

UC Davis

UC Davis Previously Published Works

Title

Extensive Genetic Diversity is Present within North American Switchgrass Germplasm

Permalink

<https://escholarship.org/uc/item/185994vs>

Journal

The Plant Genome, 11(1)

ISSN

1940-3372

Authors

Evans, Joseph

Sanciango, Millicent D

Lau, Kin H

et al.

Publication Date

2018-03-01

DOI

10.3835/plantgenome2017.06.0055

Peer reviewed

# Extensive Genetic Diversity is Present within North American Switchgrass Germplasm

Joseph Evans, Millicent D. Sanciangco, Kin H. Lau, Emily Crisovan, Kerrie Barry, Chris Daum, Hope Hundley, Jerry Jenkins, Megan Kennedy, Govindarajan Kunde-Ramamoorthy, Brieanne Vaillancourt, Ananta Acharya, Jeremy Schmutz, Malay Saha, Shawn M. Kaeppler, E. Charles Brummer, Michael D. Casler, and C. Robin Buell\*

## Abstract

Switchgrass (*Panicum virgatum* L.) is a perennial native North American grass present in two ecotypes: upland, found primarily in the northern range of switchgrass habitats, and lowland, found largely in the southern reaches of switchgrass habitats. Previous studies focused on a diversity panel of primarily northern switchgrass, so to expand our knowledge of genetic diversity in a broader set of North American switchgrass, exome capture sequence data were generated for 632 additional, primarily lowland individuals. In total, over 37 million single nucleotide polymorphisms (SNPs) were identified and a set of 1.9 million high-confidence SNPs were obtained from 1169 individuals from 140 populations (67 upland, 65 lowland, 8 admixed) were used in downstream analyses of genetic diversity and population structure. Seven separate population groups were identified with moderate genetic differentiation [mean fixation index (*F<sub>st</sub>*) estimate of 0.06] between the lowland and the upland populations. Ecotype-specific and population-specific SNPs were identified for use in germplasm evaluations. Relative to rice (*Oryza sativa* L.), maize (*Zea mays* L.), soybean [*Glycine max* (L.) Merr.], and *Medicago truncatula* Gaertn., analyses of nucleotide diversity revealed a high degree of genetic diversity (0.0135) across all individuals, consistent with the outcrossing mode of reproduction and the polyploidy of switchgrass. This study supports the hypothesis that repeated glaciation events, ploidy barriers, and restricted gene flow caused by flowering time differences have resulted in distinct gene pools across ecotypes and geographic regions. These data provide a resource to associate alleles with traits of interest for forage, restoration, and biofuel feedstock efforts in switchgrass.

Plant Genome 11:170055  
doi: 10.3835/plantgenome2017.06.0055

© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Core Ideas

- This was the largest panel of switchgrass genetic diversity generated to date.
- The Gulf coast of the United States is the center of genetic diversity for switchgrass.
- There was a genetic bottleneck in upland switchgrass.

**S**WITCHGRASS is an important perennial grass species being developed as a biofuel feedstock crop. Switchgrass is a polyploid species native to North America, with two main ecotypes across its habitat: upland switchgrass, found largely in the northern United States and southern Canada, and lowland switchgrass, found along the eastern seaboard and southern United States and extending

J. Evans, M.D. Sanciangco, K.H. Lau, E. Crisovan, B. Vaillancourt, C.R. Buell, Department of Energy (DOE) Great Lakes Bioenergy Research Center, Michigan State Univ., East Lansing, MI 48824; J. Evans, (current address), DuPont Pioneer, Johnston, IA 50131; J. Evans, M.D. Sanciangco, K.H. Lau, E. Crisovan, B. Vaillancourt, C.R. Buell, Dep. of Plant Biology, Michigan State Univ., East Lansing, MI 48824; K. Barry, C. Daum, H. Hundley, M. Kennedy, J. Schmutz, Dep. of Energy, Joint Genome Institute, Walnut Creek, CA 94598; G. Kunde-Mamamoorthy, Dep. of Energy, Joint Genome Institute, Walnut Creek, CA 94598; G. Kunde-Mamamoorthy (current address), The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032; J. Jenkins, J. Schmutz, HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806; A. Acharya, Univ. of Georgia, Athens, GA 30602; M. Saha, Noble Research Institute, Ardmore, OK 73401; S.M. Kaeppler, M.D. Casler, DOE Great Lakes Bioenergy Research Center, Univ. of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706; S.M. Kaeppler, Dep. of Agronomy, Univ. of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706; E.C. Brummer, Univ. of California, Davis, CA 95616; M.D. Casler, USDA-ARS, U.S. Dairy Forage Research Center, 1925 Linden Dr., Madison, WI 53706. Received 23 June 2017. Accepted 19 Sept. 2017. \*Corresponding author (buell@msu.edu).

**Abbreviations:** ARF, auxin response factors; *F<sub>st</sub>*, fixation index; GO, Gene Ontology; SNP, single nucleotide polymorphism.

into northern Mexico. Upland switchgrass populations are either tetraploid ( $2n = 4x = 36$ ) or octoploid ( $2n = 8x = 72$ ), whereas lowland switchgrass is primarily tetraploid with widespread aneuploidy as reported by Costich et al. (2010). In general, upland switchgrass exhibits better winter hardiness, earlier flowering, and less total biomass accumulation than lowland switchgrass (Casler, 2005; Casler et al., 2007; Vogel et al., 2004). Previous work has characterized the diversity of northern switchgrass at a genomic level (Evans et al., 2015, Lu et al., 2013), but analyses of southern switchgrass populations across a broader geographic range have been limited to a small number of polymerase chain reaction-based markers (Zalapa et al., 2011). Development of the 1.6 million HapMapv1 SNP set and interrogation of the 537 member Northern Switchgrass Association Panel (Evans et al., 2015) revealed five population groups and reduced genetic diversity in the upland compared with the lowland populations, consistent with the hypothesis that the repeated glaciations in northern latitudes have resulted in reduced genetic diversity in northern latitudes (Soltis et al., 1997).

The key goal of switchgrass biomass improvement projects is the development of adapted cultivars capable of producing high biomass yield under harsh winter temperatures in northern regions or under high temperatures, periodic droughts, and weathered soils in southern regions. An improved understanding of the genetic diversity and population structure of switchgrass germplasm would enable informed selection of parents to produce better hybrids with optimal performance. Switchgrass is also used frequently for ecological restoration and in such cases, knowledge of genetic relatedness may increase the success of identifying locally adapted accessions for the target areas.

A reference genome for switchgrass derived from 'AP13', an individual genotype selected from the 'Alamo' cultivar, which is a lowland tetraploid switchgrass cultivar native to Texas, is available. It is approximately 1230 Mbp in size and encodes ~98,000 predicted genes (phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\_Pvirgatum, accessed 25 Oct. 2017). Switchgrass is a primarily outcrossing species and exhibits inbreeding depression, resulting in a high level of heterozygosity in the genome (Sharma et al., 2012). Switchgrass also contains many repetitive elements within the nuclear genome, with estimates placing the amount of repetitive DNA at up to 60% (Sharma et al., 2012). To minimize sequencing of repetitive DNA, an exome capture sequencing probe set was developed that targeted approximately 50 Mb of the exonic regions (Evans et al., 2014) and used to estimate diversity (Evans et al., 2015), identify loci associated with flowering time (Grabowski et al., 2017), and develop genome prediction models (Ramstein et al., 2016).

In this study, exome capture sequencing was used to explore the genic space of 1169 switchgrass individuals, 537 of which were from the Northern Switchgrass Association Panel (Evans et al., 2015) and 632 were from two separate panels of lowland switchgrass, the Southern

Switchgrass Association Panel (Acharya, 2014) and the Supplemental Southern Switchgrass Association Panel (this study). In total, the 1169 individuals span 140 populations. The sequencing captured a combined total of ~5,300 Gbp of sequence data yielding a 1,878,584 SNP set, HapMapv2, which had coverage across all 1169 individuals. With the panels encompassing a robust representation of North American switchgrass, seven population groups were identified, enabling genetic diversity estimates in lowland and upland switchgrass. Characterization of distinct gene pools in these seven population groups will provide a foundation for more targeted research in biofuel feedstock development and prairie restoration.

## Materials and Methods

### Description of Population and DNA Isolation

Metadata associated with the Southern Switchgrass Associational Panel and Supplemental Southern Switchgrass Association Panel are provided in Table 1. In brief, materials with listed PI numbers originated from the National Plant Germplasm System (<http://www.ars-grin.gov/npgs/>, accessed 25 Oct. 2017) and the others were collected during 2008. Material without PI numbers represent wild switchgrass populations. Plants used in the Southern Switchgrass Association Panel were grown initially at the University of Georgia, Athens, GA; subsequently, clones were propagated and planted at the Noble Foundation in Ardmore, OK. The Supplemental Southern Switchgrass Association Panel was developed during two collecting trips during August and September 2011 and based on collection of whole leaves from at least 10 widely dispersed plants from each accession. Leaf punches were taken for DNA extraction and sequencing as described previously (Acharya, 2014). For the Supplemental Southern Switchgrass Association panel, leaf punches were taken from switchgrass plants located near the listed global positioning system coordinates and DNA was extracted using the cetyl trimethylammonium bromide protocol (Saghai-Marooof et al., 1984).

### Exome Capture, Read Alignment, and Read Processing

Exome capture sequencing was performed on the individuals with the Roche-NimbleGen SeqCap EZ kit and protocol (Roche-Nimblegen, Madison, WI) as described previously (Mascher et al., 2013), with the exception that Kapa Biosystems reagents were used for library preparation. Capture was performed using the Roche-NimbleGen '120911\_Switch-grass\_GLBRC\_R\_EZ\_HX1' probe set (Roche-NimbleGen). Capture and sequencing steps were all performed by the Department of Energy Joint Genome Institute, Walnut Creek, CA. Individuals were sequenced on an Illumina HiSeq 2000 platform (Illumina, San Diego, CA), generating 150-nt paired-end reads; 12 samples were sequenced per lane on the flow-cell, yielding approximately 3.2 Gb of sequencing data per sample and a total of ~5 Tb of sequencing data. In

**Table 1. Metadata on populations within the Northern Switchgrass Association, Southern Switchgrass Association, and Supplemental Switchgrass Association panels used in this study.**

Population name	Association panel	Ecotype (STRUCTURE) <sup>†</sup>	Pedigree	Individuals (n)	State or province	Elevation (m asl)	Latitude	Longitude	Reference <sup>‡</sup>
'196' (PI 337553)	Southern	Upland	Natural population	15	Rafaela Experiment Station, Argentina				Acharya 2014
'Alamo' (PI 422006)	Southern	Lowland	Bred cultivar	15	TX		28.33	98.12	Acharya 2014
'AM-314-MS-155' (14; PI 421999)	Southern; Supplemental Southern	Lowland	Natural population	19	AR; Pangburn, AR		35.42; 35.42664	91.84; 91.836	Acharya 2014; This study (5)
'API3 (Alamo)'	Reference genotype	Lowland	Single genotype derived from Alamo	2	TX		28.33	98.12	Missaoui et al. 2005
'Big Branch'	Supplemental Southern	Lowland	Natural population	5	LA		30.27438	89.9552	This study
'Biloxi'	Supplemental Southern	Lowland	Natural population	5	MS		30.77678	88.76443	This study
'Biloxi One'	Supplemental Southern	Lowland	Natural population	3	MS		30.73199	88.78432	This study
'Biloxi Two'	Supplemental Southern	Lowland	Natural population	2	MS		30.74736	88.78172	This study
'Blackwell' (PI 421520)	Northern; Southern	Upland	Natural track cultivar	25	OK	320	36.76; 36.71	97.24; 97.08	Lu et al. 2013 (10); Acharya 2014 (15)
'Blair'	Supplemental Southern	Lowland	Natural population	5	OK		34.84099	99.36858	This study
'BN-10860-61' (PI 315724)	Southern	Upland	Natural population	13	KS		38.73	98.23	Acharya 2014
'BN-11357-63' (PI 315727)	Southern	Lowland	Natural population	15	NC		35.73	78.85	Acharya 2014
'BN-12323-69' (PI 414070)	Southern; Supplemental Southern	Lowland	Natural population	17	KS		38.81	98.27	Acharya 2014 (15); This study (2)
'BN-13645-64' (PI 315728)	Southern	Admixed	Natural population	14	NC (donated MD)		34.71	80.16	Acharya 2014
'BN-14668-65' (PI 414065)	Southern; Supplemental Southern	Lowland	Natural population	16	AR; Pangburn, AR		35.42; 35.42664	91.84; 91.836	Acharya 2014 (16); This study (2)
'BN-14669-92' (PI 315725)	Southern	Admixed, Lowland	Natural population	2	MS		33.95	89.67	Acharya 2014
'BN-18758-67' (PI 414068)	Southern	Upland	Natural population	12	KS		39.08	96.63	Acharya 2014
'BN-8358-62' (PI 315723)	Southern	Admixed, Lowland	Natural population	15	NC		35.03	79.55	Acharya 2014
'BN-8624-67' (PI 414067)	Southern	Admixed, Upland	Natural population	14	NC		35.47	80.05	Acharya 2014
'Brookville'	Supplemental Southern	Lowland	Natural population	4	MS		33.201	88.569	This study
'Bullis'	Supplemental Southern	Admixed, Lowland	Natural population	3	TX		29.749389	98.575556	This study
'Cajun Prairie'	Supplemental Southern	Admixed, Lowland	Natural population	1	LA		30.49919	92.40636	This study
'Campo'	Supplemental Southern	Upland	Natural population	5	CO		37.00852	102.74671	This study
'Carthage' (PI 421138)	Northern; Southern	Admixed, Upland	Natural track cultivar	21	NC	120	35.31; 35.2	79.3; 79.47	Lu et al. 2013 (8); Acharya 2014 (13)
'Cave-in-Rock'	Northern	Upland	Natural track cultivar	10	IL	120	37.47	88.16	Lu et al. 2013
'Cimarron River'	Supplemental Southern	Admixed, Upland	Natural population	5	KS		37.07472	102.00962	This study
'Citrus-Co-FL'	Southern	Admixed, Lowland	Natural population	3	FL		28.9	82.59	Acharya 2014
'Dacotah'	Northern	Upland	Natural track cultivar	8	ND	520	46.38	100.94	Lu et al. 2013
'ECS-1'	Northern	Lowland	Natural population	5	NJ	18	39.82	74.53	Lu et al. 2013
'ECS-10'	Northern	Upland	Natural population	6	PA	310	40.95	79.62	Lu et al. 2013
'ECS-11'	Northern	Upland	Natural population	6	PA	430	41.4	79.77	Lu et al. 2013
'ECS-12'	Northern	Upland	Natural population	7	NY	2	42.72	73.83	Lu et al. 2013
'ECS-2'	Northern	Upland	Natural population	6	OH	200	41.58	83.67	Lu et al. 2013
'ECS-6'	Northern	Lowland	Natural population	2	MD	1	38.08333333	75.33333333	Lu et al. 2013
'Falcon' (PI 642190)	Southern	Upland	Bred cultivar	14	NM		36.45	103.18	Acharya 2014
'Forewinds'	Supplemental Southern	Admixed, Upland	Natural population	4	KS		37.17704	101.4022	This study
'Forgan'	Supplemental Southern	Admixed, Upland	Natural population	5	KS		37.0166	100.49197	This study
'Fort Polk'	Supplemental Southern	Lowland	Natural population	5	LA		30.97482	93.16632	This study
'Grand Bay'	Supplemental Southern	Lowland	Natural population	5	MS		30.45051	88.65633	This study
'Greenville'	Supplemental Southern	Lowland	Natural population	5	AL		31.618889	86.532778	This study
'Grenville' (PI 414066)	Southern	Admixed, Upland	Bred cultivar	13	NM		36.59	103.62	Acharya 2014
'Harrell'	Supplemental Southern	Admixed	Natural population	5	MS		32.33648	89.43923	This study
'High Tide'	Northern	Lowland	Natural track cultivar	11	MD	5	39.61	76.15	Lu et al. 2013
'HSP-FL'	Southern	Lowland	Natural population	15	FL		28.14	82.22	Acharya 2014
'Intracoastal'	Supplemental Southern	Lowland	Natural population	3	LA		29.80047	92.13931	This study
'Kanlow' (PI 421521)	Northern; Southern	Lowland	Bred cultivar	23	OK; OK (developed in KS)	250	35.26	96.18	Lu et al. 2013 (11); Acharya 2014 (12)

(cont'd.)

**Table 1. Continued.**

Population name	Association panel	Ecotype (STRUCTURE) <sup>†</sup>	Pedigree	Individuals (n)	State or province	Elevation (m asl)	Latitude	Longitude	Reference <sup>‡</sup>
'Kisatchie'	Supplemental Southern	Lowland	Natural population	5	LA		31.01393	93.04425	This study
'KY1625' (PI 431575)	Northern; Southern	Upland	Natural track cultivar; Natural population	20	WV	750	37.94	80.99	Lu et al. 2013 (10); Acharya 2014 (10)
'Lake E.T.'	Supplemental Southern	Lowland	Natural population	5	OK		34.73458	98.51987	This study
'LBJ'	Supplemental Southern	Lowland	Natural population	5	TX		33.40871	97.66816	This study
'Lueders'	Supplemental Southern	Lowland	Natural population	5	TX		32.75211	99.6242	This study
'Meade State Park'	Supplemental Southern	Admixed, Upland	Natural population	1	KS		37.17	100.44	This study
'Middle Springs'	Supplemental Southern	Admixed, Upland	Natural population	5	KS		37.11332	101.9268	This study
'Mineral Wells'	Supplemental Southern	Lowland	Natural population	5	TX		32.7977	98.18715	This study
'Mt. Dora'	Supplemental Southern	Upland	Natural population	5	NM		36.56362	103.56075	This study
'Mustang Lake'	Supplemental Southern	Admixed, Lowland	Natural population	3	TX		28.244722	96.809167	This study
'OSSP-FL'	Southern	Lowland	Natural population	15	FL		27.18	82.46	Acharya 2014
'P33'	Supplemental Southern	Lowland	Natural population	5	TX		32.7174	98.66863	This study
'Panola'	Supplemental Southern	Admixed	Natural population	6	AL		32.912272	88.247381	This study
'Pasco-Co-FL'	Southern	Admixed	Natural population	12	FL		28.6	82.4	Acharya 2014
'Pathfinder'	Northern	Admixed, Upland	Bred cultivar	10	NE	370	41.2	96.5	Lu et al. 2013
PI422016	Southern	Admixed, Lowland	Natural population	13	FL				Acharya 2014
'Pitkin'	Supplemental Southern	Lowland	Natural population	5	LA		30.9548	93.16072	This study
'PMT-785' (PI 422003)	Southern	Admixed, Lowland	Natural population	13	FL				Acharya 2014
'Possum Kingdom'	Supplemental Southern	Lowland	Natural population	5	TX		32.8818	98.56618	This study
'Post'	Supplemental Southern	Admixed	Natural population	5	TX		33.20674	101.25124	This study
'Reydon'	Supplemental Southern	Lowland	Natural population	6	OK		35.65006	99.92577	This study
'Roby'	Supplemental Southern	Lowland	Natural population	5	TX		32.76107	100.30605	This study
'Roswell'	Supplemental Southern	Lowland	Natural population	3	NM		33.61756	104.36784	This study
'Sabine'	Supplemental Southern	Lowland	Natural population	5	LA		30.84485	93.56693	This study
'Santa Rosa'	Supplemental Southern	Admixed, Upland	Natural population	1	NM		34.9388	104.69086	This study
'Shelter'	Northern	Upland	Natural track cultivar	9	WV	250	39.41	81.2	Lu et al. 2013
'SNF'	Southern	Lowland	Natural population	14	SC		34.53	81.63	Acharya 2014
'SP-Bluff'	Southern	Admixed, Lowland	Natural population	12	FA		32.87	84.48	Acharya 2014
'Stuart' (PI 422001)	Southern	Lowland	Bred cultivar	15	FL		27.2	80.25	Acharya 2014
'Summer' (PI 642191)	Southern	Upland	Bred cultivar	13	SD		42.68	96.68	Acharya 2014
'Sunburst'	Northern	Admixed, Upland	Bred cultivar	9	SD	340	42.6	92.6	Lu et al. 2013
'SW102'	Northern	Upland	Natural population	10	WI	210	44.01666667	91.48333333	Lu et al. 2013
'SW109'	Northern	Upland	Natural population	10	WI	320	44.26666667	89.66666667	Lu et al. 2013
'SW110'	Northern	Upland	Natural population	10	WI	320	44.2	89.66666667	Lu et al. 2013
'SW112'	Northern	Upland	Natural population	10	WI	240	43.46666667	89.43333333	Lu et al. 2013
'SW114'	Northern	Upland	Natural population	8	WI	280	42.56666667	90.4	Lu et al. 2013
'SW115'	Northern	Upland	Natural population	7	WI	280	42.56666667	90.4	Lu et al. 2013
'SW116'	Northern	Upland	Natural population	10	WI	210	43.2	90.45	Lu et al. 2013
'SW122'	Northern	Upland	Natural population	7	WI	210	43.2	90.33333333	Lu et al. 2013
'SW123'	Northern	Upland	Natural population	11	WI	240	42.78333333	88.3	Lu et al. 2013
'SW124'	Northern	Upland	Natural population	10	WI	180	42.55	87.8	Lu et al. 2013
'SW127'	Northern	Upland	Natural population	9	WI	270	42.9	88.45	Lu et al. 2013
'SW128'	Northern	Upland	Natural population	9	WI	250	42.85	88.63333333	Lu et al. 2013
'SW129'	Northern	Upland	Natural population	10	WI	320	44.33333333	89.6	Lu et al. 2013
'SW31'	Northern	Upland	Natural population	8	IN	260	40.3	86.22	Lu et al. 2013
'SW33'	Northern	Admixed, Upland	Natural population	9	IN	250	40.45	86.18	Lu et al. 2013
'SW38'	Northern	Upland	Natural population	4	IN	250	40.1	86.72	Lu et al. 2013
'SW40'	Northern	Upland	Natural population	8	IN	180	41.63	87.43	Lu et al. 2013
'SW43'	Northern	Upland	Natural population	8	MI	290	42.3	84.28	Lu et al. 2013
'SW46'	Northern	Upland	Natural population	9	MI	174	42.5	82.57	Lu et al. 2013
'SW49'	Northern	Upland	Natural population	7	MN	300	43.8	91.83	Lu et al. 2013
'SW50'	Northern	Admixed, Upland	Natural population	5	MN	400	46.2	94.42	Lu et al. 2013

(cont'd.)

Table 1. Continued.

Population name	Association panel	Ecotype (STRUCTURE)†	Pedigree	Individuals (n)	State or province	Elevation (m asl)	Latitude	Longitude	Reference‡
'SW51'	Northern	Admixed, Upland	Natural population	7	MN	320	44.53	95.08	Lu et al. 2013
'SW58'	Northern	Upland	Natural population	10	MN	310	44.32	93.93	Lu et al. 2013
'SW63'	Northern	Upland	Natural population	6	NY	300	42.93	78.18	Lu et al. 2013
'SW64'	Northern	Upland	Natural population	11	OH	330	40.6	80.67	Lu et al. 2013
'SW65'	Northern	Upland	Natural population	8	OH	320	40.55	80.67	Lu et al. 2013
'SW781'	Northern	Lowland	Natural population	6	NY	1	40.6	74.13	Lu et al. 2013
'SW782'	Northern	Upland	Natural population	7	VA	510	38.48	78.52	Lu et al. 2013
'SW786'	Northern	Upland	Natural population	7	MI	190	43	86	Lu et al. 2013
'SW787'	Northern	Upland	Natural population	10	MI	180	43.09	86.25	Lu et al. 2013
'SW788'	Northern	Lowland	Natural population	11	NY	1	40.68	74.01	Lu et al. 2013
'SW789'	Northern	Admixed	Multisite synthetic	7	AR & MS		various	various	Lu et al. 2013
'SW790'	Northern	Lowland	Bred cultivar	5	MS	80	34.125	89.03333333	Lu et al. 2013
'SW793'	Northern	Lowland	Natural population	5	NY	11	40.5	74.22	Lu et al. 2013
'SW795'	Northern	Lowland	Natural population	9	NY	1	40.6105	74.08188333	Lu et al. 2013
'SW796'	Northern	Lowland	Natural population	9	NY	1	40.62	74.18	Lu et al. 2013
'SW797'	Northern	Lowland	Natural population	8	NY	35	40.72	73.58	Lu et al. 2013
'SW798'	Northern	Lowland	Natural population	6	NY	3	41.04	71.93	Lu et al. 2013
'SW799'	Northern	Lowland	Natural population	7	NY	3	41.04	71.93	Lu et al. 2013
'SW802'	Northern	Lowland	Natural population	5	NY	35	40.72	73.58	Lu et al. 2013
'SW803'	Northern	Lowland	Natural population	9	NY	3	41.04	71.93	Lu et al. 2013
'SW805'	Northern	Lowland	Natural population	9	NY	3	41.02	72.01	Lu et al. 2013
'SW806'	Northern	Lowland	Natural population	9	NY	1	40.94	72.28	Lu et al. 2013
'SW808'	Northern	Upland	Natural population	10	WV	450	39.68	79.81	Lu et al. 2013
'SW809'	Northern	Upland	Natural population	10	WV	670	39.41	79.66	Lu et al. 2013
'SWFWMD-FL'	Southern	Lowland	Natural population	5	FL		27.9	81.59	Acharya 2014
'SWG32'	Northern	Lowland	Natural population	7	IL	250	42.33	89.02	Lu et al. 2013
'SWG39'	Northern	Lowland	Natural population	8	IA	400	43.44	92.38	Lu et al. 2013
'T16971' (PI 476296)	Southern	Upland	Natural population	15	MD		39.28	77.39	Acharya 2014
'T2086' (PI 476290)	Southern	Admixed, Lowland	Natural population	13	NC		34.22	77.94	Acharya 2014
'T2099' (PI 476291)	Southern	Upland	Natural population	8	MD		38.88	78.57	Acharya 2014
'T2100' (PI 476292)	Southern	Admixed, Upland	Natural population	12	AR		35.3	94.04	Acharya 2014
'T2101' (PI 476293)	Southern	Admixed	Natural population	10	NJ		39.22	74.99	Acharya 2014
'T4613' (PI 476294)	Southern	Upland	Natural population	11	CO		38.48	102.78	Acharya 2014
'T4614' (PI 476295)	Southern	Upland	Natural population	11	CO		39.04	104.63	Acharya 2014
'Talladega'	Supplemental Southern	Upland	Natural population	2	AL		32.961389	87.324444	This study
'Timber'	Northern	Lowland	Multisite synthetic	7	NC, SC, & GA		various	various	Lu et al. 2013
'Union'	Supplemental Southern	Admixed, Lowland	Natural population	5	MS		32.679925	89.211686	This study
'VS16 (Summer)'	Mapping population parent	Upland	Single genotype derived from Summer	1	SD		42.68	96.68	Missaoui et al. 2005
'White Plains'	Supplemental Southern	Admixed	Natural population	1	GA		33.13787	84.83723	This study
'Wilmer'	Supplemental Southern	Admixed, Lowland	Natural population	4	MS		30.82062	88.43385	This study
'Wolf Bay'	Supplemental Southern	Lowland	Natural population	4	AL		30.34547	87.62416	This study
'Woods'	Supplemental Southern	Admixed, Upland	Natural population	5	KS		37.16977	101.13034	This study
'WS4U'	Northern	Upland	Multisite synthetic	8	North Central US		various	various	Lu et al. 2013
'WS98-SB'	Northern	Admixed, Upland	Natural population	8	WI	280	45.08	92.83	Lu et al. 2013

† Ecotype. STRUCTURE shows the major ecotype (Upland or Lowland) from the STRUCTURE analysis. Individuals were classified into specific population groups if their  $q$ -value was  $>0.65$  for that group or Admixed if the  $q$ -value was  $<0.65$ . In this table, an Admixed population was further classified as Admixed-Upland or Admixed-Lowland if the individuals of that population, albeit with a low  $q$ -value, showed ancestry from only the Upland-specific or from only the Lowland-specific population groups, and Admixed-Admixed if the  $q$ -value was too low to be defined from only the Upland-specific population group or only the Lowland-specific population group or showed signal from both the Upland and Lowland population groups.

‡ Numbers in parentheses represents the number of individuals assayed from the Southern Switchgrass Association Panel.

total, the average depth was  $\sim 12.5\times$  for all positions with coverage; filtering of positions with less than  $5\times$  coverage resulted in an average depth of  $23.5\times$ . Read quality was assessed with FastQC 0.10.0 ([www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc), accessed 25 Oct. 2017). Reads were processed with CUTADAPT version 1.1 (Martin, 2011) and adaptor sequences and polymerase chain reaction primers removed, but bases with quality below 20 were trimmed and all reads shorter than 35 bases were discarded. The resulting reads were aligned to the hard-masked *P. virgatum* version 1.1 genome (phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\_Pvirgatum, accessed 25 Oct. 2017) using Bowtie version 0.12.7 (Langmead et al., 2009) in single-end mode. Unanchored contigs in the reference genome were assigned to “Chromosome Unknown” scaffolds (designated ChrUn1–ChrUn15) to accelerate processing. Only uniquely aligned reads with a maximum of a single mismatch in the first 35 nt were used for subsequent analysis.

### Single Nucleotide Polymorphism Detection and Analysis

Reads meeting the filtering and alignment criteria were sorted and indexed with the SORT and INDEX functions of the SAMTools software package (version 0.1.18) (Li et al., 2009). The SAMTools MPILEUP command with the `-BD` and `-C0` options was used to generate pileup files, which were then processed with custom Perl scripts to identify sequence polymorphisms. A position was determined as polymorphic if one individual at that position possessed a nonreference allele in at least two reads, with a minimum total depth of five reads, resulting in a set of 37,266,859 polymorphic loci. Alleles that were only represented by single reads were removed as sequencing errors. To generate the high-confidence set of loci used for phylogenetic and population structure analysis, the dataset was filtered to remove loci that were not biallelic, all loci with no reads in any individual, any position with fewer than five reads in 10 or more individuals, and all loci failing to meet the 5% or the minimum of two read polymorphisms. Additionally, at least two individuals were required to meet this criterion to retain the locus, resulting in a final set of 1,878,584 polymorphic loci.

### Population Structure Analysis

Population structure analysis was performed with the STRUCTURE program (Pritchard et al., 2000) using 35,857 loci randomly selected from HapMapv2 using the Perl `rand()` function. STRUCTURE was run for predicted numbers of populations ( $K$ -values) from 5 to 11, and was run six times for each  $K$ -value with a burn-in period of 10,000 iterations and 2000 Monte Carlo iterations. The `inferalpha`, `computeprubs`, and `freqscorr` corrections were set to 1 and all other options were left at the default values. Analysis was performed with the admixture model with no prior population knowledge. The results from each of the multiple  $K$ -runs were then aligned and merged with the CLUMPP version 1.1.2 (Jakobsson and

Rosenberg, 2007) and POPHELPER version 2.2.1 packages in the R version 3.3.2 environment (Francis, 2017; Jakobsson and Rosenberg, 2007). The best number of  $K$  clusters were determined using the  $\Delta K$  method [Eq. 1], defined as the ratio of the absolute value of the second-order rate of change of the likelihood distribution to the SD, as detailed by Evanno et al. (2005).

$$\Delta K = \frac{m(|L''K|)}{s[L(K)]} \quad [1]$$

The true value of  $K$  is the modal value of the distribution (Supplemental Fig. S1) was determined to be  $K = 9$  across seven distinct population groups with admixed proportions (Supplemental Fig. S2). Individuals were assigned to groups with a  $q$ -value threshold of  $>0.65$ .

To further characterize population differences between and within the lowland and upland ecotypes, the HapMapv2 SNP was further filtered by minor allele frequency ( $>0.01$ ), which resulted in a total of 472,452 SNPs. Estimates of population differentiation were computed between the lowland and upland ecotypes and for each pair of population groups based on Weir and Cockerham's  $F_{st}$  statistic (Weir and Cockerham, 1984) implemented in VCFtools version 0.1.12b (Danecek et al., 2011).

Phylogenetic analysis was performed on all SNPs in HapMapv2 using Phylyp 3.695 (Felsenstein, 1989). Genetic distances were calculated with the 'gendist' function with default parameters, and a neighbor-joining tree was created with the 'neighbor' function with default parameters. The resulting tree was visualized and colored according to their population group membership as determined by STRUCTURE via the `qqtree` package (Yu et al., 2017) implemented in the R environment.

Nucleotide diversity ( $\pi$ ) was calculated as described in Nei (1972) using Python scripts with the formula:

$$\pi = \sum_{ij} x_i x_j \pi_{ij} = 2 \times \sum_{i=2}^n \sum_{j=1}^{i-1} x_i x_j \pi_{ij} \quad [2]$$

A diversity index was calculated across all individuals from the three diversity panels jointly and separately from a subset that included individuals assigned to upland-only and lowland-only populations according to the STRUCTURE analysis. A position was considered heterozygous only if both alleles were represented by at least 25% of the reads; otherwise, the position was considered homozygous for the predominant allele. Ecotype was determined on the basis of the STRUCTURE results with individuals classified as admixed excluded from the ecotype-specific analysis. An allele was required to be represented by at least 25% of the reads at a position to be considered in the nucleotide diversity analysis.

Genome-wide identity-by-state pairwise distances for the  $1169 \times 1169$  individuals were computed on the filtered dataset using the `cluster-mds-plot` option in PLINK

(Chang et al., 2015; Purcell et al., 2007). To generate a more independent set of SNPs for multidimensional scaling analysis, the filtered dataset was further pruned by removing SNPs that were in linkage disequilibrium in a given window size of 50 SNPs, a window increment of five SNPs and a  $r^2 > 0.05$ . The MDS plot was redrawn with the ggplot (Wickham, 2009) package in R version 3.3.2.

### Population Group-Specific SNPs

To remove the variants likely to be caused by sequencing error, the less prevalent allele in a biallelic call was removed if it was supported by fewer than 12.5% of the reads overlapping the particular SNP in the particular sample. Single nucleotide polymorphisms that had no nonreference allele call in any individual that could be assigned to a population group were removed. Individuals were scored as present or absent for the nonreference allele at each SNP. Individuals that could not be placed into any population group were ignored. Filtering was done with two different sets of criteria. In one filter, SNPs were retained if the nonreference allele was present in 80% or more of the individuals in the population group(s) or ecotype being compared but present in only 20% or fewer of individuals in all other populations groups or ecotype. In another filter, SNPs were retained if the nonreference allele was found in any number of individuals but completely absent in all other comparators. Venn diagrams were drawn with the R package VennDiagram (Chen and Boutros, 2011) and manually edited.

ANNOVAR version 2016-02-01 (Wang et al., 2010) was used to identify genes with SNPs that were then filtered for SNPs specific for particular ecotypes or population groups with the 80–20% criteria described above. Gene Ontology (GO) enrichment was calculated with the web-based tool, agriGO version 1.2 (Du et al., 2010) and its pre-computed GO annotations for *P. virgatum* version 1.1 gene models. Enrichment was tested via Fisher's exact test with *P*-value adjustment via the false discovery rate (Benjamini and Yekutieli, 2001) at a significance level of 0.05. Clustering and visualization of enriched GO terms was performed with the web tool REVIGO (Supek et al., 2011) with the SimRel measure for semantic similarity and the whole UniProt database for calculating GO term sizes.

## Results and Discussion

### Exome Capture Sequencing and SNP Detection

Genetic analyses of exome capture sequencing data with the Northern Switchgrass Association Panel comprising 537 individuals (536 individuals plus the reference genotype AP13) from 66 populations (45 upland, 20 lowland, 1 unknown) have been previously described (Evans et al., 2015). To provide a broader representation of lowland switchgrass, a Southern Switchgrass Association Panel (Acharya, 2014), which was composed of 447 individuals (445 individuals plus 'AP13' and 'VS16') from 36 populations (15 upland, 18 lowland, 3 unknown) and a newly generated Supplemental Southern Switchgrass

Association Panel composed of 185 individuals from 45 populations (10 upland, 31 lowland, 4 unknown) (Table 1), was used. Four populations are common between the Northern and Southern Association panels and three populations are common between the Southern and Supplemental Southern panels (Table 1). The total composition of the combined panels represents 1169 individuals (1166 samples plus two AP13 reference genotype samples and one VS16 sample (parent of AP13 × VS16 mapping population) from 140 unique populations (67 upland, 65 lowland, 8 admixed) (Table 1). For the Southern Switchgrass Association Panel and Supplemental Southern Switchgrass Association panels, exome capture sequencing was performed as described previously (Evans et al., 2015) and combined with data from the Northern Switchgrass Association Panel (Evans et al., 2015) to call variants across the full set of 1169 individuals. Initial SNP detection yielded 37,266,859 positions with sequence polymorphisms; after filtering, 1,878,584 high-confidence biallelic loci remained (Supplemental File S1); this SNP set is referred to as the Switchgrass HapMapv2 dataset.

### Structure of Switchgrass Populations in the United States

Previous switchgrass diversity analyses attempted to identify and characterize genetic diversity across most of the range of switchgrass in the United States. These studies have shown that the Gulf Coast region is the center of diversity for switchgrass but also that previous characterizations were geographically incomplete (Evans et al., 2015; Lu et al., 2013; Zhang et al., 2011a, 2011b). Combining two existing panels (the Northern Association Switchgrass Panel and the Southern Association Switchgrass Panel) with a new supplemental southern association panel into a single diversity analysis that provides representation of wild populations, multisite synthetic populations, cultivars, and those from the USDA National Plant Germplasm System collection (Table 1) permitted a more robust estimation of the genetic diversity of switchgrass. Previous results from SNP-based population analyses based on mostly northern accessions indicated the presence of three upland and two lowland population groups (Evans et al., 2015). In this study, with  $K = 9$  clusters determined to be the most likely in the STRUCTURE analysis (Pritchard et al., 2000) (Supplemental Fig. S2), one population group was present with less than 1% probability in all samples, as observed previously by Evans et al. (2015) and is most probably caused by sequencing error. Of the remaining eight populations, four upland population groups, three lowland population groups, and one admixed population (Fig. 1) were identified. Of the 1169 individuals, 62 were classified as belonging to the Lowland Central population group, 134 to the Lowland North group, 212 to the Lowland South group, 101 to the Upland East group, 66 to the Upland Montane group, 156 to the Upland North group, and 160 to the Upland West group, with 278 considered admixed (Supplemental File S2).

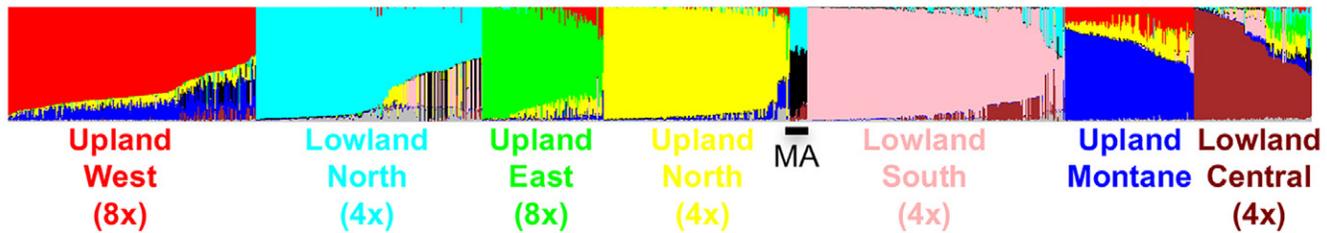


Fig. 1. Identification of seven population groups in US switchgrass. A subset of the 1.9 million single nucleotide polymorphisms (SNPs) from HapMapv2 were subjected to STRUCTURE analysis (Pritchard et al., 2000), resulting in the identification of seven population groups that correspond to ecotype, ploidy, and geographic distribution. Upland population groups (red, green, yellow, and blue) and lowland population groups (aqua, pink, maroon) show varying degrees of admixture. Black represents the admixed population group. Mixed ancestry (MA) represents individuals with ancestry from the Lowland North group and an unknown group, but with a  $q$ -value threshold of  $<0.65$ .

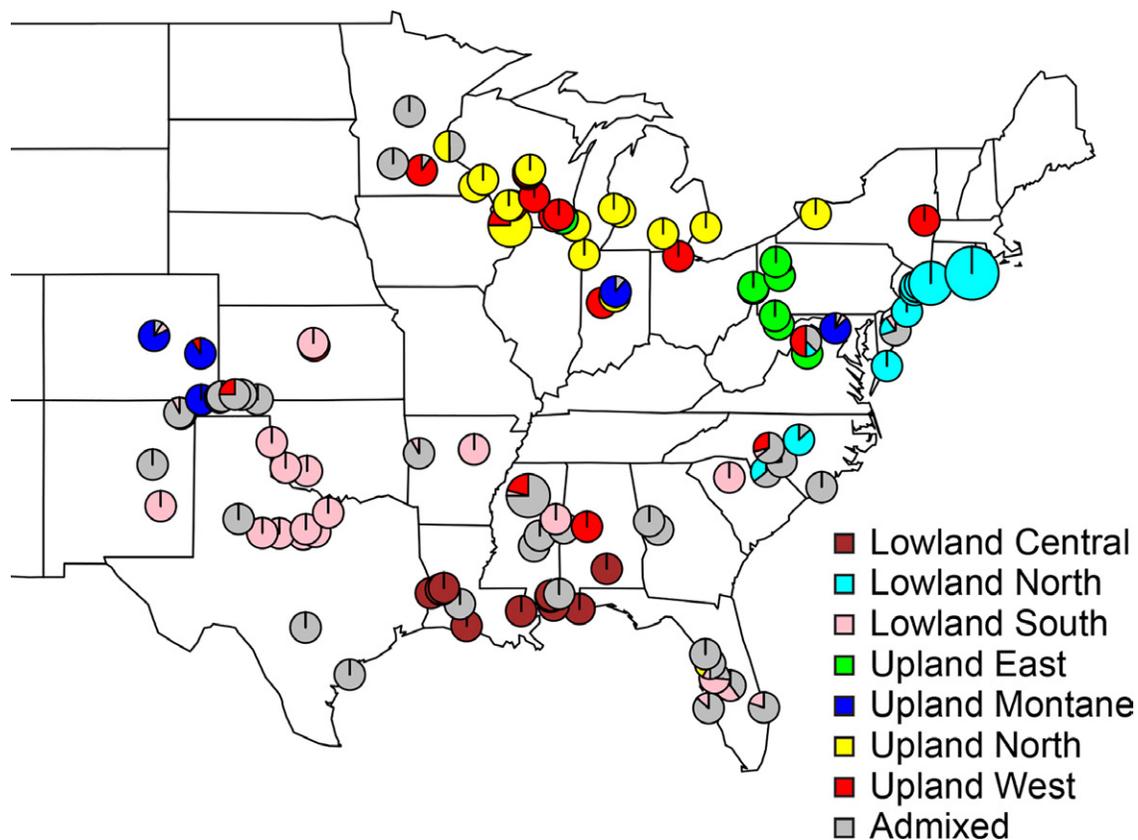


Fig. 2. Distribution of switchgrass population groups across the United States. Pie charts represent 117 populations and their per-individual population group membership as determined by STRUCTURE (Pritchard et al., 2000) and were plotted on the basis of geographical distribution. Colors were derived from population groups identified by STRUCTURE as shown in Fig. 1 (Upland West, red; Upland East, green; Upland North, yellow; Upland Montane, blue; Lowland North, aqua; Lowland South, pink; Lowland Central, maroon; admixed, black). Circle sizes reflect merged population groups that could not be depicted individually because of their overlapping geographic origins. A subset of populations (23) were omitted from the figure because their global positioning system coordinates are not available or could not be verified, or they were synthetic populations or transplant populations.

Two of the lowland population groups overlap with those previously identified (Evans et al., 2015; Lu et al., 2013), representing population groups from the eastern US seaboard (Lowland North) and the southern United States (Lowland South). These population groups have a larger representation in the two southern association panels included in this study. The new lowland population group (Lowland Central) represents accessions from

the south central United States and the Florida peninsula (Fig. 2). Two population groups (Upland East and Upland North) overlap with previously identified population groups (Evans et al., 2015; Lu et al., 2013). With the increased samples in this study, the Upland West population group (Evans et al., 2015) has been separated into two new population groups (Upland Montane and Upland West).

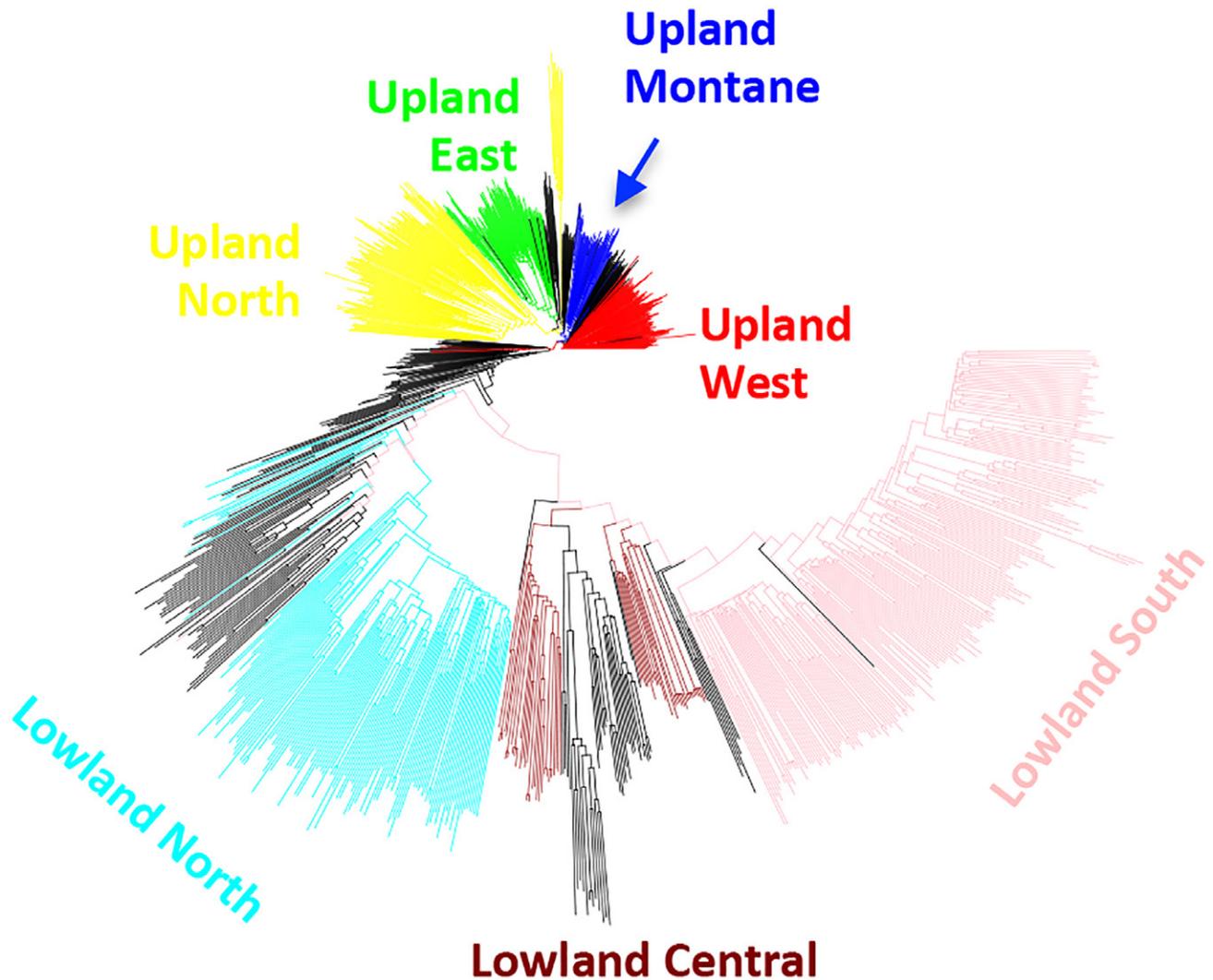


Fig. 3. Genetic distance dendrogram for 1169 switchgrass individuals examined in this study. With the neighbor-joining method in Phylip, 1.9 million HapMapv2 single nucleotide polymorphisms (SNPs) were used to measure the genetic distance among all 1169 individuals from the three association panels. Individuals are colored according to their population group assignment as shown in Fig. 1 (Upland West, red; Upland East, green; Upland North, yellow; Upland Montane, blue; Lowland North, aqua; Lowland South, pink; Lowland Central, maroon; admixed, black) in which an individual had to have >65% membership in a population group to be assigned to that group. Four distinct upland population groups and three lowland population groups are evident, with substantially higher genetic distances within lowland individuals and population groups than those observed in upland population groups and individuals.

To verify the results of the STRUCTURE analysis, the full HapMapv2 SNP set was used to calculate the genetic distance between all members of the combined switchgrass panels from which a neighbor-joining tree was constructed (Fig. 3, Fig. S3). Consistent with the STRUCTURE results, members from each upland population group formed a tight cluster that was separate from the three lowland population groups, which formed a far larger and more distributed cluster (Fig. 3). Upland population groups showed a much lower level of genetic distance between individuals and between populations than the lowland population groups, consistent with the hypothesis that upland switchgrass is derived from lowland switchgrass and originated from a smaller, more recent pool of germplasm (Zhang et al., 2011a, 2011b). The seven population groups and the admixed samples

also show cleared differentiation when plotted with a multidimensional scaling analysis of the linkage disequilibrium-pruned dataset (Fig. 4).

The amount of genetic differentiation between the two major ecotypes and the population groups was also computed via Weir and Cockerham's  $F_{st}$  estimates (Weir and Cockerham, 1984). Moderate differentiation with a mean  $F_{st}$  estimate of 0.06 (Table 2) was observed between lowland and upland population groups. Within the lowland population groups,  $F_{st}$  values ranged from 0.05 to 0.07, with the central and southern lowland populations being more closely related ( $F_{st} = 0.05$ ). In contrast,  $F_{st}$  values within the upland population groups showed only little genetic differentiation between populations, with  $F_{st}$  values ranging from 0.02 to 0.04 (Table 2).

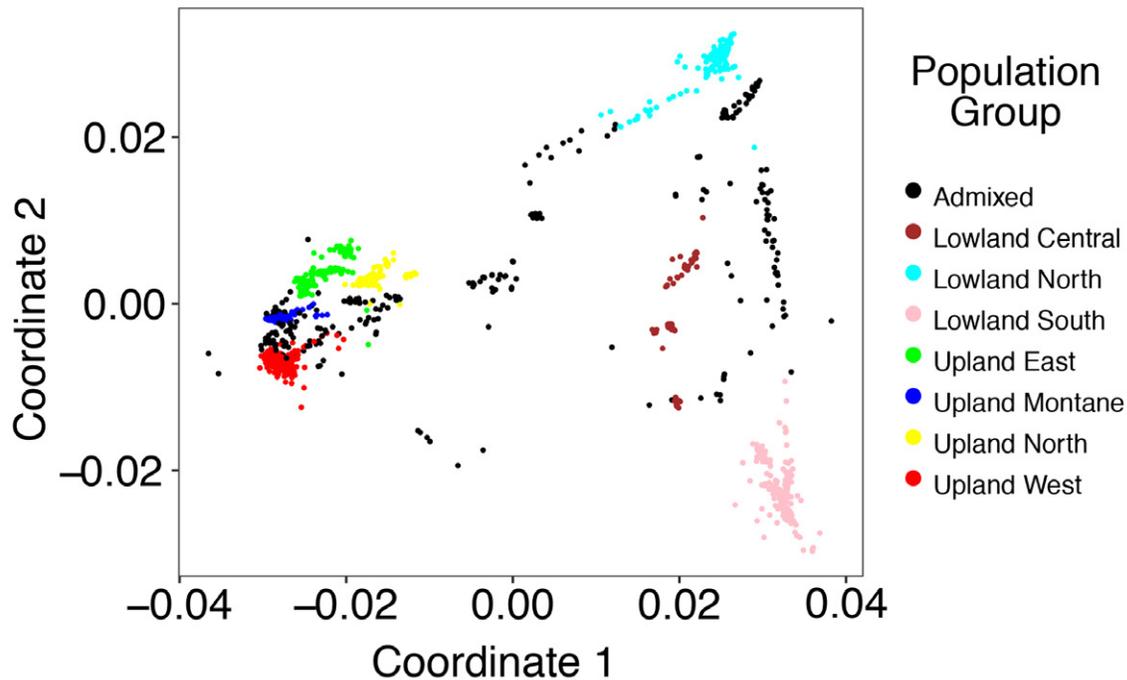


Fig. 4. Multidimensional scaling plot of the genetic distances of switchgrass individuals from the three diversity panels. Pairwise distances for all 1169 individuals were calculated and plotted with the ggplot package in R (Wickham, 2009). Color coding is the same as that for the seven populations identified by STRUCTURE in Fig. 1: Upland West, red; Upland East, green; Upland North, yellow; Upland Montane, blue; Lowland North, aqua; Lowland South, pink; Lowland Central, maroon; admixed, black.

**Table 2. Pairwise fixation index values for each pair of switchgrass population groups from a  $K = 9$  grouping from the STRUCTURE analyses, wherein group membership was defined with a  $q$ -value of  $>0.65$ .**

	Lowland Central	Lowland North	Lowland South	Upland East	Upland Montane	Upland North	Upland West	Admixed
Lowland Central	—	0.055	0.048	0.070	0.073	0.071	0.070	0.028
Lowland North	—	—	0.069	0.076	0.085	0.070	0.077	0.036
Lowland South	—	—	—	0.087	0.096	0.072	0.078	0.048
Upland East	—	—	—	—	0.043	0.030	0.044	0.038
Upland Montane	—	—	—	—	—	0.034	0.023	0.027
Upland North	—	—	—	—	—	—	0.033	0.029
Upland West	—	—	—	—	—	—	—	0.028

Upland switchgrass evolved as a separate ecotype from the lowland type about 1 to 1.3 million years ago (Huang et al., 2003; Zhang et al., 2011b). During this time, there have been 10 to 12 continental glaciation events, each of which has compressed the range of the species to a narrow geographic region (Bintanja and van der Wal, 2008). The center of diversity for switchgrass may shift slightly between glacial maxima and minima, but it has remained essentially in the Gulf Coast region for well over a million years. Following each glacial maximum, the glaciers retreated and temperate habitats reappeared, allowing plants and animals to gradually recolonize North America from the Gulf Coast to the Arctic Ocean, a process that required thousands of years and was accomplished only in fits and starts. We hypothesize that the upland ecotype of switchgrass was the only variant that was sufficiently hardy to survive this punctuated recolonization process.

The early flowering trait of upland switchgrass was one of the most important adaptive traits for the recolonization process to move so far north. Indeed, when evaluated in common-garden experiments near 42°N latitude, local upland ecotypes are up to 6 wk earlier in flowering time than southern lowland ecotypes (Casler, 2012, Casler et al., 2012). Ploidy may also have played a role in this process but it was not solely responsible for the ability of upland switchgrass to colonize the humid-temperate regions of North America, as evidenced by the small frequency of tetraploid upland populations (Lu et al., 2013, Zhang et al., 2011b). Although this study did not include extensive ploidy determination, several previous reports have shown that tetraploid accessions are fairly rare in the northern range of switchgrass (Lu et al., 2013, Zhang et al., 2011b), suggesting that the octoploid form was better able to colonize northern habitats across a broad geographic range, as they are hypothesized to be more adaptive (Stebbins, 1985, Symonds et al., 2010).

## Population Structure is Largely Explained by Geography

The seven distinct population groups identified in this study can be well explained by their geographic distribution. The majority of the populations included in this dataset have all of their members (individuals sampled from the same location) identified with one ecotype (a threshold  $q$ -value of  $>0.65$  in the STRUCTURE analysis); relaxing this threshold to  $q > 0.5$  would have resolved most of the individuals classified as admixed to one of the subgroups. For some admixed populations, all members reflect shared relationships with both ecotypes, mostly from adjacent population groups. For example, members of ‘Bullis’, ‘PMT-785’, and ‘Mustang-Lake’ have proportions of ancestry from the Lowland South and Lowland Central groups, which could have been a result of their geographic proximity with gene flow via seed dispersal. Some individuals from the Lowland North group also have shared ancestry with the Lowland South group, which could have been caused by switchgrass migrating from the south to the northeast via shipments of dry hay (Zhang et al., 2011b). Conversely, individuals from Lowland North are isolated from upland switchgrass by the Appalachian Mountains, and this geographic barrier could have prevented gene flow between these groups. The close genetic distance between the upland populations can be accounted for by their shared genetic variations when switchgrass migrated from the south. The repeated vicariant events have impacted the gene flow in the upland switchgrass populations with consequent differences in ploidy and in flowering time being factors of reproductive isolation (Grabowski et al., 2017; Zhang et al., 2011b).

This dataset also includes populations known to have been selected for breeding; these have clustered together with their respective ecotypes as expected. ‘Sunburst’ is a bred cultivar with ancestry mostly from the Upland North and Upland Montane groups. ‘Cave-in-Rock’, ‘Shelter’, and ‘KY1625’ are natural track cultivars mainly from the Upland East group. ‘Blackwell’ is from the Upland West group. ‘Carthage’ is admixed, but the majority of its ancestry can be traced to the Upland West group (the  $q$ -values from STRUCTURE analysis ranged from 0.42 to 0.89). ‘SW789’ is a multisite synthetic with mixed ancestry mostly from the Lowland Central, Lowland South, and Upland North groups. The results of the STRUCTURE (Fig. 1) analysis also recovered a population group with mixed ancestry comprised mainly of the Lowland North group and an unknown population group that includes individuals from the ‘OSSP-FL’, ‘BN-11357–63’, ‘Pasco-Co-FL’, ‘HSP-FL’, ‘Stuart’, ‘SWFWMD-FL’, and ‘Citrus-Co-FL’ populations. Sampling more individuals from these populations and around their surrounding regions would provide more resolution about this enigmatic population group.

**Table 3. Nucleotide diversity values for upland, lowland, and all switchgrass individuals.**

Species	Diversity value†
<i>Panicum virgatum</i>	
Upland individuals	0.0102
Lowland individuals	0.0134
All individuals	0.0135
<i>Glycine max</i> (cultivated)	0.0019
<i>Glycine max</i> (wild)	0.0030
<i>Medicago truncatula</i>	0.0043
<i>Oryza sativa indica</i>	0.0016
<i>Oryza sativa japonica</i>	0.0006
<i>Oryza sativa</i> (landraces)	0.0024
<i>Zea mays</i> (improved)	0.0048
<i>Zea mays</i> (landraces)	0.0049

† Diversity values were obtained from Branca et al. (2011) for *Medicago truncatula*; Huang et al. (2010) for rice, and soybean; and Hufford et al. (2012) for maize.

## Genetic Diversity Estimations

Nucleotide diversity (Nei, 1972) was calculated for the HapMapv2 SNP set (Table 3). As a whole, the switchgrass individuals had a diversity value of 0.0135, which is substantially higher than those reported for *M. truncatula* (Branca et al., 2011), rice (Huang et al., 2010), soybean (Lam et al., 2010), and maize (Hufford et al., 2012), reflective of the polyploid and outcrossing nature of switchgrass as well as the highly undomesticated state of the majority of populations examined in this study. In switchgrass breeding programs, natural populations are selected for improved agronomic traits with limited cycles of selection or breeding (Casler et al., 2012). Notably, upland switchgrass had substantially lower nucleotide diversity than lowland switchgrass (Table 3), consistent with the hypothesis of decreased genetic diversity in more northern latitudes (Hewitt, 1996; Soltis et al., 1997). The reduced genetic diversity in upland switchgrass is also consistent with a population bottleneck, which could explain the current distribution of natural populations attributable to repeated North American glaciation events (Zhang et al., 2011b).

## Genetically Distinct Individuals within Populations and Diagnostic Polymorphisms

Switchgrass breeding populations consist of individual plants collected from the same geographical site and tend to be similar genetically. However, this work identified multiple populations with genetic outliers that clustered with different population groups and sometimes even different ecotypes (Supplemental Table S1). These outlier individuals were observed mainly in populations from the Southern Switchgrass Association Panel but were also present in the ‘High-Tide’, ‘SWG39’, and ‘SW790’ populations from the Northern Switchgrass Association Panel. Seed sources for establishment of native prairie or savanna habitats are wind, birds, and mammals (Cheplick, 1998). As such, remnant prairie or savanna sites may contain unexpected levels of diversity that can include

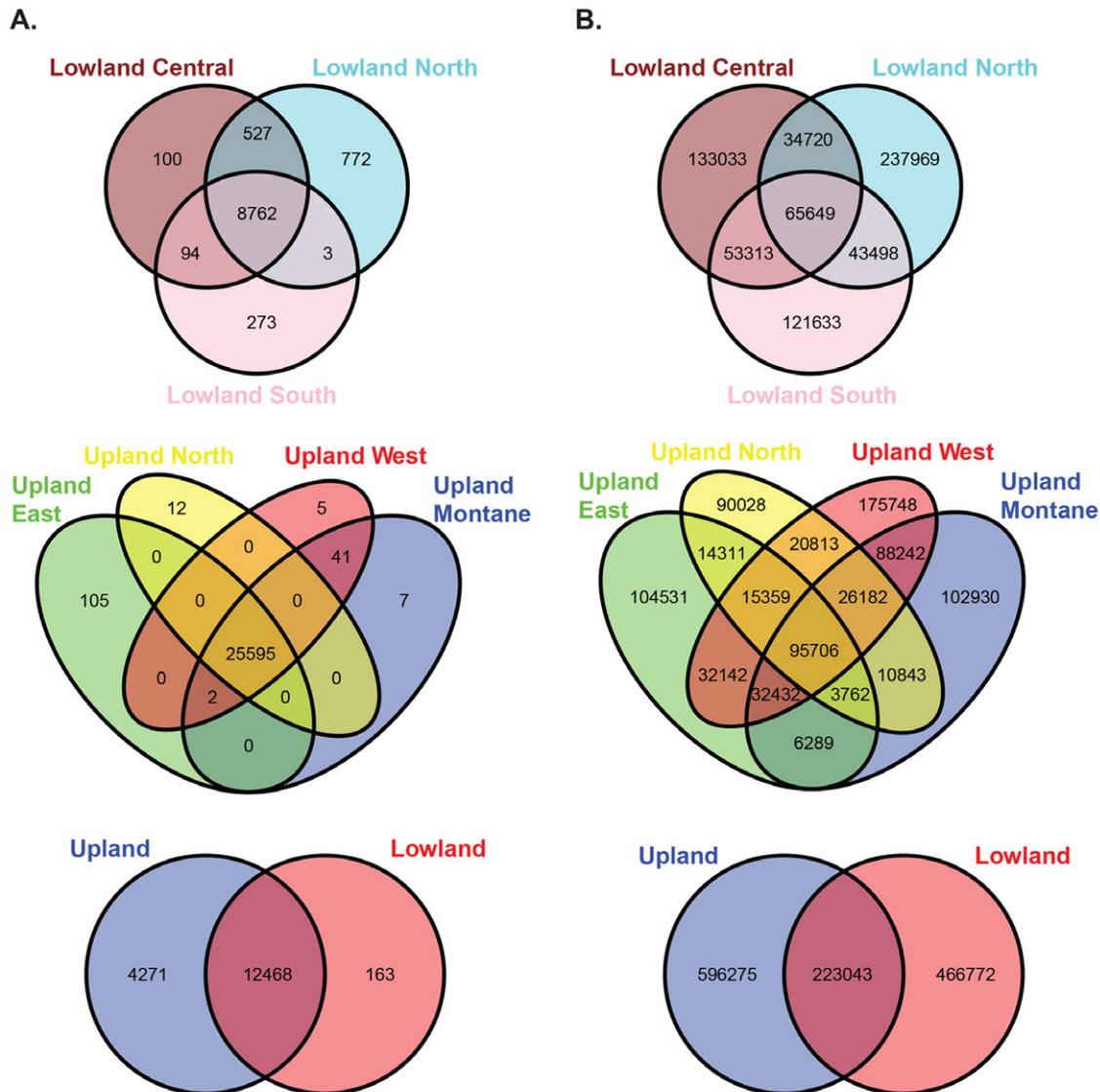


Fig. 5. Single nucleotide polymorphisms (SNPs) with the presence or absence of genotypes specific for certain switchgrass population groups or ecotypes. Comparisons were made for (top) lowland groups, (middle) upland groups, and (bottom) between lowland and upland ecotypes. (A) The SNPs were considered to be specific for an ecotype or population group(s) if the nonreference allele was present in  $\geq 80\%$  of the individuals but only  $\leq 20\%$  of other groups. (B) The SNPs were considered specific for an ecotype or population group(s) if the nonreference allele was present in at least one individual but completely absent in other groups.

multiple ploidies, multiple ecotypes, or genotypes with a range of adaptive traits (Lu et al., 2013; Zhang et al., 2011a, 2011b). Of course, the off-types may represent stock contamination such as that of the Alamo population, which was obtained from the National Plant Germplasm System; this has since been rectified. These off-types may also reflect contamination during sample collection or mislabeling at the time of planting the experiment or when samples were taken for DNA extraction, or could represent pollen contamination during seed propagation, a plausible scenario, given the tendency of switchgrass to outcross. Further work will be required to identify the cause of these heterogeneous switchgrass populations.

To facilitate development of informative markers for future efforts with switchgrass genetics and breeding, SNPs were identified where the alternate allele was

found frequently in specific population groups ( $\geq 80\%$  of individuals) but rarely in others ( $\leq 20\%$  of individuals) (Fig. 5A; Supplemental File S3). These initial filtering criteria resulted in very few population group-specific SNPs in comparisons between upland population groups (Fig. 5A), consistent with the reduced genetic distance (Fig. 3 and Fig. 4) and sequence diversity (Table 2) within the upland populations. Correspondingly, more specific SNPs were identified in comparisons between lowland population groups and between upland and lowland ecotypes, reflecting the greater genetic distance between these comparators. The Lowland North and Lowland South groups had higher levels of specific SNPs than the Lowland Central group, consistent with tight genetic clustering of their individuals (Fig. 4). Overall, filtering for  $\geq 80\%$  presence in one group and  $\leq 20\%$

in other groups removed many SNPs (the total was 1,286,090 SNPs after removing those supported by very few reads) (Fig. 5A) because occurrence of the alternate allele tended to be low. In contrast, filtering instead for the presence in a population group at any frequency and complete absence in any other group retained many more SNPs (Fig. 5B), suggesting that most SNPs within these switchgrass panels are recent mutations occurring after the formation of ecotypes and population groups.

Among the SNPs present in  $\geq 80\%$  in upland individuals and  $\leq 20\%$  in lowland individuals, significant enrichment was observed for multiple GO terms (Fig. 6; Supplemental File S4). Of special interest were genes associated with the GO terms ‘response to hormone’, ‘protein–chromophore linkage’, and ‘phospholipid metabolism’ (Supplemental Table S2). Interestingly, all of the enriched genes associated with the ‘response to hormone’ term are homologs of *Arabidopsis thaliana* (L.) Heynh. auxin response factors (ARF), which represent all genes within this GO category (Supplemental Table S2; Supplemental File S4). Mutant studies in *A. thaliana* have elucidated the functions of several ARFs that span a wide range of auxin-regulated biological processes, some of which could affect upland switchgrass adaptation. For example, in *A. thaliana*, ARF2 limits seed and leaf size by reducing cell divisions, shortens flowering time, and promotes the onset of leaf senescence (Li et al., 2004; Lim et al., 2010; Schruoff et al., 2006), whereas ARF7 functions redundantly with ARF19 to increase leaf size by maintaining large cell size (Wilmoth et al., 2005). Thus the switchgrass ARF homologs may be associated with phenological and morphological differences observed between upland and lowland populations.

The genes annotated as ‘protein–chromophore linkage’ are homologs of *PHYTOCHROME A*, *PHYTOCHROME B*, and *PHYTOCHROME C* in *A. thaliana* (Supplemental Table S2), which are proteins that mediate responses to light signals. In rice, double-mutant combinations of *phyA*, *phyB*, and *phyC* reduce flowering time by  $\sim 40\%$  when grown in the field under long daylength (Takano et al., 2005). Furthermore, natural variation in *PHYTOCHROME C* in *A. thaliana* is associated with flowering time variation, and *PHYTOCHROME C* haplotypes are correlated with latitudinal clines (Balasubramanian et al., 2006). The involvement of phytochrome genes in the control of flowering time in upland switchgrass is consistent with a recent switchgrass genome-wide association study that indicated a role for the photoperiod response gene *FLOWERING LOCUS T*, in flowering time and a correlation of flowering time with habitat latitude, suggesting an important role for photoperiod sensitivity in controlling flowering time (Grabowski et al., 2017).

Enrichment of genes associated with the ‘phospholipid metabolism’ term could involve cold adaptation in upland switchgrass. Cold temperatures reduce membrane fluidity and freezing can destabilize cell membranes. Alterations to cell membrane composition is a component of cold acclimation in plants (Uemura et al., 1995). Enrichment

in two genes encoding aminoalcoholphosphotransferase homologs (Supplemental Table S2) may be associated with optimized cell membrane composition in upland switchgrass. Two major components of cell membranes are phosphatidylcholine and phosphatidylethanolamine, two phospholipids synthesized by aminoalcoholphosphotransferase. Overexpression of *aminoalcoholphosphotransferase* in *Brassica napus* L. leads to more cold-tolerant plants, and *A. thaliana aminoalcoholphosphotransferase* is induced by cold treatment and abscisic acid, a phytohormone involved in cold response (Qi et al., 2003). Future work could test (e.g., by targeted mutagenesis) whether these genes are involved in conferring traits associated with uplands such as smaller plant size, earlier flowering time, and cold tolerance.

## Conclusions

This study expands on previous assessments of genetic diversity in switchgrass by including 632 individuals from two additional southern switchgrass association panels to complement the 537 individuals from the northern association panel, thereby providing a more robust assessment of North American switchgrass. Using exome capture sequencing, a high-density SNP set (HapMapv2) containing  $\sim 1.9$  million SNPs was generated. Coupled with the broader representation of switchgrass germplasm in these three association panels, this study supported previously identified population groups in North American switchgrass while permitting the discovery of two new population groups. These results also demonstrated the extent of diversity within lowland versus upland switchgrass, with the Gulf Coast region being the center of diversity for switchgrass, whereas the limited diversity within upland switchgrass is suggestive of a genetic bottleneck. This study further supports the hypothesis that repeated glaciation events, ploidy barriers, and restricted gene flow caused by flowering time differences have resulted in distinct gene pools across ecotypes and geographic regions. These data can be used to guide breeding and restoration efforts, understand the genes and molecular mechanisms involved in important traits relevant to switchgrass production and adaptation, and further the understanding of the evolution of extant switchgrass diversity.

## Data Access

All sequencing reads from the Southern Switchgrass Association Panel and the Supplemental Southern Switchgrass Association Panel used in this study are available from the National Center for Biotechnology Information under project BioProject PRJNA324429 (to be made available on publication). Reads from the previously genotyped Northern Switchgrass Association Panel are available in the National Center for Biotechnology Information under BioProject PRJNA280418. The HapMapv2 matrix and genetic distance dendrogram files in Newick format are available on the Dryad Digital Repository (10.5061/dryad.mp6cp).

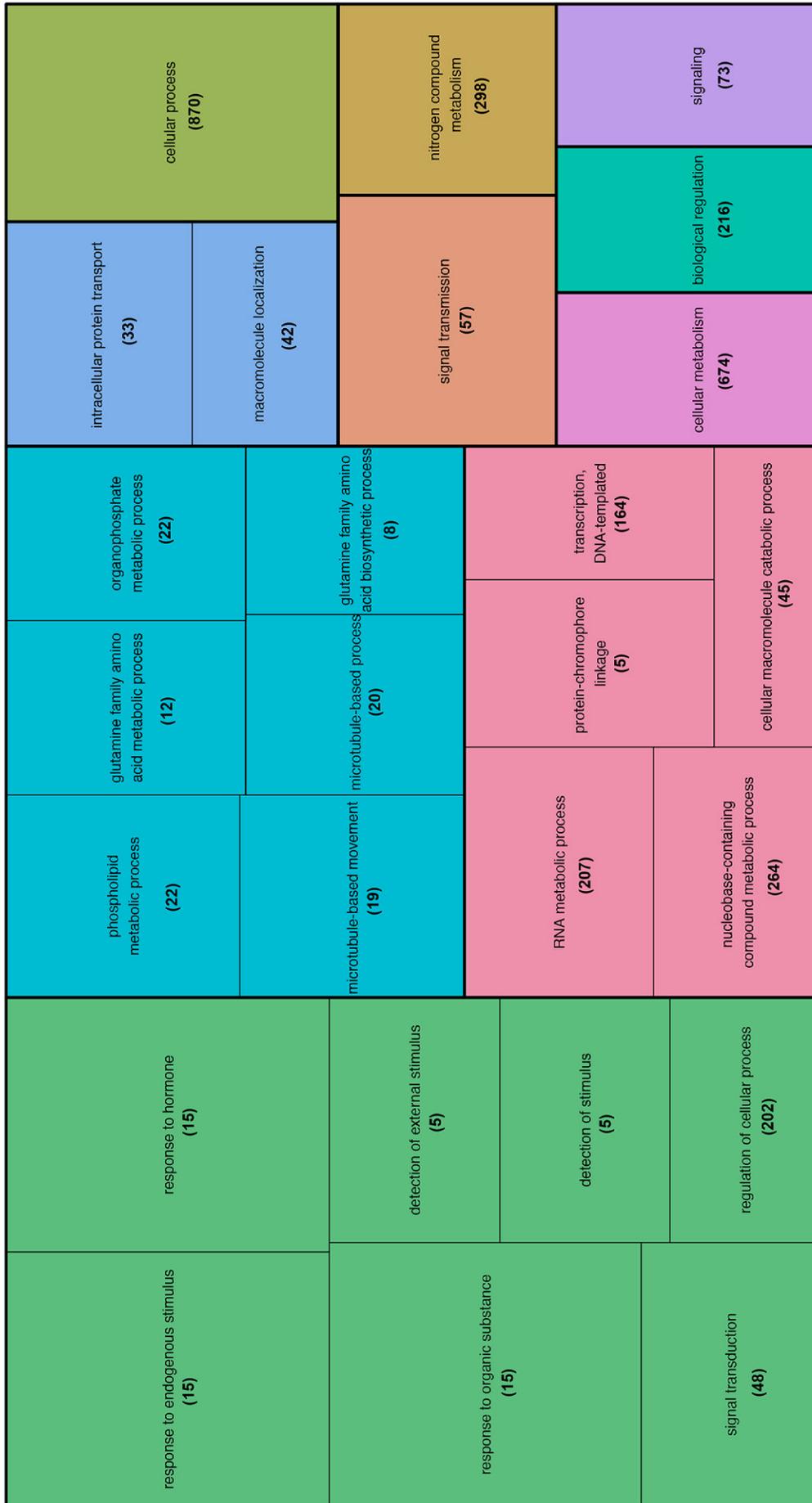


Fig. 6. Enriched gene ontology (GO) biological processes for switchgrass single nucleotide polymorphisms (SNPs) present in >80% of upland individuals and < 20% of lowland individuals after GO redundancy reduction with REVIGO (Supek et al., 2011). Each box represents a cluster of highly semantically similar GO terms and boxes of the same color indicate loosely related GO terms. The size of each box is proportional to  $-\log_{10}$  of its false discovery rate-adjusted  $p$ -value for enrichment. Numbers in parentheses indicate the number of genes.

## Supplemental Information

Supplemental Table S1. Off-type individuals.

Supplemental Table S2. A subset of the significant gene ontology terms calculated from the genes that overlap single nucleotide polymorphisms that are present in 80% of upland individuals but less than 20% in lowland individuals.

Supplemental Fig. S1. The absolute value of the second-order rate of change of the likelihood distribution ( $|L''(K)| = |L'(K+1) - L'(K)|$ ) (left) and

$$\Delta K (K = \frac{m(|L''(K)|)}{s[L(K)]}) \text{ (right) show that the modal}$$

value of the distribution is  $K = 9$  in this study.

Supplemental Fig. S2. Population structure differentiation from the STRUCTURE analysis for  $K$  groups 5 to 11. Each plot shows the consensus run generated by CLUMPP version 1.1.2 from multiple repeats for each  $K$ -value from the STRUCTURE analysis, with default colors from the POPHELPER version 2.2.1 package showing distinct groups.

Supplemental Fig. S3. The same neighbor-joining tree as shown in Fig. 3 computed with the 1.9 million HapMapv2, with 1169 individuals as tip labels, from the three association panels and presented in ladderized form. Individuals are colored according to the population group assignment as shown in Fig. 1 and Fig. 3, with Upland West (red), Upland East (green), Upland North (yellow), Upland Montane (blue), Lowland North (aqua), Lowland South (pink), Lowland Central (maroon), and admixed (black).

Supplemental File S1. HapMapv2 SNP calls in 1169 accessions of switchgrass. This file is ~2 GB and will be made available via the Dryad Digital Repository on publication.

Supplemental File S2. Membership of individuals in the major population groups. Available as a compressed Excel file.

Supplemental File S3. Significantly enriched GO terms and associated switchgrass genes. Available as a compressed Excel file.

Supplemental File S4. Significantly enriched GO terms for genes overlapping population groups or ecotype-specific single nucleotide polymorphisms based on the 80–20% criteria. Available as a compressed Excel file.

## Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

## Acknowledgments

This work was funded by the Department of Energy (DOE) Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE BioEnergy Science Center (DOE BER Office of Science DE-AC05-00OR22725) through subcontracts to the Noble Research Institute and University of Georgia. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US DOE under Contract No. DE-AC02-05CH11231.

## References

- Acharya, A.R. 2014. Genetic diversity, population structure, and association mapping of biofuel traits in southern switchgrass germplasm. Univ. of Georgia, Athens, GA.
- Balasubramanian, S., S. Sureshkumar, M. Agrawal, T.P. Michael, C. Wessinger, J.N. Maloof, et al. 2006. The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat. Genet.* 38:711–715. doi:10.1038/ng1818
- Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29:1165–1188.
- Bintanja, R., and R.S.W. van der Wal. 2008. North American ice-sheet dynamics and the onset of the 100,000-year glacial cycles. *Nature* 454:869–872. doi:10.1038/nature07158
- Branca, A., T.D. Paape, P. Zhou, R. Briskine, A.D. Farmer, J. Mudge, et al. 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* 108:864–870.
- Casler, M.D. 2005. Ecotypic variation among switchgrass populations from the northern USA. *Crop Sci.* 45:388–398. doi:10.2135/cropsci2005.0388
- Casler, M.D. 2012. Switchgrass breeding, genetics, and genomics. In: A. Monti, editor, *Switchgrass*. Springer, New York. p. 29–53.
- Casler, M.D., R.B. Mitchell, and K.P. Vogel. 2012. Switchgrass. In: C. Kole et al., editors, *Handbook of bioenergy crop plants*. Taylor & Francis, New York. p. 563–590.
- Casler, M.D., K.P. Vogel, C.M. Taliaferro, N.J. Ehlke, J.D. Berdahl, E.C. Brummer, et al. 2007. Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Sci.* 47:2249–2260. doi:10.2135/cropsci2006.12.0780
- Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* 4:1–16. doi:10.1186/s13742-015-0047-8
- Chen, H., and P.C. Boutros. 2011. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. doi:10.1186/1471-2105-12-35
- Cheplick, G.P. 1998. Seed dispersal and seedling establishment in grass populations. In: G.P. Cheplick, editor, *Population biology of grasses*. Cambridge Univ. Press, Cambridge, UK. p. 84–105.
- Costich, D.E., B. Friebe, M.J. Sheehan, M.D. Casler, and E.S. Buckler. 2010. Genome-size variation in switchgrass (*Panicum virgatum*): Flow cytometry and cytology reveal rampant aneuploidy. *Plant Genome* 3:130–141. doi:10.3835/plantgenome2010.04.0010
- Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi:10.1093/bioinformatics/btr330
- Du, Z., X. Zhou, Y. Ling, Z. Zhang, and Z. Su. 2010. agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38:W64–70. doi:10.1093/nar/gkq310
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Evans, J., E. Crisovan, K. Barry, C. Daum, J. Jenkins, G. Kunde-Ramamoorthy, et al. 2015. Diversity and population structure of northern switchgrass as revealed through exome capture sequencing. *Plant J.* 84:800–815. doi:10.1111/tpj.13041
- Evans, J., J. Kim, K.L. Childs, B. Vaillancourt, E. Crisovan, A. Nandety, et al. 2014. Nucleotide polymorphism and copy number variant detection using exome capture and next-generation sequencing in the polyploid grass *Panicum virgatum*. *Plant J.* 79:993–1008. doi:10.1111/tpj.12601
- Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* 5:164–166.
- Francis, R.M. 2017. Pophelper: An R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 17:27–32. doi:10.1111/1755-0998.12509

- Grabowski, P.P., J. Evans, C. Daum, S. Deshpande, K.W. Barry, M. Kennedy, et al. 2017. Genome-wide associations with flowering time in switchgrass using exome-capture sequencing data. *New Phytol.* 213:154–169. doi:10.1111/nph.14101
- Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc. Lond.* 58:247–276. doi:10.1111/j.1095-8312.1996.tb01434.x
- Huang, S., X. Siu, R. Haselkorn, and P. Gornicki. 2003. Evolution of switchgrass (*Panicum virgatum* L.) based on sequences of the nuclear gene encoding plastic acetyl-CoA carboxylase. *Plant Sci.* 164:43–49. doi:10.1016/S0168-9452(02)00327-8
- Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42:961–967. doi:10.1038/ng.695
- Hufford, M.B., X. Xu, J. van Heerwaarden, T. Pyhajarvi, J.M. Chia, R.A. Cartwright, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44:808–811. doi:10.1038/ng.2309
- Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806. doi:10.1093/bioinformatics/btm233
- Lam, H.M., X. Xu, X. Liu, W. Chen, G. Yang, F.L. Wong, et al. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42:1053–1059. doi:10.1038/ng.715
- Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi:10.1186/gb-2009-10-3-r25
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. doi:10.1093/bioinformatics/btp352
- Li, H., P. Johnson, A. Stepanova, J.M. Alonso, and J.R. Ecker. 2004. Convergence of signaling pathways in the control of differential cell growth in *Arabidopsis*. *Dev. Cell* 7:193–204. doi:10.1016/j.devcel.2004.07.002
- Lim, P.O., I.C. Lee, J. Kim, H.J. Kim, J.S. Ryu, H.R. Woo, et al. 2010. Auxin response factor 2 (ARF2) plays a major role in regulating auxin-mediated leaf longevity. *J. Exp. Bot.* 61:1419–1430. doi:10.1093/jxb/erq010
- Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, et al. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215. doi:10.1371/journal.pgen.1003215
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12. doi:10.14806/ej.17.1.200
- Mascher, M., T.A. Richmond, D.J. Gerhardt, A. Himmelbach, L. Clissold, D. Sampath, et al. 2013. Barley whole exome capture: A tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* 76:494–505. doi:10.1111/tj.12294
- Missaoui, A.M., Paterson, A.H., Bouton, J.H. 2005. Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. *Theor. Appl. Genet.* 110:1372–1383. doi:10.1007/s00122-005-1935-6
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* 106(949):283–292. doi:10.1086/282771
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575. doi:10.1086/519795
- Qi, Q., Y.F. Huang, A.J. Cutler, S.R. Abrams, and D.C. Taylor. 2003. Molecular and biochemical characterization of an aminoalcoholphosphotransferase (AAPT1) from *Brassica napus*: Effects of low temperature and abscisic acid treatments on AAPT expression in *Arabidopsis* plants and effects of over-expression of *BnAAPT1* in transgenic *Arabidopsis*. *Planta* 217:547–558. doi:10.1007/s00425-003-1031-6
- Ramstein, G.P., J. Evans, S.M. Kaeppler, R.B. Mitchell, K.P. Vogel, C.R. Buell, et al. 2016. Accuracy of genomic prediction in switchgrass (*Panicum virgatum* L.) improved by accounting for linkage disequilibrium. *G3 (Bethesda)* 6:1049–1062. doi:10.1534/g3.115.024950
- Saghai-Marouf, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley—Mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad USA* 81:8014–8018. doi:10.1073/pnas.81.24.8014
- Schruff, M.C., M. Spielman, S. Tiwari, S. Adams, N. Fenby, and R.J. Scott. 2006. The *AUXIN RESPONSE FACTOR 2* gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133:251–261. doi:10.1242/dev.02194
- Sharma, M.K., R. Sharma, P.J. Cao, J. Jenkins, L.E. Bartley, M. Qualls, et al. 2012. A genome-wide survey of switchgrass genome structure and organization. *PLoS One* 7(4): e33892.
- Soltis, D.E., M.A. Gitzendanner, D.D. Strenge, and P.E. Soltis. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Syst. Evol.* 206:353–373. doi:10.1007/BF00987957
- Stebbins, G.L. 1985. Polyploidy, hybridization, and the invasion of new habitats. *Ann. Mo. Bot. Gard.* 72:824–832. doi:10.2307/2399224
- Supek, F., M. Bosnjak, N. Skunca, and T. Smuc. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:E21800. doi:10.1371/journal.pone.0021800
- Symonds, V.V., P.S. Soltis, and D.E. Soltis. 2010. Dynamics of polyploidy formation in *Tragopogon* (Asteraceae): Recurrent formation, gene flow, and population structure. *Evolution* 64(7):1984–2003. doi:10.1111/j.1558-5646.2010.00978.x
- Takano, M., N. Inagaki, X. Xie, N. Yuzurihara, F. Hihara, T. Ishizuka, et al. 2005. Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. *Plant Cell* 17:3311–3325. doi:10.1105/tpc.105.035899
- Uemura, M., R.A. Joseph, and P.L. Steponkus. 1995. Cold acclimation of *Arabidopsis thaliana* (effect on plasma membrane lipid composition and freeze-induced lesions). *Plant Physiol.* 109:15–30. doi:10.1104/pp.109.1.15
- Vogel, K.P., R.B. Mitchell, G. Sarath, H.G. Jung, B.S. Dien, and M.D. Casler. 2004. Switchgrass. In: L.E. Moser, et al., editors, Warm-season (*C<sub>4</sub>*) grasses. ASA-CSSA-SSSA, Madison, WI. p. 561–588.
- Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:E164. doi:10.1093/nar/gkq603
- Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Wickham, H. 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.
- Wilmoth, J.C., S. Wang, S.B. Tiwari, A.D. Joshi, G. Hagen, T.J. Guilfoyle, et al. 2005. NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. *Plant J.* 43:118–130. doi:10.1111/j.1365-313X.2005.02432.x
- Yu, G., D. Smith, H. Zhu, Y. Guan, and T.T. Lam. 2017. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8:28–36. doi:10.1111/2041-210X.12628
- Zalapa, J.E., D.L. Price, S.M. Kaeppler, C.M. Tobias, M. Okada, and M.D. Casler. 2011. Hierarchical classification of switchgrass genotypes using SSR and chloroplast sequences: Ecotypes, ploidies, gene pools, and cultivars. *Theor. Appl. Genet.* 122:805–817. doi:10.1007/s00122-010-1488-1
- Zhang, Y., J.E. Zalapa, A.R. Jakubowski, D.L. Price, A. Acharya, Y. Wei, et al. 2011a. Natural hybrids and gene flow between upland and lowland switchgrass. *Crop Sci.* 51:2626–2641. doi:10.2135/cropsci2011.02.0104
- Zhang, Y., J.E. Zalapa, A.R. Jakubowski, D.L. Price, A. Acharya, Y. Wei, et al. 2011b. Post-glacial evolution of *Panicum virgatum*: Centers of diversity and gene pools revealed by SSR markers and cpDNA sequences. *Genetica (The Hague)* 139:933–948.