

2011

Single-locus EST-SSR markers for characterization of population genetic diversity and structure across ploidy levels in switchgrass (*Panicum virgatum* L.)

Miki Okada

University of California - Davis

Christina Lanzatella

USDA-ARS, Genomics and Gene Discovery Research Unit

Christian M. Tobias

USDA-ARS, Genomics and Gene Discovery Research Unit, christian.tobias@ars.usda.gov

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

Okada, Miki; Lanzatella, Christina; and Tobias, Christian M., "Single-locus EST-SSR markers for characterization of population genetic diversity and structure across ploidy levels in switchgrass (*Panicum virgatum* L.)" (2011). *Publications from USDA-ARS / UNL Faculty*. 1314.

<http://digitalcommons.unl.edu/usdaarsfacpub/1314>

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

Single-locus EST-SSR markers for characterization of population genetic diversity and structure across ploidy levels in switchgrass (*Panicum virgatum* L.)

Miki Okada · Christina Lanzatella ·
Christian M. Tobias

Received: 6 July 2010 / Accepted: 11 October 2010 / Published online: 12 November 2010
© Springer Science+Business Media B.V. (outside the USA) 2010

Abstract Polyploidy, ploidal variation between populations, and aneuploidy within some populations complicate population genetic analyses in switchgrass. We report 21 genic-simple sequence repeat marker loci with single-locus disomic segregation in tetraploids and apparently tetrasomic inheritance in octoploids, thus allowing population genetic analyses across ploidy levels. Based on 472 individuals sampled over four tetraploid and eight octoploid cultivars, six to 55 alleles were detected per locus with an average of 24.1. Genetic diversity was greater in octoploids than tetraploids, as expected from polysomic inheritance. One tetraploid cultivar displayed comparable diversity to the least diverse octoploid cultivars, suggesting breeding history or population history in the native stands of origin may have also affected within-cultivar diversity. Amplicon number at each locus and population relationships suggest

autopolyploid origin of octoploids within upland tetraploids with significant cultivar differentiation. However, model-based Bayesian clustering of individuals indicated that closely related octoploid cultivars are difficult to identify, possibly due to slowed differentiation by polysomic inheritance. The analysis of the sampling effect indicated addition of loci is more effective for cultivar identification than more individuals sampled per cultivar. Discriminating power of loci tended to correlate with their variability. The eight loci with greatest discriminatory power within tetraploids were equally successful as 20 loci at identifying the four tetraploid cultivars. The set of markers reported in this study are useful for characterization of switchgrass germplasm and identifying population structure for association studies.

Keywords Bioenergy · Cultivar identification · Microsatellite marker · *Panicum virgatum* · Polyploidy · Population genetic diversity

Electronic supplementary material The online version of this article (doi:10.1007/s10722-010-9631-z) contains supplementary material, which is available to authorized users.

M. Okada · C. Lanzatella · C. M. Tobias (✉)
USDA-ARS, Genomics and Gene Discovery
Research Unit, Western Regional Research Center,
800 Buchanan Street, Albany, CA 94710, USA
e-mail: christian.tobias@ars.usda.gov

Present Address:

M. Okada
Department of Plant Sciences, University of California,
Davis, One Shields Avenue, Davis, CA 95616, USA

Introduction

Switchgrass is a once-dominant native component of the tallgrass prairie in North America with a wide geographical distribution from Central America to Canada (Hitchcock and Chase 1950). It is a warm season perennial C4 grass that has been planted for forage and soil conservation (Vogel and Jung 2001),

and is now regarded as a promising bioenergy crop in the US and Europe (Lewandowski et al. 2003; McLaughlin and Kszos 2005). Morphologically distinct lowland and upland ecotypes are recognized within the species (Porter 1966). The two ecotypes are also well differentiated based on molecular markers (Gunter et al. 1996; Hultquist et al. 1996; Missaoui et al. 2006; Narasimhamoorthy et al. 2008). Ploidy levels varying from diploid to duodecaploid have been reported (Nielsen 1944), and ploidal variation including aneuploids may be present within populations (Barnett and Carver 1967; Brunken and Estes 1975; McMillan and Weiler 1959; Narasimhamoorthy et al. 2008). However, the lowland populations are generally tetraploid, and the upland ecotype consists of populations that are either predominantly octoploid or tetraploid (Hopkins et al. 1996; Hultquist et al. 1997). Postzygotic incompatibility reproductively isolates tetraploids and octoploid plants based on very little seed set in controlled interploidy crosses (Martinez-Reyna and Vogel 2002). Switchgrass cultivars are derived from one or more collections of known geographical origins and include both upland and lowland ecotypes as well as tetraploid and octoploid populations (Alderson et al. 1995).

Polyploids are commonly classified as autopolyploid or allopolyploid according to the mode of origin, and alternative inheritance patterns occur in the two types of polyploids (Ramsey and Schemske 2002; Stebbins 1947). In general, polysomic inheritance is found in autopolyploids originating from genome duplications within a species, whereas the inheritance is disomic in allopolyploids with subgenomes derived from hybridization between divergent taxa. Also, diploidized paleopolyploids that switched from polysomic inheritance have diploid meiotic behavior (Le Comber et al. 2010; Wolfe 2001). In switchgrass, tetraploids have disomic inheritance based on bivalents observed in meiotic chromosomes in tetraploids (Barnett and Carver 1967) and F1 hybrids between upland and lowland tetraploids although rare univalents occurred (Martinez-Reyna et al. 2001). Analyses of molecular marker segregation in tetraploid linkage mapping populations (Missaoui et al. 2005; Okada et al. 2010) also indicate a high degree of preferential pairing at meiosis. On the other hand, multivalents observed in octoploid upland plants (Barnett and Carver 1967;

Brunken and Estes 1975) indicate polysomic inheritance in octoploids. Thus, in switchgrass, distinct inheritance patterns occur in the tetraploids and octoploids reflecting the difference in the mode or the timing of origin.

Because switchgrass is largely self-incompatible (Martinez-Reyna and Vogel 2002) and cross-bred, a cultivar is a population of plants with similar but diverse genotypes that are distinct from other cultivars based on the frequency of favorable alleles, as in most forage crops (Barker and Warnke 2001). Thus, population genetic methods are appropriate for quantifying diversity within and between cultivars and for identification of cultivars. In particular, codominant markers such as SSRs are useful because they enable estimates of allelic diversity at single loci. Single-locus codominant markers provide population genetic insights useful in germplasm conservation efforts, including detection of reduction in allelic variation and heterozygosity due to demographic bottlenecks (Nei et al. 1975) and non random mating (Jarne and David 2008). In addition, for autopolyploids, single-locus marker variation in populations of different ploidy levels within a species would provide insights into the evolution of ploidal variation (Parisod et al. 2010).

In polyploids, alleles at single SSR loci are difficult or impossible to ascertain because of simultaneous PCR amplification of duplicated loci (Obbard et al. 2006). However, contrary to the expectation, the linkage maps of lowland tetraploid switchgrass constructed with SSR and STS markers indicated significant subgenome differentiation, suggesting that PCR primer pairs for single-locus disomic markers may not be rare (Okada et al. 2010). We report a set of single-locus EST-SSR markers that have been placed on the linkage maps of lowland tetraploid switchgrass. The objectives of the study were to determine if the allelic variation detected at the markers is consistent with segregation at single-loci in diverse sample of cultivars representing both ecotypes and ploidy levels and to use the markers to quantify (1) the population genetic diversity within and among the sampled cultivars of different ploidy levels, (2) the clustering of individual multilocus genotypes according to the source cultivars, and (3) the effect of the number of loci and individuals per cultivar sampled on cultivar identification.

Materials and methods

Plant material

The seeds for 12 cultivars were obtained from the Germplasm Resources Information Network (<http://www.ars-grin.gov/>). The 12 cultivars included four and eight upland and lowland cultivars and variation in ploidy level, and 37–40 individuals were sampled per cultivar (Table 1). Although ploidal variation may be present within cultivars, since populations with ploidal variation are predominantly tetraploid or octoploid in switchgrass, the cultivars will be referred to as simply tetraploid or octoploid cultivars. Two pairs of cultivars of known close relationships were included among the 12 cultivars. ‘Shawnee’ was derived from ‘Cave-in-Rock’ as the base population (Vogel et al. 1996). ‘Trailblazer’ was developed from two base populations that share origins with ‘Pathfinder’ (Vogel et al. 1991). The seed were germinated on filter paper in petri dishes. The seedlings were transplanted into soil and grown in the greenhouse. Young leaves were harvested from each plant, dried, and stored in silica gel at room temperature prior to DNA extraction. Total genomic DNA was extracted using the CTAB method of Chen and Ronald (1999).

Table 1 Switchgrass (*Panicum virgatum*) cultivars included in the analysis

Cultivar	Accession number	Ecotype	Ploidy	Origin ^a
Alamo	PI 422006	Lowland	4x	Texas
Kanlow	PI 421521	Lowland	4x	Oklahoma
Dacotah	PI 537588	Upland	4x	North Dakota
Summer	PI 642191	Upland	4x	Nebraska
Blackwell	Grif 16409	Upland	8x	Oklahoma
Caddo	PI 476297	Upland	8x	Oklahoma
Cave-in-Rock	PI 469228	Upland	8x	Illinois
Falcon	PI 642190	Upland	8x	New Mexico
Grenville	PI 414066	Upland	8x	New Mexico
Pathfinder	PI 642192	Upland	8x	Nebraska and Kansas
Shawnee	PI 591824	Upland	8x	Illinois
Trailblazer	PI 549094	Upland	8x	Nebraska and Kansas

^a Origins according to Alderson et al. (1995)

SSR marker analysis

SSR markers used in the study were selected from EST-SSR markers designed using switchgrass expressed sequence tags and placed on linkage maps (Okada et al. 2010; Tobias et al. 2006, 2008). A panel of 64 individuals consisting of 16 plants from each of four cultivars, ‘Alamo’, ‘Kanlow’, ‘Blackwell’, and ‘Trailblazer’, was used to screen the markers for clean amplification and ease of scoring. The map positions were considered to avoid tightly-linked markers. Twenty-one EST-SSR loci were selected that displayed apparently diploid segregation in tetraploid cultivars with a maximum of two peaks observed per individual and up to four peaks in octoploid cultivars. The 21 apparently single-locus EST-SSR markers selected are on 14 of the 18 linkage groups in the tetraploid maps (Okada et al. 2010). Map distance between markers on the same linkage group ranged from 17.3 to 80.3 cM. PCR was carried out according to Tobias et al. (2006) with modifications: in a total volume of 5 μ L containing approximately 5 ng of template DNA, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 2 mM MgCl₂, 0.1% Triton X-100, 1% BSA, 200 μ M dNTPs, 0.5 μ M reverse primer, 0.033 μ M M13-tailed forward primer, and 0.5 μ M FAM-, VIC-, or NED-labeled M13 primers, and 0.1 μ L Taq polymerase. An MJ Research PTC-225 thermocycler was used with cycling conditions consisting of a 2-min initial denaturation step at 95°C followed by 32 cycles of 25 s at 95°C, 25 s at 54–56°C, 1 min at 72°C, and a final extension step of 30 min at 72°C. PCR products were polyethylene glycol (PEG)-precipitated, washed in 80% ethanol and resuspended in deionized formamide at 50°C for 10 min prior to capillary electrophoresis on an ABI3730xl DNA analyzer (Applied Biosystems, Foster City, CA) with PET-labeled size standards of 133–533 bp. PCR amplicons were sized using GENEMAPPER v. 3.7 (Applied Biosystems, Foster City, CA).

Data analysis

The 21 loci were tested for disomic versus tetrasomic modes of inheritance in tetraploid cultivars using the Bayes factor test (Catalán et al. 2006). Each EST-SSR locus within each tetraploid cultivar was analyzed as in the cases in which the PCR probe

amplifies a single locus in. The genotypes were scored as a single locus under both disomic and tetrasomic assumptions, and the Bayes factor was calculated using a spreadsheet (MS EXCEL).

To estimate genetic diversity within cultivars, number of alleles (A), gene diversity (H_E) (Nei 1978), and F_{IS} (Weir and Cockerham 1984) were calculated based on the genotype data using SPAGeDi version 1.3 (Hardy and Vekemans 2002). Allele frequencies using incomplete genotypes are calculated by giving equal weight to each allele recorded in each individual by SPAGeDi. Thus, in octoploid cultivars with no allele dosage data in partially heterozygous individuals, frequencies of common alleles may be underestimated and rare alleles overestimated slightly. The number of alleles detected, A , on the other hand is not affected by the absence of data on allele dosages in partial heterozygotes. To examine the partitioning of genetic variation among all cultivars and within and between ecotypes and ploidy levels, F_{ST} (Weir and Cockerham 1984) and R_{ST} (Slatkin 1995) were calculated based on the genotype data using SPAGeDi. To assess the pairwise differentiation between cultivars, F_{ST} (Weir and Cockerham 1984) and R_{ST} were calculated in SPAGeDi. Significance was tested using 1,000 permutations. To visualize the relationships among cultivars, a Neighbor-Joining dendrogram was constructed using PHYLIP version 3.63 (Felsenstein 2005) based on standard genetic distance (Nei 1978) calculated between each pair of cultivars by SPAGeDi.

Feasibility of cultivar identification was assessed using the set of 21 EST-SSR markers and the 12 cultivars sampled by quantifying the clustering of individual genotypes with respect to their source cultivars using STRUCTURE version 2.3.3 (Pritchard et al. 2000). STRUCTURE is a model-based Bayesian method to identify K clusters of individuals based on multilocus genotypic data by maximizing Hardy–Weinberg equilibrium and minimizing linkage disequilibrium within clusters. The individuals are simultaneously assigned to the identified clusters with probability q , which is the proportion of that individual's ancestry from each of the K clusters. The clustering analysis in STRUCTURE was carried out on the genotypic data on the octoploid cultivars using the RECESSIVEALLELES option to account for genotype ambiguity in the data (Falush et al. 2007).

STRUCTURE was run five times for each value of K ranging from one to the number of cultivars plus

one, using 50,000 burn-in runs and 100,000 Markov chain Monte Carlo replications. The admixture model and correlated allele frequencies (Falush et al. 2003) were used. Parameters were set to the default values. When there was a clear maximum value for $\ln P(D)$, the estimated posterior probability of the data for a given K , it was used to determine K . Otherwise, ΔK , an ad hoc quantity related to the second order rate of change of $\ln P(D)$ with respect to K was used to infer K at the uppermost hierarchical level of population structure (Evanno et al. 2005). The identified clusters were tested further for substructuring, using the same conditions as described above on the individuals with $q \geq 0.9$ and 0.8.

To examine the effect of the sampled number of loci and individuals per cultivar on cultivar identification, varying numbers of randomly selected loci and individuals were analyzed. Randomly selected individuals were treated as known members of the cultivar by setting USEPOPINFO = 1, and K was set to the number of cultivars analyzed. The remaining individuals within the cultivars were analyzed as having unknown origins (USEPOPINFO = 0) to be assigned to one or more of the K clusters. The results were summarized over ten independently created data sets for each particular combination of the numbers of loci and individuals per cultivar. The relative discriminatory power of each locus to assign individuals to source cultivars was determined with the software WHICHLOCI (Banks et al. 2003) using 1,000 resampled data with sample size of 40, allele frequency differential for population assignment, 90% assignment accuracy, and assignment stringency of LOD 5.0.

Results

Pattern of segregation

The 21 EST-SSR loci produced up to two and four distinct alleles per individual over 159 individuals in the tetraploid and 313 individuals in the octoploid cultivars, respectively, with only a few exceptions. Greater than two or four distinct alleles per individual were detected at low frequencies of 0.060 and 0.061% in tetraploids and octoploids, respectively.

The 21 loci were tested for disomic over tetrasomic mode of inheritance within tetraploid cultivars

based on the Bayes factor test (Catalán et al. 2006). The comparison of posterior probabilities for the two competing hypotheses of disomic and tetrasomic inheritance patterns favored the former, as twice the natural logarithm of the Bayes factor was greater than 10, at all 21 loci in each of the four tetraploid cultivars (Online Resources Table S1). For the octoploid cultivars, the same test was not performed because interpretation of the data under both assumptions of disomic versus tetrasomic inheritance necessary for the analysis was not possible. Under the disomic assumptions, genotypes of octoploid individuals at the homoeologous loci that were assumed to exist for the purpose of encoding the data could not be ascertained because of shared alleles between the loci, as expected from polysomic inheritance. In an attempt to objectively estimate allele dosages in partial heterozygotes, peak area ratios were used following Esselink et al. (2004) under the tetrasomic assumption. However, estimation of dosage failed because of wide variations in relative peak areas within pairs of amplicons. Also, some pairs of alleles simply did not occur together multiple times in fully heterozygous individuals within the sample to allow the estimation of dosage.

Genetic diversity and structure

The EST-SSR loci used in the study included nine and 12 loci with dinucleotide and trinucleotide repeat motifs, respectively (Table 2). The total number of alleles detected per locus over 472 individuals from 12 cultivars varied from six to 55 across loci with an average of 24.1 per locus. The EST-SSR loci with a dinucleotide and trinucleotide repeat motifs detected averages of 38.7 and 13.2 alleles per locus, respectively. The loci with dinucleotide repeat motif detected a significantly greater number of alleles than loci with trinucleotide motifs (Mann–Whitney U test, $P = 0.00033$).

All loci were polymorphic in all cultivars with the exception of sww2024 in ‘Dacotah’ and sww1394 and sww2165 in ‘Summer’. The number of alleles per locus within cultivars varied across cultivars from 4.1 per locus in ‘Summer’ to 14.0 in ‘Falcon’ (Table 3). Accordingly, gene diversity varied from 0.515 in ‘Summer’ to 0.784 in ‘Falcon’, as expected. Octoploid cultivars as a group had a significantly greater number of alleles and gene diversity within cultivars

Table 2 The EST-SSR loci selected from PCR probes placed on tetraploid linkage maps (Okada et al. 2010; Tobias et al. 2006, 2008)

Locus	SSR motif	Linkage group	Amplicon size range (bp)	Total number of alleles
sww88	AG	Va	138–244	55
sww659	CT	IIIb	260–307	27
sww2059	AC	IIa	186–272	43
sww2068	GA	VIIa	52–250	53
sww2558	GA	Ia	173–223	39
sww2980	AG	VIIIb	153–233	38
sww3005	TC	VIIa	162–252	41
sww3027	CT	VIb	370–438	32
sww3046	GT	VIb	170–232	20
sww545	AGC	VIb	344–364	9
sww1394	TCG	VIIa	191–224	11
sww1611	TGC	Ia	217–232	6
sww1729	GAG	VIa	187–214	26
sww1918	GAA	IVa	212–244	11
sww2024	GGC	IIIb	202–217	6
sww2036	CCG	IVa	174–201	13
sww2079	AGC	IVb	182–230	15
sww2165	GCG	IIB	171–205	13
sww2175	AGC	IXa	196–223	8
sww2260	CTC	IXa	500–536	28
sww2376	CTG	Vb	215–230	12

than the tetraploids (Wilcoxon signed rank test across loci for A , $P = 3.7 \times 10^{-5}$ and H_E , $P = 4.8 \times 10^{-7}$). Among the tetraploid cultivars, ‘Alamo’, had the highest A and H_E which were comparable to an octoploid cultivar, Cave-in-Rock, with the lowest diversity among the octoploids. Cultivar ‘Alamo’ also had the greatest number of private alleles, followed by cultivars ‘Falcon’, ‘Grenville’, ‘Blackwell’, and ‘Kanlow’. Cultivar pairs ‘Cave-in-Rock’ and ‘Shawnee’ as well as ‘Pathfinder’ and ‘Trailblazer’ with known shared ancestry in the breeding history had a relatively small number of private alleles, as expected. The least number of private alleles was found in cultivars ‘Dacotah’, ‘Summer’, and ‘Shawnee’ with only one private allele.

Significant genetic differentiation was indicated by F_{ST} and R_{ST} among all cultivars (Table 4). Although among-cultivar variation is highly significant, 89% of the total genetic variation was found within cultivars based on F_{ST} among all populations. All pairwise F_{ST}

Table 3 Genetic diversity detected within cultivars over 21 EST-SSR loci

Cultivar	Ploidy	N	A	H_E	F_{IS}	Total alleles per individual	Private alleles
Alamo	4x	40	10.7	0.724	0.085**	34.8	24
Kanlow	4x	39	7.8	0.632	0.009 ^{ns}	34.0	9
Dacotah	4x	40	6.5	0.515	0.120**	30.4	1
Summer	4x	40	4.1	0.499	0.051*	30.7	1
Blackwell	8x	40	13.5	0.781	-0.096**	56.4	10
Caddo	8x	40	13.0	0.782	-0.094**	56.1	4
Cave-in-Rock	8x	39	9.7	0.735	-0.117**	52.8	6
Falcon	8x	39	14.0	0.784	-0.106**	57.0	18
Grenville	8x	40	12.2	0.762	-0.098**	55.9	12
Pathfinder	8x	40	11.0	0.758	-0.109**	55.6	3
Shawnee	8x	38	10.8	0.756	-0.079**	53.0	1
Trailblazer	8x	37	11.2	0.768	-0.097**	54.6	5

N sample size, A number of alleles detected, H_E gene diversity (Nei 1978), F_{IS} inbreeding coefficient, calculated using SPAGeDi version 1.3 (Hardy and Vekemans 2002)

^{ns} $P \geq 0.05$; * $P < 0.05$; ** $P < 0.001$

Table 4 Genetic differentiation detected at 21 EST-SSR loci between ecotypes and among switchgrass cultivars estimated using SPAGeDi version 1.3 (Hardy and Vekemans 2002)

Level	No. of groups	F_{ST}^a	R_{ST}
All cultivars	12	0.112	0.105
Between upland and lowland ecotypes	2	0.136	0.200
Within lowland ecotype	2	0.145	0.163
Within upland ecotype	10	0.078	0.048
Between ploidy	2	0.042	0.045
Within tetraploids	4	0.275	0.326
Within octoploids	8	0.048	0.028

^a All F_{ST} and R_{ST} values are significant at $P < 0.0000$

values were significant after Bonferroni correction ($P < 0.01$) (Table 5). Pairwise F_{ST} values were notably lower than others in cultivar pairs with recorded shared ancestry in the breeding history: ‘Cave-in-Rock’/‘Shawnee’ and ‘Pathfinder’/‘Trailblazer’. The lowest pairwise F_{ST} was observed between cultivars ‘Blackwell’ and ‘Caddo’, both originating from Oklahoma. Pairwise R_{ST} was not significant for the three above pairs of cultivars after Bonferroni correction ($P \geq 0.05$) and in many other pairs of upland octoploid cultivars, mostly involving cultivars ‘Blackwell’, ‘Caddo’, ‘Falcon’, ‘Grenville’,

‘Pathfinder’, and ‘Trailblazer’, suggesting recent divergence.

F_{ST} and R_{ST} between ecotypes were greater than between ploidy (Table 4), indicating that tetraploid upland cultivars are more closely related to the upland octoploids than to the lowland tetraploids. Accordingly, within tetraploids (including upland and lowland ecotypes) F_{ST} and R_{ST} were substantially greater than within upland ecotypes (including tetraploids and octoploids). The Neighbor-Joining dendrogram of pairwise standard genetic distance between cultivars also show greatest differentiation between the ecotypes and a close genetic relationship between upland tetraploids and octoploids, consistent with F_{ST} and R_{ST} (Fig. 1). Two pairs of cultivars with known shared breeding history, ‘Pathfinder’/‘Trailblazer’ and ‘Cave-in-Rock’/‘Shawnee’, clustered together as expected. Consistent with low pairwise F_{ST} , cultivars ‘Blackwell’ and ‘Caddo’ also clustered closely together. Also, as expected the upland and lowland tetraploids clustered separately. Interestingly, however, upland octoploid as a group did not cluster together. In addition, upland tetraploid ‘Summer’ from Nebraska was more closely related to the upland tetraploid cultivar ‘Dacotah’ from North Dakota than to the octoploids, ‘Pathfinder’ and ‘Trailblazer’, from similar geographical origins of Nebraska/Kansas.

Table 5 Pairwise F_{ST} (upper half) and R_{ST} (lower half) between cultivars calculated using SPAGeDi version 1.3 (Hardy and Vekemans 2002)

	Alamo	Dacotah	Kanlow	Summer	Blackwell	Caddo	CIR	Falcon	Grenville	Pathfinder	Shawnee	Trailblazer
Alamo	–	0.261	0.163	0.269	0.142	0.138	0.172	0.176	0.162	0.174	0.151	0.166
Dacotah	0.281	–	0.478	0.142	0.123	0.138	0.057	0.077	0.098	0.138	0.061	0.095
Kanlow	0.145	0.351	–	0.492	0.250	0.238	0.337	0.309	0.286	0.275	0.322	0.311
Summer	0.298	0.163	0.364	–	0.135	0.164	0.069	0.120	0.094	0.142	0.054	0.125
Blackwell	0.144	0.147	0.193	0.168	–	0.017 ^{ns}	0.047	0.024	0.014 ^{ns}	0.015 ^{ns}	0.044	0.011 ^{ns}
Caddo	0.135	0.147	0.189	0.164	0.005	–	0.066	0.034	0.025	0.025	0.056	0.031
CIR	0.180	0.161	0.224	0.162	0.078	0.079	–	0.037 ^{ns}	0.029	0.054	–0.001 ^{ns}	0.033
Falcon	0.158	0.111	0.217	0.123	0.042	0.050	0.077	–	0.008 ^{ns}	0.016 ^{ns}	0.035 ^{ns}	0.012 ^{ns}
Grenville	0.154	0.128	0.200	0.143	0.035	0.035	0.092	0.043	–	0.004 ^{ns}	0.025	0.009 ^{ns}
Pathfinder	0.148	0.140	0.210	0.168	0.020	0.023	0.091	0.048	0.046	–	0.054	0.010 ^{ns}
Shawnee	0.161	0.141	0.212	0.155	0.051	0.054	0.009	0.056	0.064	0.060	–	0.037
Trailblazer	0.139	0.158	0.198	0.177	0.018	0.023	0.084	0.050	0.048	0.012	0.054	–

All F_{ST} and R_{ST} values were significant at $P < 0.01$ after Bonferroni corrections unless denoted ns ($P \geq 0.05$)

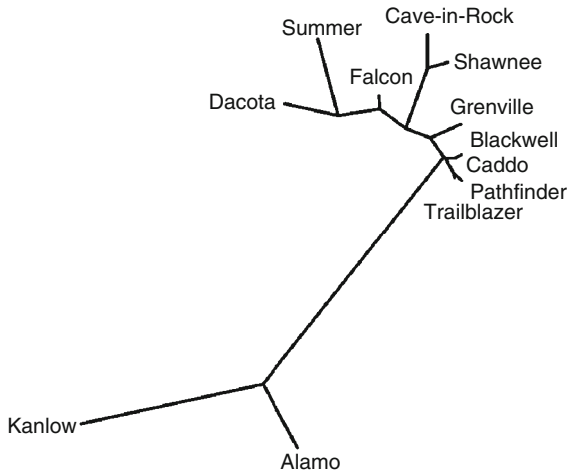


Fig. 1 Neighbor-joining tree based on standard genetic distance (Nei 1978) between cultivars calculated using SPAGeDi version 1.3 (Hardy and Vekemans 2002)

Cultivar identification

The clustering of 472 individuals sampled from 12 cultivars by STRUCTURE using the 21 EST-SSR loci resulted in gradual increase in the probability of the data, $\ln P(D)$, as the number of clusters, K , increased (Fig. 2), with no clear maximum, indicative of hierarchically structured populations (Evanno et al. 2005). Thus, ΔK was used to infer K at the uppermost hierarchical level of population structure. Clusters identified were tested for substructure using

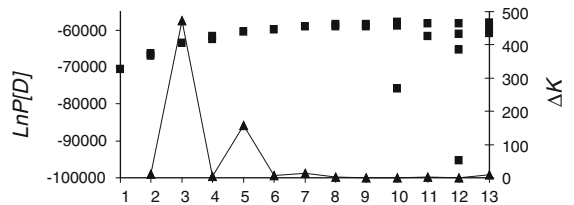


Fig. 2 Plot of the log likelihood of the multilocus genotypic data, $\ln P(D)$, shown as squares for five runs at each value of K clusters, and the second order rate of change in $\ln P(D)$, ΔK , shown as triangles as a function of the number of clusters, K , from a model-based Bayesian clustering analysis (Pritchard et al. 2000; Evanno et al. 2005) on 472 individuals from 12 switchgrass cultivars

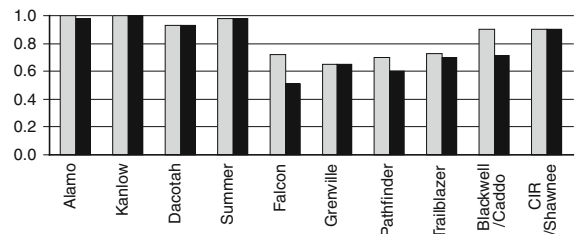


Fig. 3 Proportion of individuals with probability of ancestry (q) greater than 0.9 and 0.8 to the clusters identified by the model-based Bayesian clustering analysis using STRUCTURE (Pritchard et al. 2000; Falush et al. 2007) and ΔK (Evanno et al. 2005) to identify the upper hierarchical structure

individuals with high assignment probability ($q \geq 0.9$ and 0.8) if the cluster consisted of more than one cultivar (Online Resources Fig. 1S). The

analysis led to identification of 10 clusters (K) at $q \geq 0.9$ or 0.8 (Fig. 3). There were no cases of incorrect identification as all individuals with high assignment probability ($q \geq 0.9$ and 0.8) assigned to the cluster defined by the source cultivar(s). Each of the four tetraploid cultivars was identified as a distinct genetic cluster, whereas two pairs of closely related octoploid cultivars ('Blackwell'/'Caddo' and 'Cave-in-Rock'/'Shawnee') resulted in two single genetic clusters at $q \geq 0.9$ or 0.8. The proportion of individuals assigning highly to the source cultivar/cluster varied from 0.93 to 1.00 at $q \geq 0.80$ and 0.9 for tetraploid cultivars, but for octoploid cultivars it was generally lower than tetraploids, varying from 0.60 to 0.90 at $q \geq 0.8$ and from 0.41 to 0.90 at $q \geq 0.9$.

The effect of the number of loci and individuals sampled per cultivar on cultivar identification was examined using tetraploid cultivars. Only tetraploid cultivars were used in the analysis because cultivar identification based on population assignment was not successful in octoploids as a group based on the full data set. Random subsamples of the data set with different number of loci and individuals were analyzed with the USEPOPINFO option in STRUCTURE. The subsamples consisted of five to 20 loci per genotype and 5–20 individuals per cultivar. Nine combinations representing different number of loci and sample size were examined with 10 randomly subsampled data sets per combination. Successful clustering of the individuals to the source cultivar varied slightly among tetraploid cultivars (Fig. 4). Individuals of 'Kanlow' and 'Summer' tended to cluster successfully with high assignment probability to source cultivars with fewer loci and individuals sampled than 'Alamo' and 'Dacotah'. The proportion of highly assigning individuals ($q \geq 0.9$ or 0.8) generally increased with greater number of loci or individuals sampled. Greater number of loci tended to improve the clustering of individuals to the source cultivar more than number of individuals per cultivar sampled, and $q \geq 0.8$ rather than 0.9 resulted in more successful cultivar identification. Correct identification of all individuals in all four cultivars was achieved when 20 loci were used.

The 21 loci were ranked with respect to the discriminatory power on the four tetraploid cultivars using WHICHLOCI at the stringency of LOD 5.0 and minimum assignment of 0.9 (Table 6). More variable

loci with greater number of alleles tended to correlate positively with discriminatory power of the locus. The eight loci identified by WHICHLOCI to be sufficient to assign all individuals to source cultivars were also used to test the clustering of individuals using STRUCTURE (Fig. 4). The eight loci with only five individuals per cultivar resulted in correct identification of all samples at $q \geq 0.8$.

Discussion

Population genetic analyses of switchgrass cultivars are complicated by polyploidy, ploidal variation between cultivars, and potential within-cultivar ploidal variation including aneuploidy. The EST-SSR marker loci reported in this paper reduced the complexity of the analysis by detecting only a single disomic or apparently tetrasomic locus in tetraploid and octoploid cultivars, respectively, thus allowing population genetic analyses of the data across all cultivars. Unfortunately, the marker loci were not suitable for assessing allele dosages in partial heterozygotes in octoploid cultivars using peak area ratios (Esselink et al. 2004). Consequently, diversity measures that require allele frequency estimation in octoploid cultivars need analysis programs that handle missing data. Nevertheless, the number of alleles per locus as a diversity measure is not affected by the absence of allele dosage information and provided an unbiased estimation of diversity across ploidy levels. Higher ploidy variants or aneuploids that increased the number of alleles per locus appeared absent or possibly present in extremely low frequencies based on the marker data, consistent with observations of within-population variation in chromosomal number (Brunken and Estes 1975). Aneuploids occurred frequently in upland octoploid populations but had regular meiotic disjunction. However, because most of the marker loci are on different chromosomes, presence of aneuploids is not a source of bias for population genetic diversity analyses.

The analysis of genetic diversity in tetraploid and octoploid cultivars using single-locus EST-SSR markers provided insight into the genome structure and origin of the octoploids. The inheritance patterns observed at the marker loci indicate an autopolyploid genome structure in the octoploids. Furthermore, the

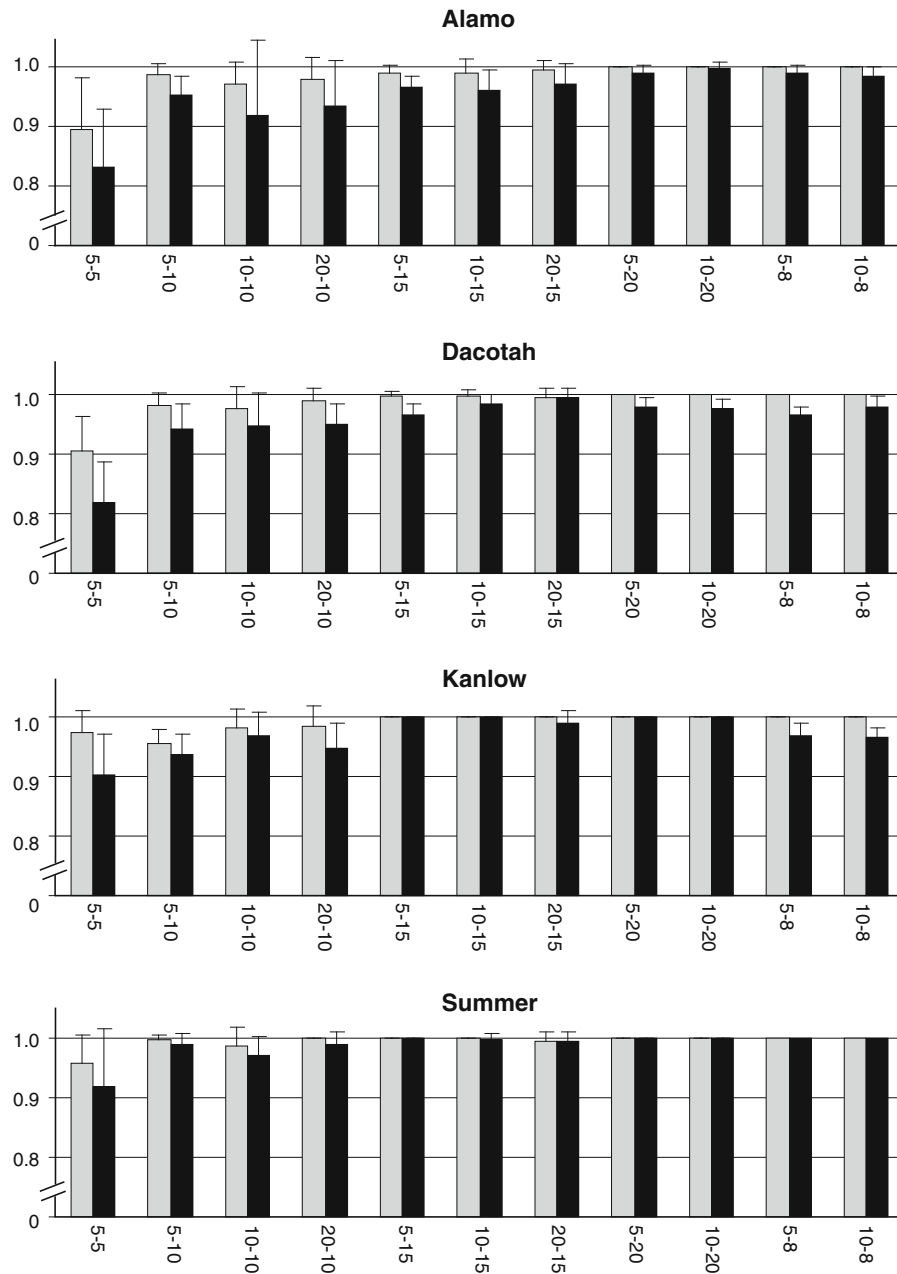


Fig. 4 Bar graphs of the average proportion of individuals with high assignment probability, $q \geq 0.8$ (gray) and $q \geq 0.9$ (black) to source cultivars over ten randomly subsampled data

sets, with various numbers of individuals subsampled followed by the number of loci, separated by a dash character on the x-axis

close relationship between upland tetraploids and octoploids (Fig. 1 and Table 4) suggests that auto-polyploidization occurred within the ancestral upland tetraploids. Within-cultivar genetic diversity was greater in octoploids than tetraploids, as expected from polysomic inheritance (Moody et al. 1993). In

addition, polysomic inheritance slows genetic drift (Moody et al. 1993) and consequently limits genetic differentiation between populations under restricted gene flow. Thus, polysomic inheritance in octoploids may have resulted in octoploids not clustering together in the Neighbor-joining dendrogram of

Table 6 EST-SSR loci ranked according to the discriminatory power within four tetraploid switchgrass cultivars using WHICHLOCI version 1.0 (Banks et al. 2003) at assignment stringency of LOD 5.0 and minimum of assignment probability of 0.9 with A, the total number of alleles per locus detected over all samples and the microsatellite repeat motif (di, dinucleotide; tri, trinucleotide)

Locus	Rank	A	Repeat motif
sww3005	1	41	di
sww2068	2	53	di
sww2260	3	28	tri
sww2558	4	39	di
sww1611	5	6	tri
sww88	6	55	di
sww1729	7	26	tri
sww2079	8	15	tri
sww2376	9	12	tri
sww3027	10	32	di
sww659	11	27	di
sww2980	12	38	di
sww2165	13	13	tri
sww3046	14	20	di
sww2175	15	8	tri
sww1918	16	11	tri
sww2059	17	43	di
sww545	18	9	tri
sww2036	19	13	tri
sww2024	20	6	tri
sww1394	21	11	tri

populations, although multiple origins of the octoploids (Soltis and Soltis 1999) are also a possible cause. Also, slowed genetic drift and greater retention of allelic diversity may have lead to the substantially smaller number of private alleles in the upland tetraploid cultivars, ‘Dacotah’ and ‘Summer’.

An exception to the expected greater diversity under polysomic in comparison to disomic inheritance was detected among the cultivars sampled in this study, suggesting that population history besides polyploidy may have influenced the genetic diversity (García-Verdugo et al. 2009) in some cultivars. Switchgrass cultivars are not removed from the native stands by many generations (Alderson et al. 1995), and cultivars appear to be representative of native populations based on neutral molecular markers (Casler et al. 2007). Thus, genetic diversity within and between switchgrass cultivars is shaped by

breeding history including hybridization and demographic bottlenecks resulting from selection or by population history in the native stands from which they originated. The diversity estimates in ‘Alamo’ were comparable to two octoploid cultivars, ‘Cave-in-Rock’ and ‘Shawnee’, ‘Shawnee’ was derived from ‘Cave-in-Rock’, so the similar level of genetic diversity is concordant with the breeding history, but a demographic bottleneck that would reduce genetic diversity is not apparent in the recorded breeding history for ‘Cave-in-Rock’ (Alderson et al. 1995). Similarly, factors that might contribute to high genetic diversity are absent in the breeding history of ‘Alamo’. Factors affecting genetic diversity in the native stands of switchgrass, especially the climatic fluctuations in the Pleistocene that likely shaped the distribution of the genetic diversity in many species native to North America (Hewitt 2000; Soltis et al. 2006) may be important in understanding the distribution of genetic diversity in switchgrass.

Greater differentiation between ecotypes than ploidy levels in switchgrass was detected in this study (Table 4, Fig. 1) as in other studies of molecular diversity in switchgrass (Gunter et al. 1996; Hultquist et al. 1996; Missaoui et al. 2006; Narasimhamoorthy et al. 2008). The upland and lowland ecotypes are also morphologically distinct but co-occur in southern portions of the tall grass prairie, including Oklahoma (Porter 1966). The co-occurring ecotypes remain reproductively isolated, and the upland ecotype in Oklahoma is more similar to northern upland populations than to the co-occurring lowland plants morphologically (Brunken and Estes 1975) and genetically (Fig. 1). The sympatry of genetically distant populations is consistent with the hypothesis by McMillan (1959) that a southern area of the distribution of switchgrass such as southern Oklahoma and northern Texas was one of the three refugia to which switchgrass populations in surrounding communities retreated during glaciation. Also, the cultivars with origins in southern states tended to have higher genetic diversity and greater number of private alleles in comparison to the cultivars from northern states (Table 3), consistent with the hypothesis of a southern refugium from which recolonization likely occurred. However, the number of cultivars sampled in this study is not adequate to assess the influence of Pleistocene oscillations on switchgrass populations.

Further studies involving remnant native populations over the distributional range of switchgrass would likely contribute to the understanding of the biogeographical and evolutionary factors that have shaped the distribution of genetic diversity in switchgrass and to effective germplasm conservation.

Successful identification of cultivars using neutral molecular markers depends on whether the cultivars are sufficiently distinct. Consistent with the outcrossing breeding system (Hamrick and Godt 1989), substantially greater within than between cultivar diversity was detected (Table 4). Accordingly, none of the private alleles was fixed within a cultivar. However, significant genetic differentiation was indicated by overall F_{ST} and pairwise F_{ST} (Tables 4, 5). On the other hand, model-based Bayesian clustering of individuals indicated that the octoploid cultivars with shared breeding history and some cultivars with similar geographical origins are not sufficiently distinct to be identified based on the markers used (Fig. 3). In contrast, cultivars with shared breeding histories were identified with high accuracies in outbreeding perennial ryegrass (*Lolium perenne* L.) using comparable numbers of SSR loci and individuals per cultivar (Kubik et al. 2001; Wang et al. 2009). The contrasting results of the success of cultivar identification may be attributed to the slowed genetic drift under polysomic inheritance (Moody et al. 1993) in octoploid switchgrass cultivars as opposed to the diploid perennial ryegrass cultivars. In switchgrass, therefore, octoploid cultivars with shared breeding history or geographical origins may be especially difficult to identify from each other, but data for closely related tetraploid cultivars are not available for comparison.

Feasibility of cultivar identification using neutral markers also depends on whether the genetic distinctness between cultivars can be detected with the sampled marker loci and individuals. The sampling effect of the number of loci and individuals per cultivar was analyzed using the tetraploid cultivars, which may be identified based on the marker variation at the 21 marker loci. The analysis indicated that increasing the number of loci and individuals improve the accuracy of cultivar identification and that increasing the number of loci rather than individuals sampled per cultivar may be more effective (Fig. 4). The 21 loci ranked with respect to the discriminatory power indicated more variable

loci tended to be more effective for cultivar identification (Table 6), but selection of loci based on empirical studies likely maximizes assignment accuracy (Banks et al. 2003). Use of loci with greater discriminating power might improve cultivar identification in switchgrass. Based on the analysis by the program WHICHLOCI, the four tetraploid cultivars can be discriminated with only eight loci. The eight loci were effective at tetraploid cultivar identification as few as five individuals sampled per cultivar. The 21 marker loci and the eight loci with greatest discriminatory power identified in this study should be useful for characterization of switchgrass germplasm and identifying population structure for association studies.

Acknowledgments We thank Melanie Harrison-Dunn and Gautam Sarath for critical review of the manuscript. The US Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.

References

- Alderson J, Sharp WC, Hanson AA (1995) Grass varieties in the United States. CRC Press, Boca Raton
- Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? *Bioinformatics* 19:1436–1438
- Barker RE, Warnke SE (2001) Application of molecular markers to genetic diversity and identity in forage crops. In: Spangeberg G (ed) *Molecular breeding of forage crops*. Kluwer Academic Publishers, Dordrecht, pp 135–148
- Barnett FL, Carver RF (1967) Meiosis and pollen stainability in switchgrass, *Panicum virgatum* L. *Crop Sci* 7:301–304
- Brunken JN, Estes JR (1975) Cytological and morphological variation in *Panicum virgatum* L. *The Southwest Nat* 19:379–385
- Casler MD, Stendal CA, Kapich L, Vogel KP (2007) Genetic diversity, plant adaptation regions, and gene pools for switchgrass. *Crop Sci* 47:2261–2273
- Catalán P, Segarra-Moragues JG, Palop-Esteban M, Moreno C, Gonzalez-Candelas F (2006) A Bayesian approach for discriminating among alternative inheritance hypothesis in plant polyploids: the allotetraploid origin of genus *Borderea* (Dioscoreaceae). *Genetics* 172:1939–1953
- Chen DH, Ronald P (1999) A rapid DNA miniprep method suitable for AFLP and other PCR applications. *Plant Mol Biol Rep* 17:53–57

- Esselink GD, Nybom H, Vosman B (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. *Theor Appl Genet* 109:402–408
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. Available at: <http://evolution.genetics.washington.edu/phylip.html>
- García-Verdugo C, Fay MF, Granado-Yela C, Rubio de Casas R, Balaguer L, Besnard G, Vargas P (2009) Genetic diversity and differentiation processes in the ploidy series of *Olea europaea* L.: a multiscale approach from subspecies to insular populations. *Mol Ecol* 18:454–467
- Gunter LE, Tuskan GA, Wullschlegel SD (1996) Diversity among populations of switchgrass based on RAPD markers. *Crop Sci* 36:1017–1022
- Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) *Plant population genetics breeding and genetic resources*. Sinauer Associates, Sunderland, pp 43–63
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2:618–620
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913
- Hitchcock AS, Chase A (1950) *Manual of the grasses of the United States*. US Department of Agriculture Miscellaneous Publication 200, Washington
- Hopkins AA, Taliaferro CM, Murphy CD, Christian D (1996) Chromosome number and nuclear DNA content of several switchgrass populations. *Crop Sci* 36:1192–1195
- Hultquist SJ, Vogel KP, Lee DJ, Arumuganathan K, Kaeppler S (1996) Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, *Panicum virgatum* L. *Crop Sci* 36:1049–1052
- Hultquist SJ, Vogel KP, Lee D, Arumuganathan K, Kaeppler S (1997) DNA content and chloroplast DNA polymorphisms among switchgrasses from remnant Midwestern prairies. *Crop Sci* 37:595–598
- Jarne P, David P (2008) Quantifying inbreeding in natural populations of hermaphroditic organisms. *Heredity* 100:431–439
- Kubik C, Sawkins M, Meyer WA, Gaut BS (2001) Genetic diversity in seven perennial ryegrass (*Lolium perenne* L.) cultivars based SSR markers. *Crop Sci* 41:1565–1572
- Le Comber SC, Ainouche ML, Kovarik A, Leitch AR (2010) Making a functional diploid: from polysomic to disomic inheritance. *New Phytol* 186:113–122
- Lewandowski I, Scurlock JMO, Lindvall E, Christou M (2003) The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass Bioenerg* 25:335–361
- Martinez-Reyna JM, Vogel KP (2002) Incompatibility systems in switchgrass. *Crop Sci* 42:1800–1805
- Martinez-Reyna JM, Vogel KP, Caha C, Lee DJ (2001) Meiotic stability, chloroplast DNA polymorphisms, and morphological traits of upland x lowland switchgrass reciprocal hybrids. *Crop Sci* 41:1579–1583
- McLaughlin SB, Kszos LA (2005) Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenerg* 28:515–535
- McMillan C (1959) The role of ecotype variation in the distribution of the central grassland of North America. *Ecol Monogr* 29:285–308
- McMillan C, Weiler J (1959) Cytogeography of *Panicum virgatum* in central North America. *Am J Bot* 46:590–593
- 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
- 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
- Moody ME, Mueller LD, Soltis DE (1993) Genetic variation and random genetic drift in autotetraploid populations. *Genetics* 134:649–657
- Narasimhamoorthy B, Saha MC, Swaller T, Bouton JH (2008) Genetic diversity in switchgrass collections assessed by EST-SSR markers. *Bioenerg Res* 1:136–146
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10
- Nielsen E (1944) Analysis of variation in *Panicum virgatum*. *J Ag Res* 69:327–353
- Obbard DJ, Harris SA, Pannell JR (2006) Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity* 97:296–303
- Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, Tobias CM (2010) Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing, and multilocus interactions. *Genetics* 185:745–760
- Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. *New Phytol* 186:5–17
- Porter CL (1966) An analysis of variation between upland and lowland switchgrass, *Panicum virgatum* L., in central Oklahoma. *Ecology* 47:980–992
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. *Annu Rev Ecol Syst* 33:589–639
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. *Trends Ecol Evol* 14:348–352

- Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Mol Ecol* 15:4261–4293
- Stebbins GL (1947) Types of polyploids: their classification and significance. *Adv Genet* 1:403–429
- Tobias C, Hayden D, 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 et al (2008) Comparative Genomics in Switchgrass Using 61,585 High-Quality Expressed Sequence Tags. *Plant Genome* 1:111–124
- Vogel KP, Jung HJG (2001) Genetic modification of herbaceous plants for feed and fuel. *Crit Rev Plant Sci* 20:15–49
- Vogel KP, Haskins FA, Gorz HJ, Anderson BA, Ward JK (1991) Registration of ‘Trailblazer’ switchgrass. *Crop Sci* 31:1388
- Vogel KP, Hopkins AA, Moore KJ, Johnson KD, Carlson IT (1996) Registration of ‘Shawnee’ switchgrass. *Crop Sci* 36:1713
- Wang J, Dobrowolski MP, Cogan NOI, Forster JW, Smith KF (2009) Assignment of individual genotypes to specific forage cultivars of perennial ryegrass based SSR markers. *Crop Sci* 49:49–58
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* 6:1358–1370
- Wolfe KH (2001) Yesterday’s polyploids and the mystery of diploidization. *Nat Rev Genet* 2:333–341

Table S1 Bayesian analysis to discriminate between disomic and tetrasomic inheritance (Catalán et al. 2006). Marginal probability of data obtained for the alternative hypotheses of disomic inheritance (D) and tetrasomic inheritance (T) and Bayes factor [B, 2 ×(D-T)] for each locus in each of the tetraploid cultivars

Locus	Alamo			Dacotah			Kanlow			Summer		
	D	T	B	D	T	B	D	T	B	D	T	B
sww88	-133.15	-308.59	350.88	-86.63	-246.19	319.12	-110.51	-289.62	358.22	-49.05	-160.42	222.75
sww545	-28.12	-114.88	173.54	-17.67	-60.14	84.94	-17.02	-66.25	98.47	-11.09	-42.86	63.53
sww659	-60.72	-187.05	252.67	-22.67	-88.94	132.53	-41.40	-145.57	208.33	-15.02	-61.31	92.58
sww1394	-33.16	-121.65	176.97	-7.34	-25.43	36.18	-13.94	-49.62	71.35	0.00	0.00	0.00
sww1611	-10.19	-43.31	66.23	-4.72	-22.07	34.69	-6.73	-18.64	23.82	-3.95	-13.78	19.67
sww1729	-73.61	-216.85	286.49	-39.80	-134.06	188.53	-26.92	-88.80	123.75	-11.07	-45.69	69.23
sww1918	-34.76	-136.23	202.94	-13.46	-64.04	101.16	-15.65	-66.32	101.34	-15.86	-44.55	57.38
sww2024	-13.22	-60.85	95.26	0.00	0.00	0.00	-15.59	-64.65	98.13	-7.22	-41.43	68.43
sww2036	-23.68	-87.38	127.42	-12.14	-47.88	71.47	-21.77	-82.20	120.86	-12.26	-342.28	660.04
sww2059	-53.00	-131.68	157.36	-85.02	-200.30	230.55	-54.23	-221.58	334.68	-25.03	-54.14	58.21
sww2068	-32.76	-114.93	164.35	-36.18	-105.49	138.62	-4.41	-19.31	29.81	-34.10	-131.12	194.03
sww2079	-21.76	-74.06	104.60	-10.82	-39.92	58.20	-13.68	-50.48	73.62	-10.84	-43.42	65.16
sww2165	-12.75	-67.65	109.80	-3.45	-8.06	9.21	-19.51	-81.35	123.69	0.00	0.00	0.00
sww2175	-69.67	-182.74	226.14	-54.82	-176.05	242.48	-26.28	-95.72	138.88	-19.30	-90.53	142.46
sww2260	-48.95	-167.16	236.42	-9.58	-21.99	24.83	-25.72	-85.09	118.74	-7.51	-33.73	52.45
sww2376	-125.17	-320.95	391.56	-74.54	-219.61	290.14	-147.95	-284.42	272.93	-36.26	-136.53	200.54
sww2558	-66.58	-203.65	274.14	-44.84	-143.75	197.82	-61.96	-192.36	260.79	-40.02	-137.33	194.62
sww2980	-100.31	-276.00	351.40	-70.15	-206.61	272.91	-63.77	-175.40	223.26	-41.44	-181.82	280.76
sww3005	-115.80	-305.71	379.81	-28.74	-94.06	130.64	-86.28	-247.00	321.44	-30.23	-116.59	172.72
sww3027	-109.20	-285.10	351.80	-40.04	-134.46	188.84	-56.63	-180.44	247.62	-30.42	-94.44	128.04
sww3046	-12.07	-47.10	70.05	-20.09	-88.87	137.56	-16.31	-68.33	104.04	-19.36	-83.17	127.60

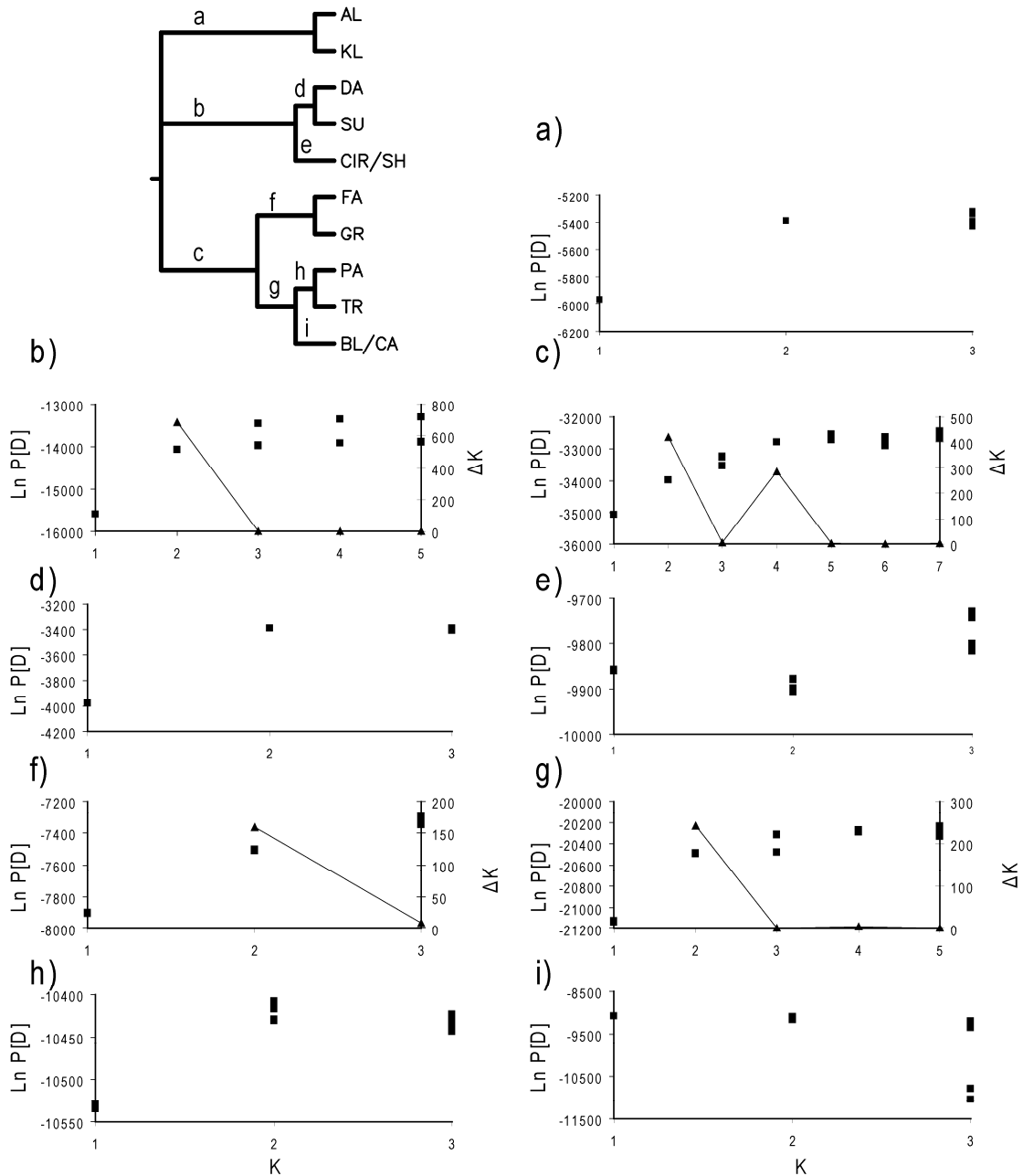


Fig. S1 A dendrogram summarizing hierarchical analyses (Evanno *et al.* 2005) subsequent to the initial analysis of 12 switchgrass cultivars and $q \geq 0.9$ as the threshold (STRUCTURE, Pritchard *et al.* 2000). The plots corresponding to the labeled nodes of the dendrogram show the log likelihood of the multilocus genotypic data, $\ln P(D)$, on the left y-axis for five runs at each value of K , and the second order rate of change in $\ln P(D)$, ΔK , as a function of the number of clusters, K , on the right y-axis if there was no clear maximum $\ln P(D)$

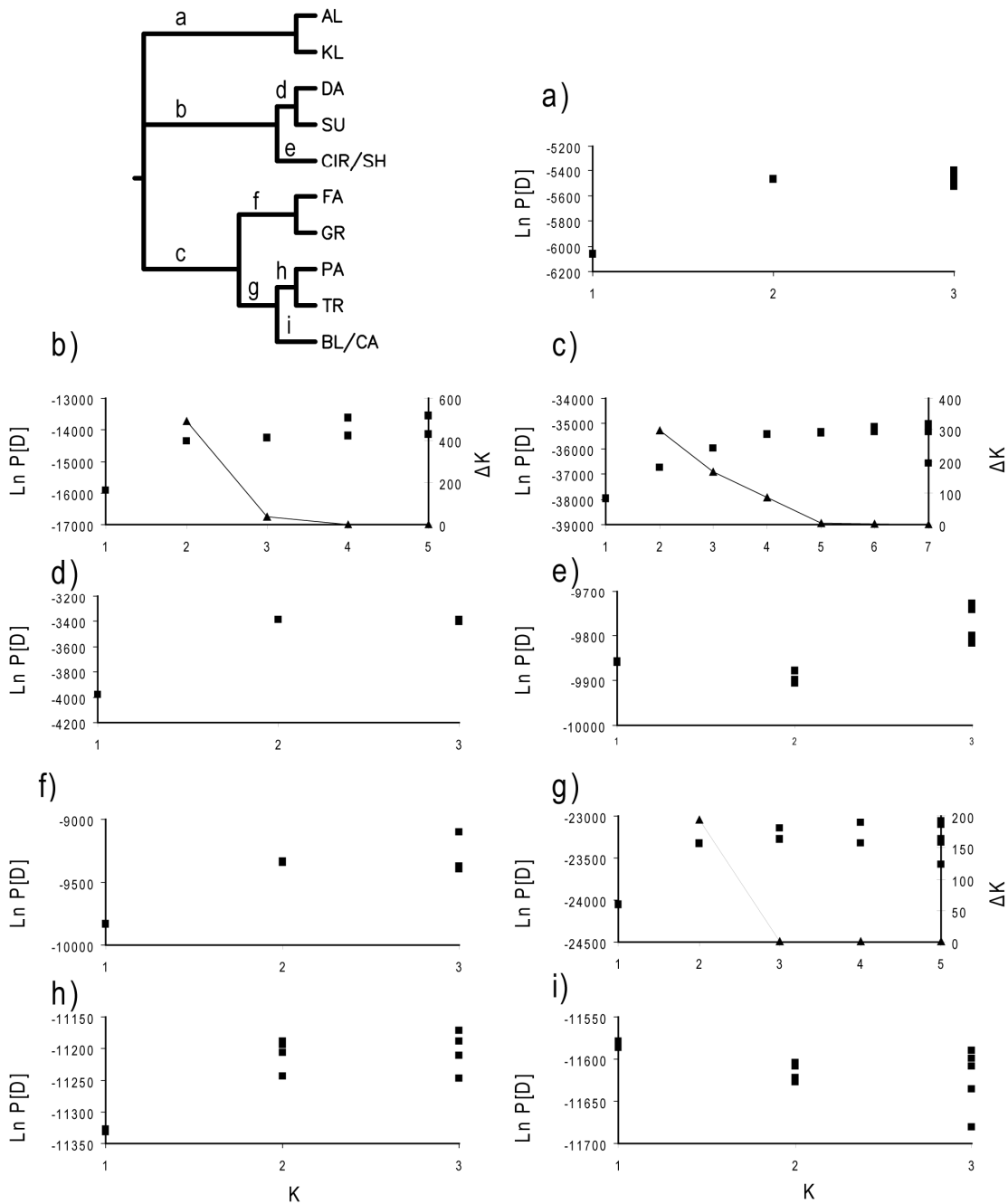


Fig. S2 A dendrogram summarizing hierarchical analyses (Evanno *et al.* 2005) subsequent to the initial analysis of 12 switchgrass cultivars using and $q \geq 0.8$ as the threshold (STRUCTURE, Pritchard *et al.* 2000). The plots corresponding to the labeled nodes of the dendrogram show the log likelihood of the multilocus genotypic data, $\ln P(D)$, on the left y-axis for five runs at each value of K , and the second order rate of change in $\ln P(D)$, ΔK , as a function of the number of clusters, K , on the right y-axis if there was no clear maximum $\ln P(D)$