Post-glacial evolution of *Panicum virgatum*: centers of diversity and gene pools revealed by SSR markers and cpDNA sequences

**Yunwei Zhang · Juan E. Zalapa · Andrew R. Jakubowski ·
David L. Price · Ananta Acharya · Yanling Wei ·
E. Charles Brummer · Shawn M. Kaeppler · Michael D. Casler**

Abstract Switchgrass (*Panicum virgatum*), a central and Eastern USA native, is highly valued as a component in tallgrass prairie and savanna restoration and conservation projects and a potential bioenergy feedstock. The purpose of this study was to identify regional diversity, gene pools, and centers-of-diversity of switchgrass to gain an understanding of its post-glacial evolution and to identify both the geographic range and potential overlap between functional gene pools. We sampled a total of 384 genotypes from 49 accessions that included the three main taxonomic groups of switchgrass (lowland 4x, upland 4x, and upland 8x) along with one accession possessing an intermediate phenotype. We identified primary centers of diversity for switchgrass in the eastern and western Gulf Coast regions. Migration, drift, and selection have led to adaptive radiation in switchgrass, creating regional gene pools within each of the main taxa. We estimate that both upland-lowland divergence and 4x-to-8x polyploidization within switchgrass began approximately 1.5–1 M ybp and that subsequent ice age cycles have resulted in gene flow between ecotype lineages and between ploidy levels. Gene flow has resulted in "hot spots" of genetic diversity in the southeastern USA and along the Atlantic Seaboard.

Keywords Switchgrass · DNA markers · Genetic diversity · Genetic structure · Post-glacial migration

The authors Yunwei Zhang, Juan E. Zalapa and Andrew R. Jakubowski contributed equally to the work described in this manuscript.

Y. Zhang
Grassland Institute, China Agricultural University, Beijing, China

J. E. Zalapa (✉)
Department of Horticulture, USDA-ARS, Vegetable Crops Research Unit, University of Wisconsin, 1575 Linden Dr., Madison, WI 53706, USA
e-mail: Juan.Zalapa@ars.usda.gov

J. E. Zalapa · S. M. Kaeppler · M. D. Casler
DOE Great Lakes Bioenergy Research Center, Madison, WI, USA

A. R. Jakubowski · D. L. Price
Department of Agronomy, University of Wisconsin, Madison, WI, USA

A. Acharya · Y. Wei · E. C. Brummer
Crop and Soil Science Department, Institute for Plant Breeding, Genetics, and Genomics, University of Georgia, Athens, GA, USA

A. Acharya · Y. Wei · E. C. Brummer
DOE BioEnergy Sciences Center, Athens, GA, USA

E. C. Brummer
The Samuel Roberts Noble Foundation, Ardmore, OK, USA

S. M. Kaeppler
Department of Agronomy, University of Wisconsin, Madison, WI, USA

M. D. Casler
USDA-ARS, U.S. Dairy Forage Research Center, Madison, WI, USA

Introduction

"The variable climatic values, different yearly dates of last killing frost in spring, yearly deviations in the precipitation pattern, and fluctuations in the yearly temperature patterns play a vital role in the selection of grassland vegetation. These deviations and fluctuations express the environmental amplitude which an individual must tolerate for continued existence, not necessarily involving reproductive processes, within a grassland community. The marked degree of variability within a species-population in grassland vegetation is a visible response to the highly variable nature of the habitat" McMillan (1959).

Switchgrass is native to the tallgrass prairie and associated ecosystems of the central and eastern USA. It is adapted to a wide range of soils, climatic conditions, and habitats east of the Rocky Mountains of the USA, northern Mexico, and southern Canada (Vogel 2004). Switchgrass has survived agricultural and urban development in thousands of prairie remnants from the Rocky Mountains to the Atlantic Seaboard, from Nova Scotia to Florida, and from Saskatchewan to New Mexico (Stubbendieck et al. 1991). These prairie remnants collectively represent an extremely rich and valuable in situ germplasm resource for switchgrass, as well as many other species.

Upland and lowland ecotypes represent the main taxonomic division within switchgrass (Vogel 2004). The phenotypes of upland and lowland ecotypes are strikingly different: lowland plants are taller in height, later in heading, have fewer tillers per plant, larger tiller diameter, and a much more bluish color than upland plants (Casler 2005; Cortese et al. 2010; Vogel 2004). Upland ecotypes tend to be more northern adapted, generally ranging from USDA Hardiness Zone 3 to 8, while lowland ecotypes tend to be more southern adapted, generally ranging from USDA Hardiness Zone 5 to 10 (Casler 2005; Vogel 2004).

All lowland accessions and cultivars evaluated to date are tetraploids with $2n = 4x = 36$ chromosomes (Vogel 2004). Upland plants ranging from $2n = 2x = 18$ to $2n = 12x = 108$ chromosomes have been reported (Barkworth et al. 2007). Octoploid (8x) is the most common ploidy level for both wild accessions and cultivars, followed by tetraploid (4x). Diploid (2x) forms are extremely rare and have only recently been re-discovered following intensive screening under highly controlled conditions (Young et al. 2010; Price et al. 2010, unpublished data). Hexaploid (6x) switchgrass plants have been frequently reported in the literature, but have not been confirmed in recent studies (Costich et al. 2010). Aneuploidy, deviation of the somatic chromosome number from the expected number for a given ploidy level, is very common in switchgrass, especially the higher ploidy levels (Costich et al. 2010).

Several marker diversity studies have identified significant patterns of DNA polymorphisms in switchgrass. These studies consistently illustrate significant differentiation between upland and lowland ecotypes, based on both nuclear DNA markers and cpDNA sequences (Cortese et al. 2010; Gunter et al. 1996; Missaoui et al. 2006; Narasimhamoorthy et al. 2008; Zalapa et al. 2011). More recently, we showed the first case of clear genetic differentiation among switchgrass ploidy levels using molecular markers (Zalapa et al. 2011). However, genetic studies showing geographic differentiation are lacking or inconsistent (Missaoui et al. 2006; Narasimhamoorthy et al. 2008; Cortese et al. 2010).

Switchgrass is thought to have survived repeated glaciation events in three refugia—a western montane refuge in the southwestern USA and Mexico, a Western Gulf Coast refuge, and an Eastern Gulf Coast refuge (McMillan 1959). Although little is known about genetic diversity, spatial differentiation, genetic isolation, or post-glacial migratory patterns of switchgrass, a general pattern has emerged, particularly with reference to woody species. Fossil pollen records from lakebed sediments have revealed a pattern of post-glacial northward migrations for dominant species of boreal and temperate forest ecosystems as well as prairie and savanna ecosystems (Clark et al. 2001; Kelley et al. 2006; Kneller and Peteet 1999; Soltis et al. 1997). Genetic diversity decreases markedly for both plant and animal species on a north–south axis in both North America and Europe, suggesting that post-glacial founder effects have significantly reduced genetic variation in northern populations of species that have a large latitudinal distribution (Hewitt 1996, 2000; Soltis et al. 1997).

Recently, we have identified patterns of DNA polymorphisms related to the geographic origin of switchgrass accessions using a selected set of SSR primers (Zalapa et al. 2011). We hypothesized that such geographic differentiation of switchgrass accessions may be related to ecological differentiation due to differential climate, daylength, or habitat and perhaps account for post-glacial migrations of switchgrass. The broad objective of the present study was to confirm the existence of switchgrass gene pools by studying remnant native populations sampled across the historical range of switchgrass in North America. Specifically, we studied switchgrass to (1) inventory the natural variability of species across its range; (2) confirm patterns of DNA diversity associated with ecotype, ploidy, and origin; (3) identify primary and secondary centers of origin of switchgrass, and (4) develop a system to classify switchgrass cultivars and natural populations into function regional gene pools for the purposes of conservation, restoration, and genetic improvement. We define gene pools as broad collections of remnant populations that share phenotypic and genotypic characteristics

with many of their neighboring populations across a geographic region.

Materials and methods

Plant materials

This study used a total 384 individuals representing 30 bred cultivars or prairie-remnant collections representing the upland ecotype (n = 226), 18 prairie-remnant collections representing the lowland ecotype (n = 151), and one collection of an intermediate/lowland phenotype (n = 7) (Table 1). Data from the eighteen cultivars used by Zalapa et al. (2011) were used herein to represent a broad range of diversity in cultivars and for direct comparison with our geographically broad sample of prairie-remnant populations (Table 1). The cultivars used can be considered highly representative of natural populations from their respective regions since they are simple seed increases of prairie-remnant populations or have undergone only one or two generations of selection, not sufficient to change their region of adaptation (Zalapa et al. 2011).

The samples from prairie-remnant populations are part of current genomics efforts at USDA-ARS in Madison, Wisconsin and at the University of Georgia in Athens, Georgia. These samples represent a survey of germplasm from throughout the species range and make up a part of two switchgrass association panels that will be used to detect quantitative trait loci for use in breeding improved cultivars for biofuel feedstock. Individual plants and/or populations within the panels were classified as upland or lowland by field inspection of phenotypic traits. The panels are preserved in field nurseries in Wisconsin and Georgia with the Wisconsin panel containing mostly upland ecotypes and the Georgia panel containing mostly lowland ecotypes.

DNA isolation and PCR

Total genomic DNA was isolated from approximately 0.5 cm^2 of leaf tissue using a DNeasy kit (QIAGEN, Valencia, CA, USA). We randomly selected 19 SSR loci from the Zalapa et al. (2011; Table 1) to test their discriminatory power. Laboratory procedures were described in detail by Zalapa et al. (2011). Briefly, polymerase chain reactions (PCR) were performed in 8 µl total volume using 3.5 µl 1× JumpStart REDTaq ReadyMix (Sigma, St. Louis, MO, USA), 2 µl 5 ng/µl genomic DNA, 1.25 µl of H$_2$O, 0.5 µl 5 µM M13-FAM/HEX primer, 0.5 µl 5 µM reverse/ 0.5 µM forward primer, 0.125 µl 5 M betaine (Sigma, St. Louis, MO, USA), and 0.125 µl 50 mg/ml bovine serum albumin (BSA) (CHIMERx, Milwaukee, WI, USA). Thermocycling conditions consisted of an initial melting step

(94°C for 3 min), followed by 30 cycles of 94°C for 15 s, 55°C for 90 s, and 72°C for 2 min, and a final elongation step (72°C for 20 min), followed by an indefinite soak at 4°C. PCR products combined with 15 µl Hi-Di formamide (Applied Biosystems, Foster City, CA, USA) and 0.5 µl of carboxy-X-rhodamine (ROX) standard (GeneFlo-625 ROX; CHIMERx, Milwaukee, WI, USA). SSR allele genotyping was performed using an ABI 3730 fluorescent sequencer (POP-6 and a 50-cm array; Applied Biosystems, Foster City, CA, USA). Amplified fragments were scored using Gene-Marker Software version 1.5 (SoftGenetics, State College, PA, USA). PCRs were repeated on approximately 10% of the samples, and we obtained 99% repeatability.

Flow cytometry

Flow cytometry was performed on leaves from individual plants prepared using the CyStain PI Absolute P kit (Partec, Swedesboro, NJ, USA) with the modifications described by Zalapa et al. (2011). The mean DNA content per plant cell for each sample was analyzed using ModFit software (Verity Software House, Topsham, ME, USA). Ploidy levels were estimated following Zalapa et al. (2011) procedures for all plants originally identified as the upland ecotype and those identified as having an intermediate or mixed phenotype (Table 1). A small number of individual plants, including all plants with an unexpected ploidy classification, were analyzed independently in a second laboratory using the methods of Costich et al. (2010). The latter group included euploid controls of known chromosome number: 2n = 2x = 18, 2n = 4x = 36, 2n = 6x = 54, and 2n = 8x = 72.

Estimates of genetic parameters

All amplified products from each SSR primer pair were treated as single-locus alleles for the purpose of estimating the following genetic parameters: genetic diversity measures, sample size (n), observed number of amplified fragments, number of amplified fragments with frequency ≥ 5% (Nf ≥ 5%), number of unique amplified fragments, average number of amplified fragments per locus (Na), total number of amplified fragments per individual, and polymorphism information content (PIC), were estimated for each locus and population (Table 2). PIC was calculated as follows: PIC $1 - \sum f_i^2$, where f_i is the frequency of the ith fragment.

cpDNA sequence analysis

A subset of 193 individuals was chosen for cpDNA sequencing. A minimum of three individuals per accession were chosen whenever possible and up to seven individual per accession were chosen for accessions that had not

Table 1 Names, abbreviations, number of individuals, and passport information for 67 switchgrass cultivars or accessions used in DNA marker evaluations

Cultivar or accession	Abbr.	n	Status[a]	Origin	Latitude	Longitude	Phenotype[b]	Seed source[c]
Shawnee	SE	4	Bred	IL	37.47	88.17	Upland	USDA-ARS
Summer	SU	20	Bred	NE	40.68	95.86	Upland	SDCIA
Sunburst	SB	13	Bred	SD	42.87	97.40	Upland	SDCIA
Trailblazer	TB	8	Bred	NE	39.92	98.07	Upland	USDA-ARS
WS4U	4U	6	Bred	WI	43.30	89.35	Upland	SW776
WSK4	K4	2	Bred	WI	35.33	96.24	Upland	SW1302
Blackwell	BL	28	SIC	OK	35.96	97.07	Upland	PI 421520
Carthage	CT	8	SIC	NC	35.35	79.45	Upland	PI 421138
Cave-in-Rock	CR	24	SIC	IL	37.47	88.17	Upland	PI 469228
Dacotah	DC	20	SIC	ND	46.38	100.94	Upland	NRCS-PMC
Forestburg	FB	17	SIC	SD	44.02	98.10	Upland	NRCS-PMC
KY1625	K16	4	SIC	WV	37.87	81.23	Upland	PI 431575
Pathfinder	PF	13	SIC	KS	39.82	98.48	Upland	USDA-ARS
Shelter	SH	9	SIC	WV	39.40	81.20	Upland	NRCS-PMC
Albany	ALB	4	Wild	NY	42.72	73.83	Upland	ECS-12
Allegheny River	ALG	4	Wild	PA	40.95	79.62	Upland	ECS-10
Camp Dawson	CDW	4	Wild	KY	38.02	82.67	Upland	SW809
Chiwaukee 1	CH1	4	Wild	WI	42.55	87.80	Upland	SW124
Columbiana	COL	4	Wild	OH	40.62	80.70	Upland	SW64
Genesee, NY	GEN	4	Wild	NY	42.99	78.15	Upland	SW63
Ipswich Prairie 2	IP2	4	Wild	WI	42.57	90.40	Upland	SW115
Jackson	JCK	4	Wild	MI	42.25	84.31	Upland	SW43
Morrison	MOR	4	Wild	MN	46.63	93.65	Upland	SW50
NRCS 9084291	U37	4	Wild	MI	42.98	86.06	Upland	NRCS-PMC
Route 72/563	R72	10	Wild	NJ	39.82	74.53	Upland	ECS-1
Staten Island	STI	8	Wild	NY	40.59	74.15	Upland	SW781
Sterling Barrens 3	SB3	4	Wild	WI	45.08	92.83	Upland	SW99
Tipton	TIP	4	Wild	IN	40.28	86.04	Upland	SW31
Toledo	TOL	4	Wild	OH	41.58	83.67	Upland	ECS-2
Wadena	WAD	4	Wild	MN	46.44	95.13	Upland	SW60
Alamo	AL	21	SIC	TX	28.33	98.12	Lowland	PI 420006
Kanlow	KA	23	SIC	OK	35.33	96.24	Lowland	PI 421521
AW-314/MS-155	AMS	7	Wild	AR	35.43	91.84	Lowland	PI 421999
BN-12323-69	L11	7	Wild	KS	38.81	98.27	Lowland	PI 414070
Pangburn	PNG	7	Wild	AR	35.43	91.84	Lowland	PI 414065
Hoffman	HOF	7	Wild	NC	35.03	79.55	Lowland	PI 315723
Hillsborough River S.P.	HRP	7	Wild	FL	28.15	82.24	Lowland	UGA-HRP
Miami	MI	2	Wild	FL	25.54	80.63	Lowland	PI 421901
Oscar Scherer S.P.	OSP	7	Wild	FL	27.19	82.46	Lowland	UGA-OSP
Pasco County	PCF	7	Wild	FL	28.33	82.42	Lowland	UGA-PCF
PI 422016	L19	7	Wild	FL	27.00	81.00	Lowland	PI 422016
PMT-785	PMT	7	Wild	FL	27.00	81.00	Lowland	PI 422003
SG5	SG5	9	Wild	OK	34.50	95.50	Lowland	NRCS-PMC
Stuart	ST	2	Wild	FL	27.20	80.23	Lowland	PI 422001
Sumter National Forest	SNF	7	Wild	SC	34.52	81.57	Lowland	UGA-SNF
T 2086	WIL	7	Wild	NC	34.24	77.94	Lowland	PI 476290
Timber	TI	13	Wild	NC	35.54	79.28	Lowland	NRCS-PMC

Table 1 continued

Cultivar or accession	Abbr.	n	Status[a]	Origin	Latitude	Longitude	Phenotype[b]	Seed source[c]
Wabasso	WB	6	Wild	FL	27.75	80.44	Lowland	PI 422000
Sprewell Bluff	SPB	7	Wild	GA	32.90	84.44	Int/Low	UGA-SPB

[a] Status: Wild = seed harvested on a prairie-remnant population that is likely to represent local germplasm; SIC = source-identified cultivar derived from a random seed increase, without conscious selection and breeding, within a prairie-remnant population; Bred = a product of one or more cycles of selection and breeding

[b] Int/Low = All plants of an intermediate phenotype, not sufficiently clear to warrant a confident phenotypic "call"

[c] Seed sources: USDA-ARS = switchgrass breeding program at Lincoln, NE; SDCIA = South Dakota Crop Improvement Association; NRCS-PMC = NRCS Plant Materials Centers (Bismarck, ND; Rose Lake, MI; Big Flats, NY; Cape May, NJ; Americus, GA; Coffeeville, MS); PI-xxxxxx (NRCS-GRIN; Germplasm Resources Information Network); MSU = Mississippi State University; ECS-xx = Ernst Conservation Seeds, Meadville, PA; SWxxx = seeds collected directly from prairie remnant site and all processing conducted by hand in Madison, WI; and UGA-xxx = seeds collected directly from prairie remnant site and all processing conducted by hand in Athens, GA

Table 2 Number of amplified fragments and polymorphism information content (PIC) for 19 SSR markers used to differentiate switchgrass cultivars and accessions

Locus	Number of amplified fragments	Unique to lowland	Unique to upland	Total unique amplified fragments, upland or lowland (%)	PIC
sww150	13	2	5	54	0.83
sww175	8	3	0	38	0.66
sww185	11	2	0	18	0.78
sww210	11	7	1	73	0.60
sww274	4	1	0	25	0.52
sww651	19	6	7	68	0.75
sww687	10	2	3	50	0.62
sww2309	19	7	7	74	0.82
sww2312	17	4	2	35	0.79
sww2320	8	1	3	50	0.70
sww2341	11	1	3	36	0.57
sww2376	15	1	4	33	0.78
sww2377	11	2	4	55	0.53
sww2385	14	2	6	57	0.85
sww2387	11	5	3	73	0.67
sww2393	21	5	3	38	0.83
sww2394	13	5	4	69	0.77
sww2403	17	9	2	65	0.84
sww2415	19	10	2	63	0.83
Mean	13.3	3.9	3.1	51	0.72

previously been analyzed or for which there was obvious phenotypic variability. Chloroplast DNA was amplified from the following five intergenic regions: trnL(UAA), trnT(UGU)-trnL(UAA) 5′, trnH(GUG)-psbA, psbJ-petA, and atpI-atpH ndhA (Zalapa et al. 2011). Sequencing procedures and thermocycling conditions for PCR were the same as Zalapa et al. (2011). Sequencing reactions were resolved on an ABI 3730 Genetic Analyzer Applied Biosystems, Foster City, CA, USA. Sequences were edited and assembled using Codon Code Aligner v. 3.5 (Codon-Code Corp., Dedham, MA, USA), and aligned using the MUSCLE algorithm (Edgar 2004).

Six switchgrass genotypes were chosen for molecular clock computations, representing four groups: upland octoploid, upland tetraploid, intermediate octoploid, and lowland tetraploid. For comparative purposes, genomic DNA was extracted from the diploid relative *P. hallii* var. *filipes* FIL2 and *P. capillary* var. Norris using a Qiagen DNEasy kit according to the manufacturer's recommendations. This DNA sample was used to prepare a fragment library using standard Roche protocols. One complete sequencing run (1 picotiter plate) was completed on a 454-Titanium platform at the Duke University Genome Sequencing & Analysis Core Facility. Reads were mapped

against the complete chloroplast genome of *P. virgatum* cv. Kanlow (provided by C. Tobias) using the Roche GC Reference Mapper v 2.5pl. The resulting assembly included 69% of the chloroplast genome (96.4 kb) with an average coverage of 34X. The complete cpDNA amplicon sequence was found for four of the five regions: for trnH(GUG)-psbA, only 150 bp were present in the *P. hallii* assembly. *P. hallii* and *P. capillari* sequence work was conducted by E. Meyer and T. Juenger, University of Texas, Austin, TX.

Complete chloroplast genome sequence data was obtained from Genbank for the following species: *Oryza sativa* (GU592207.1), *Zea mays* (NC_001666.2), *Sorghum bicolor* (EF115542.1), and *Saccharum offinarum* (AE009947.2). The five chloroplast regions used in our analysis were aligned to the chloroplast genome of each species using the MUSCLE algorithm (Edgar 2004). Sequence data from the five regions was concatenated using Mesquite 2.74 (Maddison and Maddison 2007).

Phylogenetic analysis of the chloroplast DNA sequences was carried out using a Bayesian MCMC performed in MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003). All analyses were run for one million iterations (sampled every 100th iteration with four chains, two simultaneous runs, and a heated chain temperature of 0.05), which was sufficient to reach stationarity. The initial 250,000 iterations were discarded for burn-in. A Bayesian 50% consensus cladogram with posterior probabilities for each clade was produced from the data using Mesquite v. 2.74 (Maddison and Maddison 2007).

Molecular clock estimates were performed using Bayesian estimates in BEAST 1.6.1 (Drummond and Rambauat 2007). The optimal model to use with the data was determined to be TN93 + Gamma (Tamura and Nei 1993) using the Bayesian Information Criterion as implemented in the program MEGA 5 (Tamura et al. 2011). While most earlier estimates of *Poaceae* divergence timing have used a 50 M ybp estimate for the divergence of *Oryza* and *Zea* (Zhong et al. 2009), recently discovered fossil evidence places this date at a minimum of 65 M ybp (Prasad et al. 2005). We set our prior distribution for divergence time of *Oryza-Zea* using a gamma distribution with an offset of 65 M ybp and a shape parameter of one. We identified the strict clock model as optimal by comparing alternative clock models using a likelihood ratio test. The tree prior was set to Speciation: Yule Process. The MCMC process was run for 10 million iterations, including a 2.5-million-iteration burn-in fraction; parameter estimates were logged every 1,000 iterations. Posterior probabilities for each clade, mean node ages, and 95% highest posterior densities (HPD) were calculated for all divergence time estimates. Only those nodes with support probability > 0.95 were reported.

Population structure

SSR data were scored in a binary format as presence (1) or absence (0) of DNA fragments. An analysis of molecular variance (AMOVA) based on 9,999 permutations was performed on all individuals to account for marker diversity associated with ecotypes, ploidy, populations, and plants (GeneAlEx) 6.4 (Peakall and Smouse 2006). Bayesian clustering algorithms available in the program STRUCTURE (v. 2.2) (Pritchard et al. 2000) were used to infer genetic pool membership for each sample. We ran STRUCTURE using 50,000 Markov chain Monte Carlo iterations with 50,000 burn-in iterations and 10 replicates per run. We used the 'admixture model' in which each individual draws a fraction of its genome from each of K subpopulations and the case of 'no prior population information'. The most likely true value of K was estimated using Bayes' rule as specified in Pritchard et al. (2000) and the DK method proposed by Evanno et al. (2005). An a posteriori AMOVA was conducted to quantify the variation associated with cluster groups determined by STRUCTURE.

Lastly, we tested the existence of founder effects along post-glacial north–south migration routes using regression analysis. We used two measures of SSR product richness to assess founder effects. For each accession, richness estimates were calculated using non-parametric estimates of incidence-based amplified fragment accumulation curves (Colwell and Coddington 1994) using the following metrics implemented in the software EstimateS (Version 8.2, R. K. Colwell, http://purl.oclc.org/estimates): the Chao 2 richness estimator (Chao 1987), the incidence-based coverage estimator of richness (Chao et al. 2000), and the second-order jackknife richness estimator (Burnham and Overton 1978; Palmer 1991). Means and standard deviations of richness estimates were calculated from 10,000 bootstrap replicates of the dataset. Because our data represents a limited sampling of individuals from each accession, with varying numbers of sampled individuals per accession, we estimated the total number of expected amplified fragments had each accession been sampled with a uniform number of individuals per, using n = 3 (Bunge and Fitzpatrick 1993). For each individual, we also computed the number of unique amplified fragments across all primer pairs. All values of SSR product richness were regressed on latitude and longitude, independently for tetraploids and octoploids.

Results

A sample of 384 switchgrass individuals was tested using SSR markers to provide genetic confirmation of ecotype

identity and to define ploidy and genetic pool membership (Table 1). The set of 19 randomly selected SSR primer pairs performed as expected according to Zalapa et al. (2011; Table 2). A total of 252 amplified fragments were detected with an average of 13.3 and a range of 4–21 amplified fragments per primer pair (Table 2). The PIC at each locus ranged from 0.52 to 0.85 with an average of 0.72 PIC over all primer pairs. The lowland (n = 157) ecotypes possessed 187 amplified fragments of which 59 were unique to members of the lowland ecotype (Table 2). Similarly, the upland ecotypes (n = 227) possessed 189 amplified fragments of which 75 were unique to members of the upland ecotype. Separating the two upland ploidy levels, the 4x (n = 75) possessed 137 amplified fragments compared to 176 amplified fragments in the 8x (Table 3). The PIC was 0.56 and 0.65 for the 4x lowland/upland and 8x upland, respectively.

STRUCTURE analyses revealed clear genetic differentiation for upland and lowland ecotypes and ploidy levels (data not presented). When only two populations (K = 2) were assumed, we identified two genetic clusters which corresponded largely to the samples classified a priori by ecotype (data not presented). Similarly, when K = 3 was assumed, our samples were divided into the three main groups: 4x lowland, 4x upland, and 8x upland (data not presented). Available flow cytometry of most samples from the switchgrass cultivars (Zalapa et al. 2011) plus a limited number of flow cytometry preparations from the prairie-remnant samples confirmed the ploidy designations by STRUCTURE.

A total of 193 individuals were successfully sequenced in all five cpDNA regions. The cpDNA analysis was based on five informative sequences, all with at least one polymorphic site. A total of 41 polymorphic sites, including the previously identified 49-bp insertion in the trnL(UAA) intron, were identified within the five intergenic regions (data not presented). Bayesian clustering of cpDNA sequences discriminated the lowland and upland switchgrass ecotypes (Fig. 1). The 49-bp indel was absent in most of the lowland ecotypes, but lowland accessions within the Gulf Coast group sometimes possessed the indel as did all but two plants classified as the upland ecotype. cpDNA sequences did not seperate ploidy levels of the upland ecotype.

Using an estimate of 65 M ybp for divergence of *Oryza* and *Zea*, we estimated divergence times of 22.3 M ybp for *Panicum* and *Zea*, 5.3 M ybp for *P. virgatum* and *P. hallii*, and 1.3 M ybp for the earliest divergence of upland and lowland ecotypes, along with the earliest 4x-to-8x polyploidization event (Fig. 2). We found a total of 21 cpDNA haplotypes in our collection of 193 individuals with complete coverage in these five cpDNA regions. Most of these haplotypes differed from each other by a single nucleotide polymorphism, or single-base-pair deletion or insertion. Two cpDNA haplotypes dominated the sample: 64 individuals (33%) had a largely upland haplotype without polymorphisms, while 79 individuals (41%) had a largely lowland haplotype without polymorphisms. For the first upland-lowland split and the first 4x-to-8x polyploidization event (Fig. 2), we were unable to resolve specific

Table 3 Genetic diversity characteristics of switchgrass ecotypes and genetic pools based on 19 SSR markers

Genotype group	Statistic[a]						
	N	Number of amplified fragments	Nf ≥ 5%	Unique amplified fragments	Na	Na per individual	PIC
All samples	384	252	107	–	13.3	36	0.66
Lowland	157	187	87	59	7.5	32	0.56
Upland	227	189	109	75	7.6	40	0.64
Upland 4x	75	137	85	2	5.5	32	0.56
Upland 8x	152	176	115	27	7.0	44	0.65
Lowland4x A	27	106	81	0	4.2	30	0.56
Lowland4x B	6	90	90	0	3.6	36	0.57
Lowland4x C	22	127	90	5	5.1	33	0.59
Lowland4x D	95	161	73	14	6.4	31	0.59
Lowland4x E	7	87	87	0	3.5	39	0.49
Upland4x A	46	108	77	2	4.3	32	0.54
Upland4x B	2	34	34	0	1.4	34	0.26
Upland4x C	27	98	70	0	3.9	32	0.49
Upland8x A	53	128	98	6	5.1	43	0.63
Upland8x B	99	159	110	15	6.4	44	0.64

[a] N = sample size; Nf ≥ 5% = number of amplified fragments with frequency ≥ 5%; Na = average number of amplified fragments per locus; Na per individual = total number of amplified fragments per individual; PIC = polymorphism information content
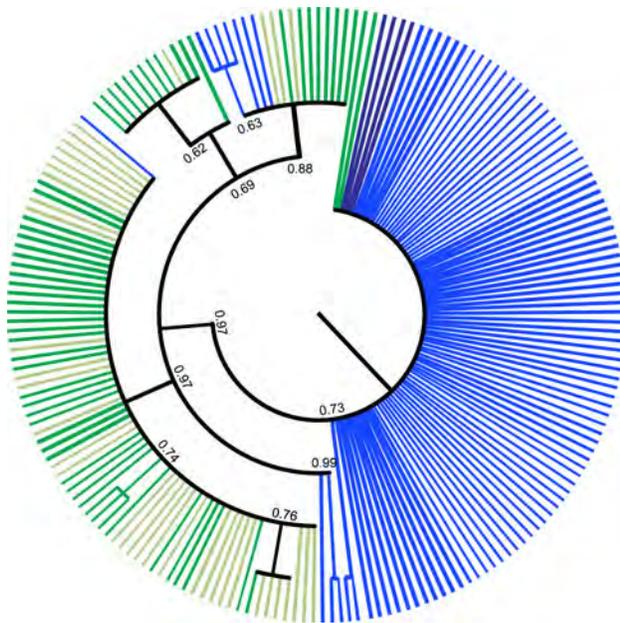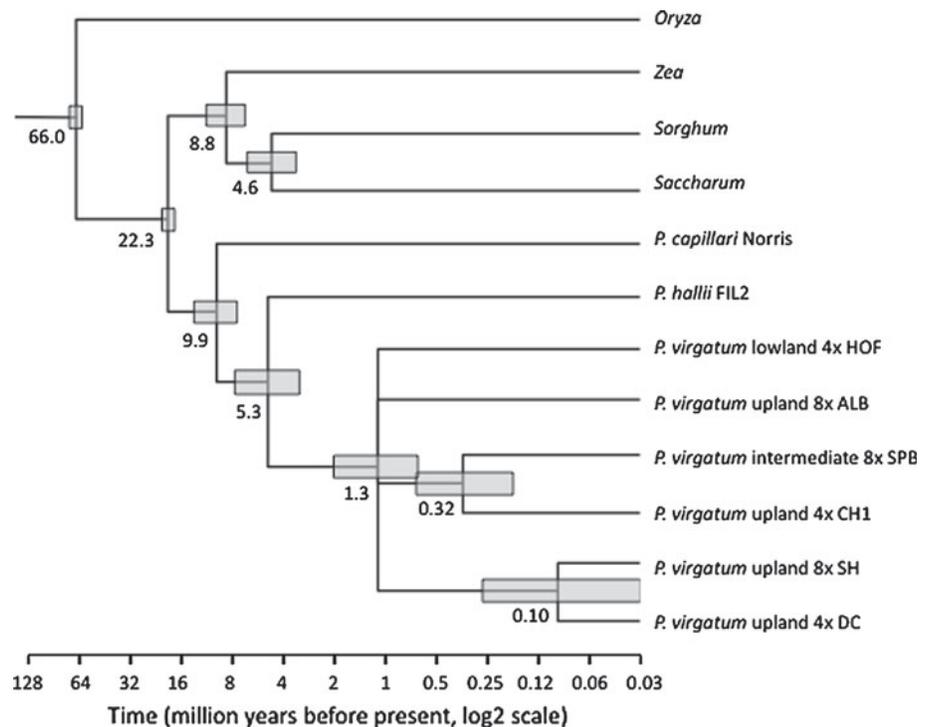
Fig. 1 Bayesian consensus tree of 193 switchgrass individuals phenotypically classified as lowland (*blue*), upland 4x (*light green*), upland 8x (*dark green*), or intermediate (*purple*), based on cpDNA polymorphisms from five regions

divergence times because of the small number of polymorphisms that characterized the remaining 50 individuals. Instead, we found evidence that both events occurred as early as 1.3 M ybp (see the split between upland and lowland and between 4x and 8x in Fig. 2), but have

probably recurred a number of times since the first of these events. We obtained complete support for an upland-intermediate split (defined by both genotype and phenotype) at 0.32 M ybp and a 4x-to-8x polyploidization event as recently as 0.10 M ybp, which appears to be our lower limit of detection using this particular molecular clock.

The most likely true value of K identified by STRUCTURE was K = 10 (Fig. 3). The analysis detected five lowland genetic pools (A–E), three 4x upland (A–C) genetic pools, and two 8x upland (A and B) genetics pools (Fig. 3). Within these three main groups, much of the clustering was associated with geographic origin. Four accessions from North Carolina and Florida had principal membership probabilities in Lowland group A. The Sprewell Bluff accession from Georgia was the only accession with primary membership probability in Lowland group B. Four accessions—one from New Jersey and three from Florida - had primary membership probabilities in Lowland group C. Nine accessions had primary membership probabilities in Lowland group D, spanning from the southern Great Plains to the Atlantic Seaboard, covering a considerable part of the adaption region for lowland ecotypes. Only one accession had a primary membership probability in Lowland group E (PMT from southern Florida).

Several upland accessions, from a wide range of locations across the northern tier of the USA had primary membership probabilities in Upland 4x group A (Fig. 3). One accession had a primary membership probability in Upland 4x group B, the broad-based northern germplasm
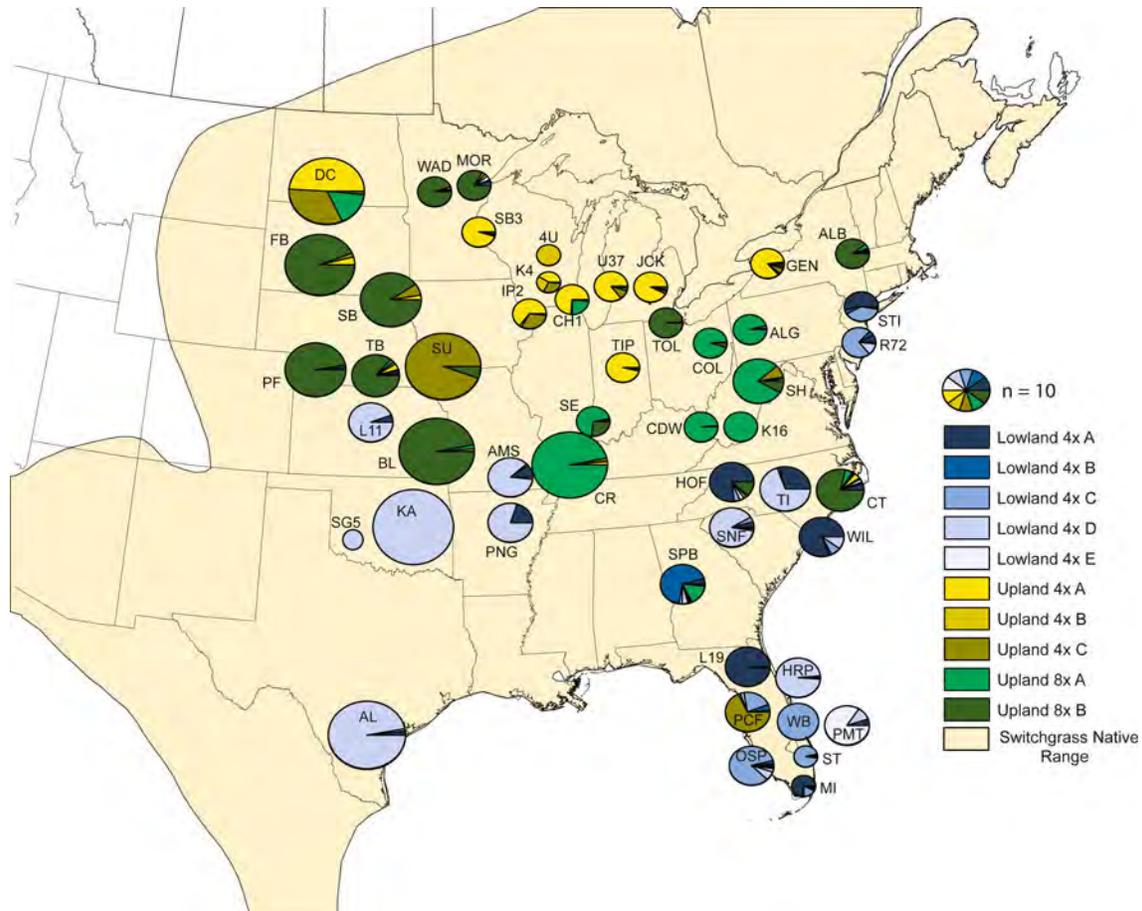
Fig. 2 Dendrogram showing estimated divergence times for six *P. virgatum* genotypes, one genotype of *P. hallii*, and four anchor species. Values shown in the figure are mean estimated divergence times and *grey boxes* are 95% highest posterior probability intervals. Note that the time scale is compressed to a log2 scale to allow visualization of evolutionary events on both small (within *P. virgatum*) and large time scales (among divergent genera). Data for *Oryza, Zea, Sorghum,* and *Saccharum* were obtained from GenBank; data for *P. hallii* were kindly provided by Eli Meyer and Tom Juenger of the University of Texas. All nodes shown are supported with a probability of 1.0. Accession abbreviations from Table 1 are shown for the six *P. virgatum* genotypes

Fig. 3 Partial map of the USA, showing the approximate location of each switchgrass cultivar or accession and the probabilities of accession membership in one of ten groups identified by STRUC-TURE analysis. The size of each *circle* represents the sample size for each accession, with n = 10 shown in the *legend*. Each cultivar or accession is identified by a two- or three-character code from Table 1. Each "slice of pie" represents the probability of membership in one of the 10 STRUCTURE groups shown in the legend

pool WS4U. Two accessions had primary membership probabilities in Upland 4x group C: Summer from Nebraska and the Pasco County accession from Florida. Seven upland accessions had primary membership probabilities in Upland 8x group A, located in the Ohio River Valley and Central Appalachian Mountain region. Lastly, ten upland accessions had primary membership probabilities in Upland 8x group B; seven of these accessions originated in the central and northern Great Plains while the other three originated in Ohio (Toledo), New York (Albany), and North Carolina (Carthage). Values of PIC were highest for the two 8x lineages and similar for all other lineages, except for upland 4x group B, which had only two members (Table 3). Only four genetic pools possessed amplified fragments that were unique to that pool: lowland 4x lineages C and D, upland 4x A, and both upland 8x lineages (5, 14, 2, 6, and 15 unique amplified fragments, respectively).

Analysis by AMOVA revealed that the SSR variation was significant ($P < 0.0001$) among ecotypes and ploidy

accounting for 17% of the total variation (Table 4) and 51% (17/33) of the marker variation among populations. In the AMOVA by genetic pool, the variation ($P < 0.0001$) among the ten genetic pools accounted for 22% of the total variation (Table 4).

SSR product richness, on an individual-plant basis, increased at more northern latitudes, as measured by an increasing number of amplified fragments across 19 SSR primer pairs (Fig. 4). Quadratic regressions were significant for individuals at both ploidy levels, differing only in the mean or intercept value. The shape of both regressions suggested richness peaks at both northern and southern extremes of the switchgrass adaptation range, with the largest peak in the north, just the opposite of expectations based solely on founder effects during post-glacial northward migrations. While there was a considerable amount of noise in these regressions, they accounted for 9–13% of the variability in number of amplified fragments and they both had highly significant P values (Fig. 4). SSR product richness, on an accession basis, was not associated with

Table 4 Analysis of molecular variance (AMOVA) for 50 switchgrass populations based on 19 switchgrass SSR primer pairs

Source of variation	df	SS	Variance component	Percentage of variation	P value
Ecotype and ploidy	2	548.65	4.02	17	<0.0001
Population/ecotype-ploidy	44	45.38	3.78	16	<0.0001
Plants/cultivar/ecotype-ploidy	331	16.07	16.07	67	<0.0001
Genetic pool[a]	9	204.23	5.28	22	<0.0001
Plants/genetic pool	374	18.37	18.37	78	<0.0001

[a] Inferred genetic pools based on STRUCTURE analysis with K = 10
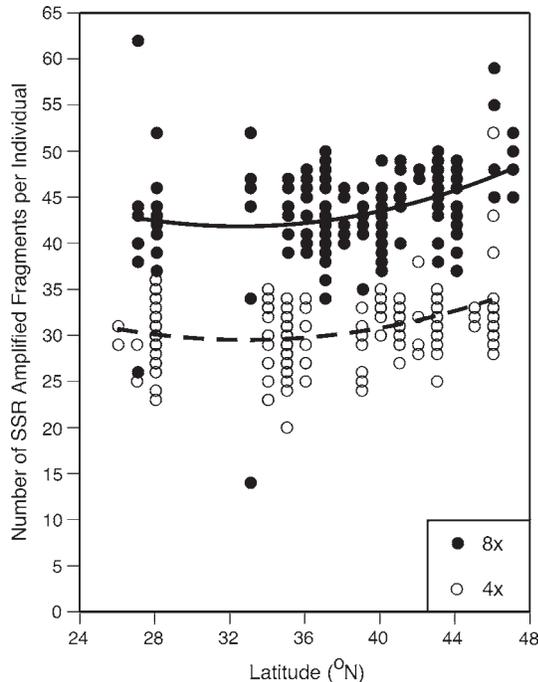
Fig. 4 Number of SSR amplified fragments, across 19 SSR primer pairs, measured on 383 individual switchgrass, plants as a function of latitude-of-origin and ploidy. One hexaploid plant was excluded from this analysis. Quadratic regression equations were: $Y = 73.39 - 1.938X + 0.0298X^2$, $R^2 = 0.09$; $P = 0.0003$ (octoploid) and $Y = 57.04 - 1.675X + 0.0255X^2$, $R^2 = 0.13$; $P < 0.0001$ (tetraploid)

Table 5 Number of individuals classified as upland or lowland switchgrass according to phenotype, cpDNA, or nuclear SSR markers

cpDNA classification	SSR classification	Phenotypic classification	
		Lowland	Upland
Lowland	Lowland	75	5[c]
Lowland	Upland	0	4[d]
Upland	Lowland	13[a]	0
Upland	Upland	4[b]	92

[a] Includes one of seven from PMT-785, four of seven from Sprewell Bluff, two of two from Miami, and six of six from Wabasso

[b] Includes four of seven from Pasco County

[c] Includes one of four from Staten Island and four of four from Route 72/563

[d] Includes one of four from Toledo, two of four from Staten Island, and one of seven from Hoffman

individuals, one each from Toledo and Hoffman, the 2 green bars that clustered with the majority of lowland individuals at the top of Fig. 1) and upland cytotype, lowland genotype, and lowland phenotype (five individuals—four from Sprewell Bluff and one from PMT-785). The presence of upland phenotypes and genotypes within the lowland cytotype and lowland phenotypes and genotypes within the upland cytotype is strong evidence of bidirectional gene flow between upland and lowland lineages.

Discussion

Geographic distribution of ecotypes and lineages

Prior to European settlement of North America, switchgrass could be found across a wide geographic range, from Mexico to Canada and southern Florida to the Canadian Maritime Provinces. The environmental gradients that occur across this region, principally photoperiod, have driven evolutionary forces to create phenotypic differentiation on two levels. The most obvious level of differentiation occurs between lowland and upland ecotypes.

latitude, regardless of the estimate (Chao2, incidence, and jackknife) or ploidy level.

Table 5 summarizes the results for cytotype, genotype, and phenotype of the 193 individuals with complete data for all three classification factors. A total of 167 individuals (86.5%) were unequivocally classified as lowland or upland by all three factors. The remaining 26 individuals fell into four groups, most of which contained individuals from multiple accessions. Of the 26 individuals in these four groups, all but one (from Toledo, OH) originated along the Atlantic Seaboard, from southern Florida to Staten Island, NY. Seven of these 26 individuals were octoploid (8x), belonging to two of the four groups: lowland cytotype, upland genotype, upland phenotype (two

Historical definitions of these two ecotypes derived from phenotypic observations on flowering time, plant height, tiller size and density, panicle size and shape, and presence or absence of bluish waxy bloom on leaf blades and sheaths (Casler 2005; Cortese et al. 2010; Vogel 2004). Lowland ecotypes are found naturally only in USDA Hardiness Zones 5–10, while upland ecotypes are found naturally in USDA Hardiness Zones 3–8. Flowering time can range from early July for extreme northern upland accessions to early October for extreme southern lowland accessions. Flowering time and photoperiod response are probably the most important traits regulating adaptive variation in switchgrass. For most cultivars, adaptation and productivity decrease significantly when the cultivar is grown in more than one hardiness zone north or south of its origin (Casler et al. 2004, 2007b).

More recent applications of DNA markers and sequencing methodologies have suggested that both cpDNA sequence data and nuclear DNA markers are highly correlated with the upland and lowland phenotypic classes. Differentiation between upland and lowland accessions is generally very obvious, representing the first level of clustering or grouping in DNA marker diversity studies (Cortese et al. 2010; Gunter et al. 1996; Narasimhamoorthy et al. 2008). These studies have generally been based on numerous DNA markers scattered throughout the genome, none of which have been singled out as specifically diagnostic or exclusively associated with the lowland-upland phenotypic polymorphism. Zalapa et al. (2011) specifically assayed 501 amplified fragments from 55 SSR primers and found that no single primer pair produced enough genetic variability (i.e., DNA fragments) to consistently and reliably discriminate the lowland or upland phenotypes. However, most likely, there are a very small number of loci responsible for the lowland-upland phenotypic polymorphism and we have yet to identify any of these genetic factors. The domestication of maize (Zea mays L.) from teosinte (Zea mays ssp. parviglumis) was largely due to a very small number of genetic polymorphisms involving both functional genes and regulatory factors (Doebley 2004). Because the phenotypic divergence between upland and lowland switchgrasses is significantly less than that between maize and teosinte, it is most likely due to a very small number of genetic factors.

One 49-bp chloroplast-genome indel has been associated with the lowland-upland phenotypic polymorphism. This sequence is present in a wide range of upland accessions, but absent in Alamo and Kanlow, two lowland accessions from the Southern Great Plains (Missaoui et al. 2006). More recently, we found that this indel is not exclusively associated with the upland phenotype (Zalapa et al. 2011). We queried seven accessions with the lowland phenotype, identifying Miami and Wabasso, both from southern Florida, as possessing this 49-bp sequence within the cpDNA trnL(UAA) intron. Despite this inconsistency in the chloroplast genome, Miami and Wabasso clustered very tightly with another Florida accession (Stuart) based on nuclear SSR markers, suggesting a distinct Florida clade of the lowland ecotype, highly differentiated from a separate Southern Great Plains clade of the lowland ecotype (Zalapa et al. 2011).

Our results have confirmed multiple lineages of switchgrass with distinct geographic patterns within both the upland and lowland ecotypes. Within the lowland ecotype, there is clearly a southern Great Plains lineage that likely derives from a glacial refuge on the coastal plain of the western Gulf Coast. The cultivars Alamo and Kanlow, as well as several other lowland accessions, are typical members of this lineage. Three eastern lowland accessions (Sumter National Forest, Timber, and Hillsborough River State Park) had high membership probabilities for the southern Great Plains lineage. All three of these accessions also had membership probabilities for other lineages, suggesting introgression and admixture from local native strains to varying degrees. Timber originated from multiple collections of switchgrass made along the Atlantic Seaboard, which is clearly reflected in its relatively high membership probabilities in two geographically diverse lineages (A and D in Fig. 3).

The remaining lowland lineages appear to derive from a highly diverse glacial refuge along the eastern Gulf Coast, extending deep into the Florida peninsula. The diversity within this region is evident on two levels: (1) genotypic variation among accessions, represented by dominant membership probabilities for four distinct lineages identified by STRUCTURE and (2) genotypic variation within accessions, represented by numerous accession with membership probabilities for several lineages, including one or more upland lineages. Observed levels of genotypic variability for these lineages extend all along the Atlantic Seaboard as far north as New Jersey and Staten Island, NY.

A model of two distinct lowland glacial refugia is in complete agreement with McMillan's conclusions based on morpho-geographic analyses of many switchgrass accessions covering much of its distributional area (McMillan 1959, 1964). We cannot speculate at this time whether these two refugia were isolated from each other or whether shared boundaries would have allowed some limited gene flow and admixture of their lowland lineages. Some limited gene flow between the two refugia could explain the presence of accessions that represent admixtures between the southern Great Plains lineage and eastern Gulf Coast lineage, but this would likely have required a more-or-less continuous distribution of lowland switchgrass all along the Gulf Coast during glacial maxima. Given the dominance of prevailing westerly winds as a mechanism of long-term

pollen transport, it is easy to imagine that very small annual movements of pollen from western strains could eventually lead to admixtures in some eastern strains as hypothesized by Casler et al. (2007a).

The picture of genotypic variation and radiation for the upland ecotype is more complex, partly due to the existence of two dominant ploidy levels, creating a robust isolation mechanism (Martinez-Reyna and Vogel 2002). Tetraploid upland switchgrass appears to have one major lineage that has spread across much of its longitudinal range from North Dakota to New York. Minor levels of differentiation are evident throughout the range of the tetraploid upland ecotype, shown as isolated membership probabilities for groups B and C in Fig. 3, with the cultivar Summer from eastern Nebraska as the only prominent outlier.

In contrast, the two upland octoploid lineages are differentiated on a geographic basis (Fig. 3). Group B is directly associated with the ancient tallgrass prairie ecosystem that dominated the Great Plains and extended into parts of central Minnesota and western Ohio (Fay and Fortier 2007). Group A is directly associated with the ancient broadleaf forest habitat, specifically mixed-mesophytic forest, which included oak savanna dominated by interspersed patches of oak forest and tallgrass prairie (Abrams 1990, 1992). Group B likely originated in the western montane refuge, with the greatest long-term adaptation to dryland and semi-arid conditions (McMillan 1959, 1964). Group A likely originated somewhere in the greater Gulf Coast refuge. Three accessions, with dominant membership in Group B, appear to be out of place (Fig. 3). The Toledo accession (TOL in Fig. 3) originated at the far eastern edge of the historical tallgrass prairie, the former "prairie peninsula" (Fay and Fortier 2007). Conversely, Carthage (CT in Fig. 3) and Albany (ALB in Fig. 3) originated a long distance from the tallgrass prairie, likely resulting from unconscious human-derived migrations.

The Carthage accession was collected near the town of Carthage, NC approximately 20 km from the entrance to Fort Bragg, a U.S. Army base that was established in 1918 (Anonymous 1967). During its first 30 years, Fort Bragg was one of the most important artillery training bases for the U.S. Army. Although mechanized infantry existed during that time, horses and mules were the primary "horsepower" for moving guns, munitions, and officers within U.S. Army artillery units. Over 1,000 horses and mules were housed at Fort Bragg during periods of maximum activity (Anonymous 1967). Fort Robinson, located in the Sand Hills of western Nebraska, is the "smoking gun" that provides a definite physical link between the tallgrass prairie and Fort Bragg. During the first half of the twentieth century, the U.S. Army bred and raised its own horses and mules at a series of Remount Stations across the western USA—Fort Robinson was the largest and most active of these Remount Stations (Buecker 2002). The U.S. Army regularly shipped horses and mules from Fort Robinson to active Army units throughout the eastern USA, always accompanied by many tons of prairie hay (Anonymous 1967; Buecker 2002). Because prairie hay is likely to be harvested at an advanced state of maturity, it will often contain ripe seeds of prairie grasses, including switchgrass, leaving very little left to link switchgrass from the Nebraska Sand Hills with rural areas surrounding Fort Bragg, North Carolina. Eerily similar, the Albany accession was collected near Albany, NY, approximately 250 km from Madison Barracks, which was connected to Fort Robinson by direct shipments of horses and hay to support an artillery unit prior to its motorization in 1934, and 160 km from the West Point Military Academy, also a heavy user of horses from Fort Robinson (Buecker 2002).

We hypothesize that the Carthage accession, and possibly the Albany accession, were established from prairie hay transported from the Sand Hills of Nebraska. These are the only two accessions for which we have any supporting evidence, direct or indirect, of human-influenced migrations. Of course, humans have been collecting and transporting switchgrass seed for well over 100 years and there are undoubtedly other instances of human-facilitated switchgrass migrations. However, invoking Occam's Razor, natural migrations facilitated by birds, mammals, and wind provides the simplest explanation for most of the observed geographic patterns observed in Fig. 3, with the obvious exceptions of the Carthage accession and perhaps the Albany accession. We do not rule out other human-induced migrations of some of the accessions in this and other studies, but we cannot identify or verify them with any current data or evidence.

Evolution and adaptive radiation

Our molecular clock computations suggest that the earliest divergence of switchgrass from the pool of diploid *Panicum* species, represented by *P. hallii* as the most similar diploid to *P. virgatum*, occurred approximately 5 M ybp. This is approximately 3 M years earlier than an estimate based on a Triticeae molecular clock that utilized genetic polymorphisms at the Acetyl-CoA carboxylase locus in the nuclear genome (Huang et al. 2003). In contrast, our estimate of the earliest taxonomic divergence within *P. virgatum* is approximately 1.3 M ybp, which agrees very closely with the Huang et al. (2003) estimate of ∼1 M ybp. Polymorphisms within the chloroplast genome of switchgrass suggest that both upland-lowland ecotype divergence and 4x-to-8x polyploidization events began to occur around 1.5–1 M ybp. These events have clearly occurred many times in the life history of this species, as

evidenced by: (1) a moderate frequency of individuals that defy unequivocal classification, (2) mixed ploidy levels or ecotype membership within individual localized collections, (3) bidirectional gene flow between 4x and 8x and between upland and lowland lineages, (4) intermediate phenotypes, and (5) the occurrence of 4x and 8x ploidy levels within both upland and lowland lineages.

During the past 1 M years, ice age cycles have occurred on an approximately 100-kyr cycle (Bintanja and van de Wal 2008), causing massive cyclic migrations of all major ecosystems, including tallgrass prairie and savanna habitats (Hewitt 2000). During glacial maxima, diverse ecosystems, ranging from tundra and taiga of northern Canada to subtropical swamps and forests of the Gulf Coast, were massively compressed into an area extending just south of the North American Great Lakes region (the southern limit of most glaciation events) to somewhere south of the current northern shoreline of the Gulf of Mexico (Deevey 1949). Long-term pollen records from soil and lakebed cores in Central Florida, the Florida panhandle, and the Coastal Plain of Georgia and the Carolinas indicate that grasses and other C4 species were most frequent in the southeastern landscape during the previous glacial maximum, ranging from 60,000 to 30,000 ybp (Grimm et al. 1993; Huang et al. 2006; LaMoreaux et al. 2009; Leigh 2008; Watts 1971; Watts et al. 1992). This was the tallgrass prairie and savanna of the last glacial maximum, likely extending west along the Gulf Coast into western Texas, New Mexico, and northern Mexico. Deglaciation was an extremely slow process, driven by thousands of years of incremental and cyclic changes in temperature, precipitation, and atmospheric CO_2 concentrations (Bintanja and van de Wal 2008) and punctuated by frequent long-term cold reversals (Berger et al. 1987; Cwynar and Levesque 1995; Jakobsen 2009; Levesque et al. 1993). Tallgrass prairie and savanna habitats gradually re-established themselves in the northern USA between 9,000 and 5,000 ybp, not in a single event, but in multiple events over thousands of years (Clark et al. 2001; Jacobson and Grimm 1986; Kelley et al. 2006). As a sessile organism, switchgrass was entirely dependent on birds and mammals as agents to facilitate migration of seeds to new locations (Ernst et al. 1992; Ocumpaugh et al. 1996; Pakeman 2001; Webb 1986). This provided a mechanism for multiple introductions of switchgrass to new colonization sites, potentially originating from a diverse array of existing sites and populations.

We hypothesize that adaptive radiation and selection during glacial minima are largely responsible for much of the phenotypic variability observed within switchgrass. Within each glaciation cycle, northward migration was a long and slow process, characterized by repeated colonization and extinction cycles. Mutations that resulted in earlier flowering and tolerance to longer daylength, allowed certain genotypes to complete cold-hardening cycles and produce ripe seed at more northern latitudes. Divergence in flowering time and photoperiod sensitivity is the basis for much of the phenotypic variability observed within the species today (Casler et al. 2007b). The vast majority of phenotypic variability in switchgrass occurs along latitudinal gradients, with earlier flowering, fewer phytomers, greater tiller density, reduced tiller size, and more compact panicle structure associated with more northern origins (Casler et al. 2004, 2007b). Much of this variability is associated with the two dominant ecotypes of switchgrass, but there is considerable within-ecotype phenotypic variability that appears to be adaptive in nature, likely resulting from selection and not associated with ploidy level per se.

Genetic bottlenecks occurred during post-glacial recolonizations by both plants and animals, many of which have persisted (Hewitt 2000). Two major geographic features are likely responsible for restricted northward migration routes, which would have enhanced bottleneck effects. The Appalachian Mountains, beginning in northern Alabama and extending north-northeast into Pennsylvania and New York would have been a significant barrier limiting westerward movement of eastern lineages and vice versa for western Gulf Coast lineages. Similarly, the east–west precipitation clines of the Great Plains, loosely centered on the 100th meridian, formed a barrier to east–west movements of drought-tolerant lineages from the western montane refuge and humid-adapted lineages from the western Gulf Coast refuge.

Our observations that allelic diversity does not decline with increased latitude-of-origin indicate that genetic bottlenecks likely occurred in switchgrass, but were overcome by other forces. Polyploidy is the rule in switchgrass and higher ploidy levels, such as 8x, are very common, particularly at higher latitudes. Polyploids are capable of storing huge amounts of cryptic genetic variability that would be useful under fluctuating environmental conditions, allowing populations to use natural selection as a long-term survival mechanism (Stebbins 1985; Symonds et al. 2010). The nearly obligate allogamous mating strategy of switchgrass encourages cross-pollination among diverse individuals and helps to preserve genetic diversity through a highly heterozygous genome (Martinez-Reyna and Vogel 2002; Vogel 2004). Pollination facilitated by wind further encourages cross-pollination between geographically and phenotypically diverse strains, leading to potentially heterotic effects from the combination of diverse and complimentary linkage blocks (Lamkey and Edwards 1999; Melchinger 1999; Vogel and Mitchell 2008). These forces resulted in a balance between migration, drift, and selection, creating phenotypic diversity across a broad landscape (McMillan 1959; Casler et al.

2007b), but allowing continued migration to balance the effects of drift, maintaining large effective population sizes across a broad landscape. Even with the loss of more than 99% of the tallgrass prairie and savanna ecosystems to habitat loss and fragmentation (Stubbendieck et al. 1991), massive amounts of genetic diversity have been preserved in nearly every accession that we have sampled (Casler et al. 2007a; Zalapa et al. 2011).

The substantially higher frequency of octoploids at northern latitudes is difficult to ignore. Early studies of polyploidy demonstrated that simple hypotheses, such as greater tolerance to environmental stresses, cannot explain adaptive differentiation between ploidy levels (Stebbins 1985) and our studies of switchgrass support this conclusion (Casler et al. 2004, 2007b). Rather, Stebbins (1985) hypothesized that higher polyploids are more capable of invading and colonizing new or disturbed habitats. Duplicate genomes present in allopolyploids would provide one level of benefit, while polysomic inheritance within each genome would create a second benefit: cryptic genetic variability to enhance the ability of populations to respond to fluctuating environments and to produce progeny capable of colonizing new environments. Because the success of invaders or colonists depends on the creation of favorable gene combinations, the allopolyploid and autopolyploid nature of octoploid switchgrass (Vogel 2004) may explain the significantly higher frequency of octoploid genotypes in habitats colonized during post-glacial migrations (Stebbins 1985). Indeed, if octoploid genotypes have a higher probability of colonization success, cyclic ice ages would have repeatedly created conditions under which these genotypes were favored over tetraploids, creating a persistent driver for regeneration and reestablishment of octoploid populations at more northern latitudes.

Hybridization between diverse switchgrass strains was further enhanced during periods of glacial maxima when the geographic distribution of switchgrass was compressed into glacial refugia. The eastern Gulf Coast most likely represents the primary center of diversity observed to date. Diversity levels in the western Gulf Coast region are considerably lower than in the eastern Gulf Coast, possibly due to insufficient sampling and coverage for the region. The eastern Gulf Coast region contains remnants of both upland and lowland ecotypes, and the only evidence to date of 8x individuals with a clear and obvious lowland phenotype (Oscar Scherer State Park, Pasco County, PMT-785, and Sprewell Bluff). Gene flow between ecotypes and ploidy levels likely occurs at a very low rate per pollination event. Large phenotypic differences in flowering time represent a significant pollination barrier between upland and lowland ecotypes (Cortese et al. 2010; Vogel 2004). Ploidy also represents a significant pollination barrier, as interploidy hybridizations are both rare and difficult to achieve even under controlled conditions (Martinez-Reyna and Vogel 2002). Gene flow from 8x to 4x can occur in switchgrass by polyembryony and parthenogenesis (Young et al. 2010) and likely occurs from 4x to 8x by 2n gametes (Harlan and de Wet 1975), but both of these processes occur at low frequencies. However, because ice ages occur on a 100-kyr cycle, long-term proximal exposure increased the opportunities for intertaxa hybridization within these refugia, so fairly rare events could lead to unusual phenotypes and genotypes in the frequencies observed in our studies.

Finally, the bulk of phenotypic and genotypic data indicate that there are clearly regional gene pools of switchgrass that serve as functional populations across a fairly broad landscape. Switchgrass does not fit the current model for tallgrass prairie and savanna habitat restoration which favors using source-identified "local" populations collected within fairly narrow geographic zones, typically defined by artificial political boundaries (Gustafson et al. 2004). With climate change looming on the near-term horizon, restoration and conservation initiatives could benefit from short-term mimicry of the natural process of tallgrass prairie and savanna restoration: develop gene pools of multiple source-identified accessions for use across broadly functional climatic and edaphic zones. Gene pools that broadly represent a functional zone (e.g., hardiness zones 3 and 4 within the ancient tallgrass prairie ecosystem; Vogel et al. 2005) as opposed to narrowly-defined local collections, would contain massive amounts of genetic variability, providing populations a significantly greater opportunity to respond to climate change with seedling recruitment and selection during times of fluctuating environments. An additional advantage of these regional gene pools would be a reduction in the number of populations that must be maintained in seed orchards, allowing members of the restoration and conservation communities to focus their efforts on many more species and species diversity per se.

References

Abrams MD (1990) Adaptations and responses to drought in *Quercus* species of North America. Tree Physiol 7:227–238

Abrams MD (1992) Fire and the development of oak forests. Bioscience 42:246–253

Anonymous (1967) Fort Bragg history 1918–1967. http://www.mybase guide.com/article/military/ft-bragg/615/Units-and-History

Barkworth ME, Anderton LK, Capels KM, Long S, Piep MB (2007) Manual of grasses for North America. Utah State University Press, Logan

Berger WH, Killingley JS, Vincent E (1987) Time scale of the Wisconsin/Holocene transition: oxygen isotope record in the Western Equatorial Pacific. Quatern Res 28:295–306

Bintanja R, van de Wal RSW (2008) North American ice-sheet dynamics and the onset of 100, 000-year glacial cycles. Nature 454:869–872

Buecker TR (2002) Fort Robinson and the American century, 1900–1948. Nebraska State Historical Society, Lincoln

Bunge J, Fitzpatrick M (1993) Estimating the number of species: a review. J Am Stat Assoc 88:364–373

Burnham K, Overton W (1978) Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65:623–633

Casler MD (2005) Ecotypic variation among switchgrass populations from the northern USA. Crop Sci 45:388–398

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

Casler MD, Stendal CA, Kapich L, Vogel KP (2007a) Genetic diversity, plant adaptation regions, and gene pools for switchgrass. Crop Sci 47:2261–2273

Casler MD, Vogel KP, Taliaferro CM, Ehlke NJ, Berdahl JD, Brummer EC, Kallenbach RL, West CP, Mitchell RB (2007b) Latitudinal and longitudinal adaptation of switchgrass populations. Crop Sci 47:2249–2260

Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43:783–791

Chao A, Hwang W, Chen Y, Kuo C (2000) Estimating the number of shared species in two communities. Stat Sin 10:227–246

Clark JS, Grimm EC, Lynch J, Mueller PG (2001) Effects of Holocene climate change on the C_4 grassland/woodland boundary in the Northern Plains, USA. Ecology 82:620–636

Colwell R, Coddington J (1994) Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc Ser B 345:101–118

Cortese LM, Honig J, Miller C, Bonos SA (2010) Genetic diversity of twelve switchgrass populations using molecular and morphological markers. Bioenerg Res 3:262–271

Costich DE, Friebe B, Sheehan MJ, Casler MD, Buckler ES (2010) Genome-size variation in switchgrass (*Panicum virgatum*): flow cytometry and cytology reveal rampant aneuploidy. Plant Genome 3:130–141

Cwynar LC, Levesque AJ (1995) Chironomid evidence for late-glacial climatic reversals in Maine. Quatern Res 43:405–413

Deevey ES Jr (1949) Biogeography of the pleistocene: Part I: Europe and North America. Geog Soc Am Bull 60:1315–1416

Doebley J (2004) The genetics of maize evolution. Annu Rev Genet 38:37–59

Drummond A, Rambauat A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolut Biol 7:214

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Ernst WHO, Veenendaal EM, Kebakile MM (1992) Possibilities for dispersal in annual and perennial grasses in a savanna in Botswana. Vegetatio 102:1–11

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:74–75

Fay J, Fortier AC (2007) The tall grass prairie peninsula: its role in shaping American culture. Stipes Publishing, Champaign

Grimm EC, Jacobson GL Jr, Watts WA, Hansen BCS, Maasch KA (1993) A 50, 000-year record of climate oscillations from Florida and its temporal correlation with the Heinrich Events. Science 261:198–200

Gunter LE, Tuscan GA, Wullshcleger SD (1996) Diversity of switchgrass based on RAPD markers. Crop Sci 36:1017–1022

Gustafson DJ, Gibson DJ, Nickrent DL (2004) Using local seeds in prairie restoration: data support the paradigm. Native Plants Spring 2005:25–28

Harlan JR, de Wet JMJ (1975) On Ö. Winge and a prayer: the origins of polyploidy. Bot Rev 41:361–369

Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Bot J Linn Soc 58:247–276

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913

Huang S, Siu X, Haselkorn R, Gornicki P (2003) Evolution of switchgrass (*Panicum virgatum* L.) based on sequences of the nuclear gene encoding plastic acetyl-CoA carboxylase. Plant Sci 164:43–49

Huang Y, Shuman B, Wang Y, Webb T III, Grimm EC, Jacobson GL Jr (2006) Climatic and environmental controls on the variation of C3 and C4 plant abundances in central Florida for the past 62, 000 years. Palaeogeog Palaeoclim Palaeoecol 237:428–435

Jacobson GT, Grimm EC (1986) A numerical analysis of Holocene forest and prairie vegetation in Central Minnesota. Ecology 67:958–966

Jakobsen BH (2009) Holocene climate change and environmental reconstruction in East Greenland. IOP Conference Series: Earth Environ Sci, vol 6, p 072028, doi:10.1088/1755-1307/6/7/072028

Kelley DW, Brachfeld SA, Nater EA, Wright HE Jr (2006) Sources of sediment in Lake Pepin on the Upper Mississippi River in response to Holocene climate changes. J Paleoclim 35:193–206

Kneller M, Peteet D (1999) Late-glacial to early Holocene climate changes from a Central Applachian pollen and macrofossil record. Quatern Res 51:133–147

Lamkey DR, Edwards JW (1999) Quantitative genetics and heterosis. In: Coors JG, Pandey S (eds) Genetics and exploration of heterosis in crops. ASA-CSSA-SSSA, Madison, pp 31–48

LaMoreaux HK, Brook GA, Knox JA (2009) Late Pleistocene and Holocene environments of the Southwestern United States from

the stratigraphy and pollen content of a peat deposit on the Georgia Coastal Plain. Palaeogeog Palaeoclim Palaeoecol 280: 300–312

Leigh D (2008) Late quaternary climates and river channels of the Atlantic Coastal Plain, Southeastern USA. Geomorph 101:90–108

Levesque AJ, Mayle FE, Walker IR, Cwynar LC (1993) A previously unrecognized late-glacial cold event in eastern North America. Nature 361:623–626

Maddison WP, Maddison DR (2007) Mesquite: a modular system for evolutionary analysis. Version 2.74. http://mesquiteproject.org

Martinez-Reyna JM, Vogel KP (2002) Incompatability systems in switchgrass. Crop Sci 42:1800–1805

McMillan C (1959) The role of ecotypic variation in the distribution of the central grassland of North America. Ecol Mono 29:285–308

McMillan C (1964) Ecotypic differentiation within four North American prairie grasses. I. Morphological variation within transplanted community fractions. Am J Bot 51:1119–1128

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) Genetics and exploration of heterosis in crops. ASA-CSSA-SSSA, Madison, pp 99–118

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 Res Crop Evol 53:1291–1302

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

Ocumpaugh WR, Archer S, Stuth JW (1996) Switchgrass recruitment from broadcast seed vs. seed fed to cattle. J Range Manage 49:368–371

Pakeman RJ (2001) Plant migration rates and seed dispersal mechanisms. J Biogeog 28:795–800

Palmer M (1991) Estimating species richness: the second-order jackknife reconsidered. Ecology 72:1512–1513

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Prasad V, Stroemberg C, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. Science 310:1170–1180

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phyolgeography of plants from the Pacific Northwest of North America. Plant Syst Evol 206: 353–373

Stebbins GL (1985) Polyploidy, hybridization, and the invasion of new habitats. Ann Missouri Bot Garden 72:824–832

Stubbendieck J, Hatch SL, Butterfield CH (1991) North American range plants. University of Nebraska Press, Lincoln

Symonds VV, Soltis PS, Soltis DE (2010) Dynamics of polyploidy formation in Tragopogon (Asteraceae): recurrent formation, gene flow, and population structure. Evolution 64:1984–2003

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol (in press)

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

Vogel KP, Mitchell RB (2008) Heterosis in switchgrass: biomass yield in swards. Crop Sci 48:2159–2164

Vogel KP, Schmer MR, Mitchell RB (2005) Plant adaptation regions: Ecological and climatic classification of plant materials. Rangeland Ecol Manage 58:315–319

Watts WA (1971) Postglacial and interglacial vegetation history of Southern Georgia and Central Florida. Ecology 52:676–690

Watts WA, Hansen BCS, Grimm EC (1992) Camel Lake: a 40,000-yr record of vegetational and forest history from Northwest Florida. Ecology 73:1056–1066

Webb S (1986) Potential role of passenger pigeons and other vertebrates in the rapid Holocene migrations of nut trees. Quatern Res 26:367–375

Young HA, Hernlem BJ, Anderton AL, Lanzatella CL, Tobias CM (2010) Dihaploid stocks of switchgrass isolated by a screening approach. Bioenerg Res 3:305–313

Zalapa JE, Price DL, Kaeppler SM, Tobias CM, Okada M, Casler MD (2011) Hierarchical classification of switchgrass using SSR and chloroplast sequences: ecotypes, ploidies, gene pools, and cultivars. Theor Appl Genet 122:805–817

Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2009) Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. PLoS One 4:e5297