–3′ with respect to the forward primer (F). *D* is Nei's diversity index value for each marker. T_a is the annealing temperature. Numbers in parentheses indicate the $MgCl_2$ concentration used for each marker.

$MgCl_2$ (Table 1), 200 μM dNTPs in the presence of 0.4 U *Taq* polymerase (Promega), 15 ng template genomic DNA, 0.5 μM marker-specific reverse primer, 0.033 μM marker-specific M13-tailed forward primer and 0.5 μM VIC- or FAM-labelled M13 primer. Reaction products were polyethylene glycol (PEG)-precipitated, (Lis 1980) washed in 80% (v/v) ethanol and resuspended in 1 mM Tris pH 8.0 and 0.1 mM EDTA. Dilutions were subjected to capillary electrophoresis on an Applied Biosystems ABI 3730xl DNA analyser along with prepared PET-labelled size standards of 133–433 bp.

A total of 96 primer pairs were designed and tested; of these 52 primer pairs gave no or very little amplification. Twelve primer pairs were eliminated due to excessive stutter artefacts, irreproducibility or an excessive number of unsuccessful amplifications. Of the 32 usable markers, no single marker discriminated between populations. We sequenced PCR products amplified with primers for one marker (5232_A03) and found that the sequence agreed with its corresponding GenBank Accession and that observed length variations were exclusively due to gain or loss of AG repeats. Results from five individuals each of the tetraploid switchgrass cultivars 'Alamo' and 'Kanlow' were used to calculate the total number of polymorphic bands, the number of bands per individual and Nei's diversity index *D* (calculated as $1 - \Sigma p_i^2$) for each marker (Weir 1990). In most cases, the samples were highly polymorphic with markers containing dinucleotide repeats being on average slightly more variable. Tetraploid switchgrass cultivars behave largely as autopolyploids with a high degree of preferential pairing between chromosomes (Missaoui *et al*. 2005). However, in this study, partial heterozygotes could not be easily discriminated based on peak height, and thus determination of deviation from Hardy–Weinberg equilibrium, and linkage equilibrium was prevented. These markers will facilitate current mapping efforts in this species and population discrimination.

Acknowledgements

Supported by the United States Department of Agriculture, Agricultural Research Service CRIS 5325-21000-013-00, NP307 Biofuel and Bioenergy Alternatives. This work was also supported in part by NIH Grant P20 RR16569 from the BRIN Program of the National Center for Research Resources and by a University of Nebraska at Kearney Research Services Council grant.

References

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *BioTechniques*, **31**, 24–28.

Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research*, **16**, 9677–9686.

Lis J (1980) Fractionation of DNA fragments by polyethylene glycol induced precipitation. *Methods in Enzymology*, **65**, 347–353.

Missaoui AM, Paterson AH, Bouton JH (2005) Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. *Theoretical and Applied Genetics*, **110**, 1372–1383.

Rozen S, Skaletsky HJ (2000) PRIMER 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, NJ.

Shure MS, Wessler N, Fedoroff NV (1983) Molecular identification of the waxy locus in maize. *Cell*, **35**, 225–233.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Research*, **11**, 1441–1452.

Tobias CM, Twigg P, Hayden DM *et al*. (2005) Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass. *Theoretical and Applied Genetics*, (published online, June 28).

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, **23**, 48–55.

Weir B (1990) *Genetic Data Analysis: Methods for Discrete Population Genetic Data Analysis*. Sinauer Associates, Inc., Sunderland, Massachusetts.