

The Switchgrass Genome: Tools and Strategies

Michael D. Casler, Christian M. Tobias, Shawn M. Kaeppler, C. Robin Buell, Zeng-Yu Wang, Peijian Cao, Jeremy Schmutz, and Pamela Ronald*

Abstract

Switchgrass (*Panicum virgatum* L.) is a perennial grass species receiving significant focus as a potential bioenergy crop. In the last 5 yr the switchgrass research community has produced a genetic linkage map, an expressed sequence tag (EST) database, a set of single nucleotide polymorphism (SNP) markers that are distributed across the 18 linkage groups, 4x sampling of the *P. virgatum* AP13 genome in 400-bp reads, and bacterial artificial chromosome (BAC) libraries containing over 200,000 clones. These studies have revealed close collinearity of the switchgrass genome with those of sorghum [*Sorghum bicolor* (L.) Moench], rice (*Oryza sativa* L.), and *Brachypodium distachyon* (L.) P. Beauv. Switchgrass researchers have also developed several microarray technologies for gene expression studies. Switchgrass genomic resources will accelerate the ability of plant breeders to enhance productivity, pest resistance, and nutritional quality. Because switchgrass is a relative newcomer to the genomics world, many secrets of the switchgrass genome have yet to be revealed. To continue to efficiently explore basic and applied topics in switchgrass, it will be critical to capture and exploit the knowledge of plant geneticists and breeders on the next logical steps in the development and utilization of genomic resources for this species. To this end, the community has established a switchgrass genomics executive committee and work group (<http://switchgrassgenomics.org/> [verified 28 Oct. 2011]).

BACKGROUND

SWITCHGRASS (*Panicum virgatum* L.) is a perennial C4 grass native to North America that is broadly adapted for growth across much of North and Central America. It has been the focus of agronomic, breeding, and genetics research since the 1950s, largely focused on development of cultivars and management systems for livestock production in the Great Plains region of the United States (Vogel, 2004). Research on switchgrass was intensified in 1992 when the U.S. Department of Energy (USDOE) chose

M.D. Casler, USDA-ARS, U.S. Dairy Forage Research Center, 1925 Linden Dr., Madison, WI 53706-1108; C.M. Tobias, USDA-ARS, Western Regional Research Center, Albany, CA 94710; S.M. Kaeppler, Dep. of Agronomy, Univ. of Wisconsin, 1575 Linden Dr., Madison, WI 53706; C.R. Buell, Dep. of Plant Biology, Michigan State Univ., East Lansing, MI 48824-1312; Z. Wang, The Samuel Roberts Noble Foundation, 2510 Sam Noble Pkwy., Ardmore, OK 73401; P. Cao, Dep. Plant Pathology, Univ. of California, Davis, and the UC-Davis Genome Center, CA, and China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute, Zhengzhou, China 450001; J. Schmutz, HudsonAlpha Institute of Biotechnology, Huntsville, AL, and Dep. of Energy Joint Genome Institute, Walnut Creek, CA; P. Ronald, The Joint Bioenergy Institute, Emeryville, Dep. Plant Pathology, Univ. of California-Davis, and the UC-Davis Genome Center, Davis, CA 95616. Received 10 Oct. 2011.

*Corresponding author (pcronald@ucdavis.edu).

Abbreviations: ARS, Agricultural Research Service; BAC, bacterial artificial chromosome; BES, bacterial artificial chromosome end sequence; BESC, Bioenergy Sciences Center; BFDP, Biofuels Feedstock Development Program; COMT, caffeic acid O-methyltransferase; EST, expressed sequence tag; FES, fosmid end sequence; GC, guanine-cytosine; GE, genetically engineered; GLBRC, Great Lakes Bioenergy Research Center; GRIN, Germplasm Resources Information Network; JBEI, USDOE Joint Bioenergy Institute; JGI, Joint Genome Institute; NGS, next-generation genetics; NPGS, National Plant Germplasm System; NRCS, Natural Resource Conservation Service; RFLP, restriction fragment length polymorphism; SGEC, Switchgrass Genetics Executive Committee; SNP, single nucleotide polymorphism; SRNF, Samuel Roberts Noble Foundation; SSR, simple sequence repeat; STS, sequence tagged site.

Published in The Plant Genome 4:273–282. Published 7 Dec. 2011.
doi: 10.3835/plantgenome2011.10.0026

© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

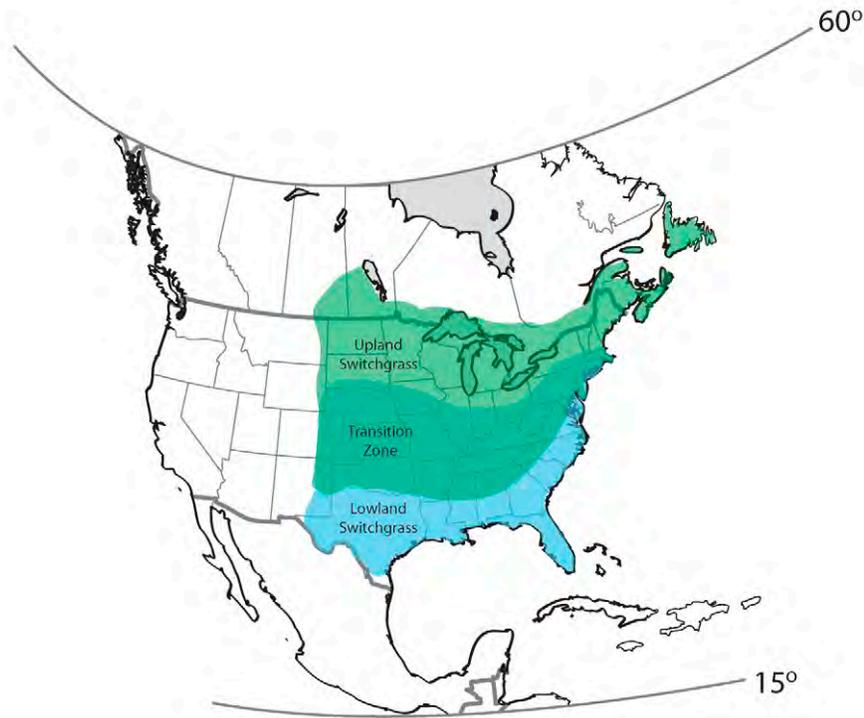


Figure 1. North American map showing the approximate adaptation ranges of upland and lowland ecotypes of switchgrass.

this species as the herbaceous model for the Biofuels Feedstock Development Program (BFDP) (Sanderson et al., 2007). That choice was based on several factors: (i) broad species adaptation from Canada to Mexico and from the Rocky Mountains to the Atlantic Ocean, (ii) relatively high biomass yields across this distributional range, (iii) high tolerance to marginal conditions, such as low fertility and drought, and (iv) the ease and simplicity of seed processing and handling. Advances in management and breeding have generated significant improvements in both biomass yield and conversion efficiency of switchgrass (Sanderson et al., 2007; Vogel, 2004).

Switchgrass Ecotypes, Ploidy, and Mating Habit

There are two principal ecotypes of switchgrass: upland and lowland. Their names reflect their origin: the upland ecotype was originally found in upland habitats, often characterized by droughty soils, while the lowland ecotype originated along riverine habitats and flood plains. Lowland switchgrass originates from the southern United States and is generally adapted to USDA hardiness zones 5 through 9 while upland switchgrass originates from the central and northern United States and is generally adapted to hardiness zones 3 through 7 (Fig. 1). Phenotypically, lowland switchgrass is taller and has longer and wider leaf blades, fewer tillers per plant, and larger stem diameter and is later in heading and flowering than upland switchgrass (Fig. 2). A bluish waxy bloom on leaf sheaths and blades is typically associated with the lowland ecotype.

Lowland cultivars of switchgrass (including Alamo, Kanlow, and Summer) are pseudotetraploid or

allotetraploid ($2n = 4x = 36$). These cultivars typically have approximately 3.1 pg per nucleus (Costich et al., 2010), implying a base haploid genome size of approximately 750 Mb. Upland cultivars (e.g., Blackwell, Cave-in-Rock, Caddo, Pathfinder, Shelter, etc.) are most frequently octoploid ($2n = 8x = 72$) and typically have 5.2 pg per nucleus, implying a base haploid genome size of approximately 650 Mb. Several tetraploid upland cultivars are available (e.g., Dacotah, Summer), but the lower ploidy is less frequent in the upland ecotype (Zhang et al., 2011a). Genetic mapping data shows that lowland switchgrass behaves more or less as a genetic diploid (disomic inheritance) with two alleles at each locus of an approximately 1.5 Gb genome rather than as true tetraploid with four alleles per locus of an approximately 750 Gb genome (tetrasomic inheritance). Therefore the (haploid) genome size of switchgrass is 1.5 Gb, an estimate that is assumed below. Tetraploid switchgrass has 18 linkage groups distributed into two highly homologous subgenomes (Okada et al., 2010).

Diploid *P. virgatum* plants have been identified through two screening approaches. Young et al. (2010) identified two individuals from a 4x-4x cross that were identified by anomalous DNA marker profiles—both individuals had fewer than expected numbers of alleles. Subsequent flow cytometry and chromosome counts confirmed that these plants were $2n = 2x = 18$ (Young et al., 2010). These authors reported that these dihaploid plants are infertile and unsuitable for genetics. Alternatively, D.L. Price and M.D. Casler (personal communication, 2011) identified numerous dihaploid individuals in several tetraploid seed lots by screening for twin seedlings. Fertile diploids would potentially be very useful in basic genome



Figure 2. Photograph of lowland switchgrass (left) and upland switchgrass (right) taken in late August near Arlington, WI.

and cytogenetic studies especially for generation of inbred lines and to assist genome-sequencing effort but all individuals have so far been completely male and female sterile (Young et al., 2010; D.L. Price and M.D. Casler, personal communication, 2011).

Aneuploidy is common in switchgrass and genome instability is particularly prevalent in octoploid accessions (Costich et al., 2010). The authors concluded that, due to the presence of mixed ploidy and aneuploidy within populations and accessions, unsubstantiated assumptions about the ploidy of individual plants is risky. This fact should not impact the genome sequencing efforts of switchgrass but will pose challenges for analysis of diversity and is a good reason why chloroplast markers (Zhang et al., 2011a, b) have been extensively relied on for studies of population structure.

Genetic diversity, rates of outcrossing, and breeding strategies are strongly influenced by anemophily (wind pollination) in switchgrass and by the presence of active prezygotic and postzygotic incompatibility systems (Martínez-Reyna and Vogel, 2002). Although well-characterized inbred lines of switchgrass do not exist for this reason, genetic analysis has indicated that the rate of self-pollination is significantly greater than zero in some genotypes. Genotypes vary in their rate of self-pollination and this rate can be as high as 50% in some individuals (E.S. Buckler and M.D. Casler, personal communication, 2011). This may be due strictly to genetics or a combination of other factors that might result in breakdown of incompatibility mechanisms. The creation

of inbred lines through repeated selfing is theoretically possible by collecting seed from bagged panicles followed by genotyping to confirm parentage and probably some selection to maintain vigor. Development of true inbred lines will require a substantial selfing or sib-mating effort accompanied by some selection pressure to ensure that inbred plants are capable of producing seed.

Switchgrass Germplasm, Diversity, and Breeding

All public and commercial germplasm of switchgrass originated from remnant prairie sites that have been preserved. There are literally thousands of these sites scattered across the species range and ranging in size from several hundred square meters to hundreds of hectares (e.g., several national grasslands, such as Cimarron, Comanche, and Rita Blanca, as well as regions such as the Flint Hills and Sand Hills). These habitats include remnants of tallgrass prairie, oak savanna, pine barrens, forest margins, and some wetlands.

Starting in the early 1990s and continuing until the present time, germplasm exploration and collection of switchgrass has been a high priority. The USDA National Plant Germplasm System (NPGS) has a small collection accessible via the Germplasm Resources Information Network (GRIN) (<http://www.ars-grin.gov/> [verified 28 Oct. 2011]). At the time of this writing, there were 174 accessions in the GRIN database but only about 60 of these represent unique accessions. Approximately 2/3 of the collection represents a narrow set of half-sib families from Union County, SD (PI numbers 642193 to 648367). The

remaining accessions represent only a fraction of the genotypic and phenotypic diversity available within the species. Numerous DNA marker studies have identified distinct patterns of regional and geographic diversity within both upland and lowland switchgrass ecotypes, leading to the a concept of regional gene pools and to the identification of recent gene flow between upland and lowland ecotypes (Zalapa et al., 2011; Zhang et al., 2011a, b).

Thousands of switchgrass accessions have been collected and currently reside in public and private germplasm collections across North America, including universities, private companies, nongovernment “seed saver” or “heirloom variety” organizations, and the USDA Natural Resource Conservation Service (NRCS) Plant Materials Centers (<http://plant-materials.nrcs.usda.gov/> [verified 28 Oct. 2011]). Some of these accessions have been released as natural-track cultivars with no breeding or selection history (e.g., Alamo, Blackwell, Cave-in-Rock, Kanlow, and many others). Others have been bulked into regional gene pools that preserve the variability but lose the connection between genotype and site of origin (e.g., Central Iowa germplasm, Southlow germplasm, MS-SG germplasm). Public availability of these accessions is limited to those that are available through NRCS or those that are shared between colleagues or collaborators. There has been no concerted effort to coordinate collection or public availability of these germplasm resources. Funding for seed preservation within the NPGS will ensure that these unique accessions are maintained and available.

Most breeding advances before the USDOE-BFDP were achieved in upland germplasm pools and focused on improving livestock production systems. Efforts to improve lowland switchgrass have intensified as a direct result of the USDOE-BFDP and its broad scope that focused on biofuel feedstock supplies throughout the entire nation. Moving southern germplasm north is a rapid and effective way to increase biomass yield in a switchgrass production system, largely due to the ability of later-flowering genotypes to continue photosynthesizing and accumulating dry matter later into the autumn. Because southern genotypes often lack the cold tolerance required at northern sites, efforts to select lowland genotypes with improved cold tolerance have intensified in recent years, moving the lowland switchgrass ecotype into the mainstream of most switchgrass breeding programs. As such, recent intensification and expansion of efforts to collect switchgrass seeds from the entire species range will form a critical foundation of future efforts to develop switchgrass into a viable bioenergy feedstock in a wide range of suitable environments (Zalapa et al., 2011; Zhang et al., 2011a, b).

Switchgrass breeding was significantly intensified in the 1990s after the USDOE decision to utilize this species as an herbaceous energy-crop model. Breeding programs multiplied from 3 to 10 during the past 15 yr and there is current breeding activity in the following states: Nebraska, South Dakota, Wisconsin, Oklahoma, Texas, New Jersey, New York, Alabama, Georgia, and

Mississippi. Most breeding objectives are focused on development of improved cultivars for biomass conversion to energy, with biomass yield as the principal trait, owing to its identification as the most important factor limiting on-farm profitability (Perrin et al., 2008). Cell-wall recalcitrance, several environmental stress tolerances, and pest resistances form the basis of additional breeding objectives in most programs. Most breeding programs focus their efforts on a regional target that spans a maximum of two or three hardiness zones within either the Great Plains region or the eastern half of the United States. While there are a small number of accessions with broad adaptation across several regions (e.g., Cave-in-Rock and Alamo), most breeding programs are working with germplasm that has the best local adaptation. As such, there is considerable demand for new genetic resources to satisfy the demand for germplasm adapted to nearly every niche or habitat east of the 100th meridian, including the genetic resources to support both traditional breeding and molecular breeding approaches.

The association of molecular markers with geographic and phenotypic differentiation combined with advances in molecular marker technologies has increased the efforts to incorporate these technologies directly into switchgrass improvement programs. For example, development and implementation of genomic selection is expected to minimally triple the rate of gain per year for traits such as biomass yield, survivorship (stress tolerance), and recalcitrance, which typically require huge field efforts for sampling and measurement combined with lengthy life cycles to collect meaningful data (Casler and Brummer, 2008). The potential benefit of genomic selection will be realized through advances in cost-effective high-throughput genotyping platforms and collection of high-quality phenotypic data across environments and in the context of relevant production systems. While phenotyping is often considered to be one of the most important factors limiting the use and effectiveness of genomic selection, several breeding programs have reached the point at which hundreds of half-sib families have been rigorously evaluated at multiple locations and across years; in these cases genomic selection could be implemented as soon as the cost, sequencing methodology, and analytical methods are available. Completion of the reference genome will allow complete interpretation of markers used in genomic selection in terms of physical and genetic proximity and by facilitating assignment of alleles to the homeologous genomes.

It is worth noting that a robust diploid congeneric of switchgrass, witchgrass (*Panicum capillare* L.), is available with a genome size of approximately 500 Mb. Due to high levels of inbreeding (T. Mitchell-Olds, personal communication, 2011), the genome of this weedy annual would provide a relatively tractable and especially valuable comparator for switchgrass. *Panicum capillare* and switchgrass diverged from their most recent common ancestor about nine million years ago (Zhang et al., 2011a).

GENETIC AND GENOMIC RESOURCES AVAILABLE FOR SWITCHGRASS

Genetic Resources

Several mapping populations and a comprehensive resource of genomic-simple sequence repeat (SSR), expressed sequence tag (EST)-SSR, sequence tagged site (STS), diverse array technology, and single nucleotide polymorphism (SNP) markers have been developed or are under development as tools for breeding switchgrass (Saha et al., 2011; Tobias et al., 2008). Missaoui et al. (2005) investigated the genomic organization and chromosomal transmission in switchgrass using restriction fragment length polymorphism (RFLP) markers and constructed the first low-density RFLP linkage map of Alamo × Summer (AP13 × VS16) switchgrass. They observed a high degree of preferential pairing between homologous chromosomes in switchgrass and, based on ratio of repulsion to coupling linkages, suggested that switchgrass is an autotetraploid with disomic inheritance. More recently, Okada et al. (2010) constructed the complete linkage maps of switchgrass arranged in nine homeologous pairs using SSR and STS markers. They observed substantial subgenome divergence through analysis of amplicons that mapped across homeologs. Analysis of repulsion to coupling phase linkages confirmed complete or near-complete disomic inheritance. These linkage maps demonstrated that the nine homeologous groups corresponded on a one-to-one basis with the nine linkage groups reported for foxtail millet [*Setaria italica* (L.) P. Beauv.].

Two association panels have been developed for marker-trait association studies. One panel is largely upland in origin, consisting of 60 northern accessions with 10 plants per accession. The upland panel was developed and evaluated as collaboration between USDA-Agricultural Research Service (ARS) (Madison, WI) and Cornell University (Ithaca, NY) (E.S. Buckler and D. Costich, personal communication, 2011). A second panel is largely lowland in origin, consisting of 48 southern accessions with 10 plants per accession. The lowland panel is currently undergoing evaluation as a partnership between the University of Georgia (Athens, GA) and The Samuel Roberts Noble Foundation (SRNF) (Ardmore, OK) (E.C. Brummer, personal communication, 2011). Subsets of both panels have been combined into DNA-marker diversity studies to identify patterns of genetic diversity across the species range and to identify plants of upland-lowland hybrid origin (Zhang et al., 2011a, b). Meta-analyses of the two association panels are currently underway as a partnership between the USDOE Bioenergy Sciences Center (BESC) (Oak Ridge, TN), Great Lakes Bioenergy Research Center (GLBRC) (Madison, WI), and SRNF using exome features mapped onto a physical linkage map of switchgrass combined with phenotypic data collected on both panels. These exome features are currently under development using EST resources of switchgrass.

A number of genetic tools have also been developed to facilitate genomic studies of switchgrass (Missaoui et al., 2005; Narasimhamoorthy et al., 2008; Okada et al., 2010; Tobias et al., 2005, 2008). Application of these tools to germplasm collections can be extremely useful in developing germplasm conservation programs and in designing efficient breeding programs, particularly for a species that has such massive phenotypic and adaptive polymorphisms across its range (Casler et al., 2007; Vogel, 2004). Initial studies using RFLP and random amplified polymorphic DNA markers revealed two prevalent ecotypes of switchgrass including tall (≥ 3 m) and thick-stemmed lowland cultivars (mostly tetraploid) with high potential productivity and shorter (1.5–2.5 m), thin-stemmed upland cultivars (mostly octaploid) with lesser biomass yield (Gunter et al., 1996; Missaoui et al., 2006).

More recent studies have demonstrated the precision with which DNA markers can be employed to identify and classify switchgrass individuals. Zalapa et al. (2011) used 55 SSR loci to classify individual plants of 19 cultivars with a success rate of 99.5% (1 plant of 192 misclassified as to its cultivar of origin). Similarly, Zhang et al. (2011a) identified two putatively local switchgrass accessions, one from New York and one from North Carolina, that were most likely inadvertently transplanted by the U.S. Army when transporting horses from northwestern Nebraska. Additionally, Zhang et al. (2011b) identified a relatively high frequency of switchgrass plants of upland × lowland hybrid origin, suggesting that glacial refugia near the Gulf Coast are diversity hot spots for switchgrass. Much of this variation has been preserved in prairie and savanna remnant habitats and is currently the focus of intensive germplasm exploration, collection, and preservation efforts all along the Gulf Coast region.

Whole Genome Sequencing

Switchgrass provides substantial challenges for whole genome assembly due to its outbred, highly heterozygous nature, the fact that it is polyploid, and the absence of robust diploid accessions. The current whole-genome sequencing effort is focused on a high-yielding, lowland, tetraploid switchgrass clone, AP13. Alamo, the cultivar from which this clone originated, is extensively distributed throughout switchgrass breeding programs in the southern United States and is a parent of several mapping populations. Because the switchgrass clone chosen as a reference is a heterozygous tetraploid with two subgenomes, several different strategies are being applied to generate the reference pseudomolecules. These include development of extensive physical clone resources [fosmids and bacterial artificial chromosome (BAC) clones, with insert sizes ranging from 40 to 200 kb] and fosmid end sequences (FES) or BAC end sequences (BESs), jumping libraries (starting at one point in the polynucleotide sequence, a distal point is located without having to first sequence the intervening bases), a dense genetic mapping effort based on next-generation skim sequencing (the partial random sequencing of a large-insert clone),

ongoing sampling with whole genome shotgun sequencing (the random generation of short DNA-sequence reads from the whole genome) utilizing both Roche 454 (Roche 454 Life Sciences, Basel, Switzerland) (achieves read lengths of 400–500 bp) and Illumina (Illumina, Inc., San Diego, CA) (previously Solexa; achieves read lengths above 100 bp) sequencing platforms, and efforts to evaluate and sequence closely related genomes that could provide an organizing principal for switchgrass.

A major challenge will be to independently assemble the two subgenomes and reach chromosome-scale contiguity for the reference. An accurate estimate of genome structure and composition (for example, guanine–cytosine [GC] content, distribution of repeat elements, percentage of coding regions, and collinearity with sequenced genomes) before full genome sequencing and assembly is needed. Generation and sequencing of BAC libraries is an efficient strategy to obtain this information and assembly of the large and complex genomes. The USDOE Joint Bioenergy Institute (JBEI) (Berkeley, CA) has led a collaborative effort to generate and characterize two additional BAC libraries representing near-complete coverage of the genome (M. Sharma, R. Sharma, and P. Ronald, personal communication, 2011). These libraries were constructed from switchgrass cultivar Alamo clone AP13. A collection of 330,297 high-quality BESs was generated from these libraries providing a basis for genome-wide survey of switchgrass genome structure and organization. Comparative mapping of full-length as well as BESs onto rice (*Oryza sativa* L.), maize (*Zea mays* L.), sorghum [*Sorghum bicolor* (L.) Moench], *Brachypodium distachyon* (L.) P. Beauv., and foxtail millet revealed extensive syntenous chromosomal regions and microcollinearity among grass genomes (M. Sharma, R. Sharma, and P. Ronald, personal communication, 2011; J. Bennetzen, personal communication, 2011). Gene annotations and analysis of BES provide an estimate of GC content, repeat elements, and SSRs in the switchgrass genome. A blast and GBrowse server was constructed to analyze sequenced full length BACs (<http://switchgrass.ucdavis.edu/> [verified 28 Oct. 2011]; M. Sharma, R. Sharma, and P. Ronald, personal communication, 2011). Recently, an *EcoRI*-generated BAC library for the SL93 2001-1 genotype of Alamo switchgrass has also been reported (Saski et al., 2011). The USDOE-Joint Genome Institute (JGI) (Walnut Creek, CA) and BESC have also developed multiple random-shear, fosmid-based libraries that will be end sequenced and employed for medium-range linking. These libraries are currently being sequenced to completion.

To address the goal of localizing genomic scaffolds within a specific subgenome, the USDOE-JGI is developing a next-generation genetics (NGS) map based on detailed resequencing of the two-clone cross AP13 × VS16, representing two random individuals from the cultivars Alamo and Summer. In addition to deeply sequencing the parents of the cross, the USDOE-JGI is resequencing 192 offspring of this cross provided by Malay Saha of SRNF. The sequence data from the

offspring will then be used to identify recombination events as compared to the genomic scaffolds and allow the scaffolds to be ordered throughout the subgenomes. This map will provide a framework to localize each version of the switchgrass genomic sequence.

The bulk of the switchgrass genomic sequence being produced by the USDOE-JGI to date has been a combination of Roche 454-based and Illumina-based sequence data. Because of relatively recent tetraploidization event that produced AP13, the subgenomes have many genomic segments that differ by fewer than 1 out of 50 base pairs. This similarity between homeologs combined with the large number of repeats in grass genomes makes complete sorting of the subgenomes with the short Illumina sequences a difficult problem. The USDOE-JGI solution is to build the accessible genomic sequence from the longer 454-based sequences and utilize 150-bp paired jumping libraries of various insert-size distances to provide the bulk of short-range linking. To construct jumping libraries, double-stranded DNA is divided into large fragments with a rare cutter and one end of each piece is ligated to a selective marker and cyclized and then cut with a frequent cutter. The marker allows isolation of the two ends of the original large piece without the intervening sequence. The combination of 454-based and jumping library-based sequence data allow scaffolds on the order of tens of kilobase pairs to be produced. These scaffolds can then be linked utilizing the extensive BES and FES resources and positioned on the NGS map to produce localized subgenome specific chromosome regions.

The goal of these sequencing efforts at the USDOE-JGI is to produce consecutive releases of the AP13 genome with progressively better localization of the subgenomes and progressively more accurately positioned genomic segments in pseudomolecules. These genome releases will immediately benefit breeding efforts and scientific analyses for ongoing switchgrass feedstock improvement. They will also provide a scaffold from which regions of importance of the switchgrass genome can be identified and selected by the community. The USDOE-JGI is committed to producing complete and accurate reference sequences for these regions of functional importance by combining the whole genome sequencing data with next-generation short-read clone-based sequences and finishing these regions to reference quality.

Expression and Transcriptomics

Significant EST resources have been generated for upland and lowland ecotypes of switchgrass (over 500K ESTs at JGI, in collaboration with C. Tobias, USDA-ARS) and large-scale pyrosequencing of millions of ESTs is in progress (JGI collaboration with SRNF and BESC). These provide the first genome-scale dataset of the switchgrass transcriptome. Additional EST sequencing is ongoing for the purposes of SNP discovery in upland populations as well as targeting of tissues engaged in mycorrhizal associations and associated with dormancy, winter-hardiness, and cold acclimation (G. Sarath, personal communication,

2011). Ongoing expression studies using Affymetrix arrays (Affymetrix, Santa Clara, CA) and RNA sequencing will considerably expedite these genomic studies. These sequence data sets have been acquired from diverse sets of germplasm. Subsequent clustering has produced results that likely overestimate the number of unique genes in switchgrass. Allelic variation, splice-site variation, and variation between homeologous sequences all contribute to this effect. For the purposes of genome annotation and gene prediction a portion of the EST sequencing efforts have been restricted to the AP13 genotype.

A liquid-phase exome-capture product (Cosart et al., 2011) is underway as a collaboration among GLBRC, BESC, SRNF, and NimbleGen (Madison, WI). This product will be used to provide genotype-by-sequence information across a wide range of individuals representing ecotypes, ploidy levels, and geographical distribution. The exome-capture array will be used to produce genotypic scores that will allow assignment of many gene models to homoologous genomes by linkage mapping in the AP13 × VS16 cross population and will be used to genotype both the northern and southern association panels before their implementation in genomic selection model development for switchgrass breeding programs.

Tissue Culture and Transformation

As one of the major experimental tools in functional genomics, genetic engineering of targeted genes is useful for revealing direct links between gene sequence and gene function (Dixon et al., 2007). Genetic engineering research offers an opportunity to generate unique genetic variation that is either absent or has very low heritability (Wang and Ge, 2006). Like many other monocot species, switchgrass is considered recalcitrant for genetic transformation. Substantial progress has been made in optimizing transformation conditions and in generating transgenic switchgrass plants in the last decade. Even though transformation efficiency is often low, transgenic switchgrass plants have been obtained by *Agrobacterium*-mediated transformation (Fu et al., 2011a, b; Li and Qu, 2011; Saathoff et al., 2011; Somleva et al., 2002, 2008; Xi et al., 2009; Xu et al., 2011) and particle bombardment (Richards et al., 2001). Most of the work focused on the use of *Agrobacterium* infection for transformation, because this method tends to result in lower copy number and fewer rearrangements than the biolistic procedure (Dai et al., 2001; Hu et al., 2003; Somleva et al., 2002). Both single copy and multicopy transgene insertion have been detected in transgenic switchgrass. Molecular characterization and segregation analysis revealed that the multicopy transgene usually resides at different loci and the segregates had various copy numbers, including single copy insertions (Xi et al., 2009). Interestingly, reversal of the expression of a silenced transgene in the T0 generation parental plant was found in segregating T1 plants with a single insert (Xi et al., 2009). Because of the rapid development of gene sequencing and cloning techniques, it has become relatively

easy to isolate and clone large numbers of genes. However, functional analysis of these genes has become a bottleneck. By identification of highly tissue-culture-responsive genotypes and by optimization of transformation parameters, a high-throughput and reproducible transformation system has been developed for the widely used switchgrass cultivar Alamo (C. Fu and Z.-Y. Wang, personal communication, 2011). However, there is still a need for the development of a highly efficient and highly reproducible protocol for a genotype-independent switchgrass transformation system that can be widely applied across the species.

Genetic transformation has been effectively utilized for the improvement of biofuel crops. In switchgrass, it has been documented that genetic manipulation of a single gene could lead to large improvement in sugar release and processing properties (Fu et al., 2011a, b; Saathoff et al., 2011; Xu et al., 2011). For example, downregulation of the caffeic acid O-methyltransferase (COMT) switchgrass gene modestly decreased lignin content, improved sugar release, and increased ethanol yield by up to 38% using conventional biomass fermentation processes (Fu et al., 2011a). In addition to increased ethanol production, the genetically engineered switchgrass also showed increased forage quality, which is beneficial for farmers since switchgrass can serve as a dual-purpose (bioenergy or forage) crop. The genetically engineered plants showed normal growth and development in the greenhouse. The only phenotypic change observed between the control and COMT-downregulated lines was the brownish to reddish color in the basal internode and its cross sections. This trait can be used as a phenotypic maker during the breeding and selection process. Furthermore, the COMT-downregulated lines require reduced pretreatment severity and 300 to 400% lower cellulase dosages for equivalent product yields using simultaneous saccharification and fermentation with yeast (Fu et al., 2011a). The COMT plants are now being tested under field conditions.

The generation of genetically engineered (GE) switchgrass with superior processing properties illustrates the feasibility and potential of developing energy crops specifically designed for industrial processing to liquid fuels. Although the GE approaches are effective and straightforward, commercialization of GE perennial outcrossing biofuel crops is complicated because of strict regulatory restrictions (Ge et al., 2011; Strauss et al., 2010). Risk assessment and the development of GE containment systems are needed for switchgrass. On the other hand, modification of certain traits, such as lignin downregulation, is unlikely to increase plant fitness. Such genetic modifications can therefore be considered low risk.

The switchgrass community is now realizing that long-term maintenance of important genetic resources is complicated by the need for clonal propagation. Substantial investment is put into sequencing and genotyping individual plants that cannot be reproduced by seed. Efficient clonal propagation and long-term preservation techniques are necessary to preserve the genetic identity and availability of

important resources. Maintaining populations as seed and individuals in situ has been adequate to this point but is inadequate to allow long-term preservation of individual genotypes. In vitro methods of clonally propagating and archiving switchgrass genotypes are available and can be scaled up to high-throughput systems (Gupta and Conger, 1999) but have not been evaluated on broad germplasm collections, as would be necessary to widely employ these methods.

Databases

With the large amount of data being accumulated, databases that integrate grass gene sequence information are needed to provide a platform for comparative genomic studies. Toward this end, researchers at the JBEI have established phylogenomics databases.

Phylogenomics is a phylogenetic approach used in comparative genomics to predict the biological functions of members of large gene families by assessing the similarity among gene products. Two phylogenomics databases for kinases (RiceKinase Database, available at <http://phylomics.ucdavis.edu/kinase/> [verified 28 Oct. 2011]) and rice glycosyltransferases (Rice Glycosyltransferase Database; <http://phylomics.ucdavis.edu/cellwalls/gt/> [verified 28 Oct. 2011]) are now available that include both rice and switchgrass sequences (Jung et al., 2010). These databases present diverse data in a phylogenetic context, including gene annotations, orthologous gene predictions, information about gene indexed mutants, as well as transcriptome data from ESTs, massively parallel signature sequencing, and microarray analyses.

The JBEI has also set up a directory of databases for plant cell wall-related enzymes (plantcellwalls.ucdavis.edu [verified 28 Oct. 2011]) so that grass researchers can more efficiently mine the wide spectrum of online databases of plant cell wall-related enzymes (Cao and Ronald, 2010). The USDOE-funded knowledgebase has recently launched a multi-institutional effort to consolidate the numerous different sources of scientific information on plants, including switchgrass, into a single integrated cyber-database (<http://science.energy.gov/news/in-the-news/2011/07-07-11/> [verified 28 Oct. 2011]).

SWITCHGRASS COMMUNITY ORGANIZATION

Increasing interest and emphasis on improving switchgrass germplasm and our knowledge of switchgrass biology and genetics has resulted in the evolution of a fairly large switchgrass community. The number of switchgrass breeding programs has increased from three in 1995 to 10 at the time of this writing, including both the public and private sectors, and the genomics and genetics group has expanded significantly as well. Efforts to improve the knowledge base regarding basic genetics and genomics have increased concomitantly.

Information on switchgrass genetics and genomics can be found through a web portal (<http://switchgrassgenomics.org/> [verified 28 Oct. 2011]) that posts a

wide range of information on switchgrass genetics and genomics research and community activities. Research projects, programs, and grant awards are summarized on the research efforts page (<http://www.switchgrassgenomics.org/research.shtml> [verified 28 Oct. 2011]) along with presentations and meeting summary reports. Summary information on public sequence resources for switchgrass is available through the Switchgrass Genome Data page (http://www.switchgrassgenomics.org/pub_sequence.shtml [verified 28 Oct. 2011]). Sources of germplasm for *Panicum virgatum* can be found on the Publicly Available Switchgrass Germplasm page (<http://www.switchgrassgenomics.org/germplasm.shtml> [verified 28 Oct. 2011]). The Switchgrass Genetics and Genomics Google Group has been set up for information distribution and individuals can self-subscribe or use the Switchgrass Genomics Email List page (<http://www.switchgrassgenomics.org/contact.shtml> [verified 28 Oct. 2011]) to become a member. Curators of the website can be reached via switchgrassgenomics@plantbiology.msu.edu to add or modify content on the website.

One key component of a research community is leadership and, analogous to the Maize Genetics Executive Committee (<http://www.maizegdb.org/mgec.php> [verified 21 Nov. 2011]), the Switchgrass Genetics Executive Committee (SGEC) (http://www.switchgrassgenomics.org/exec_committee.shtml [verified 21 Nov. 2011]) is tasked with providing information, coordinating meetings, facilitating research efforts among the community, and highlighting research needs to funding agencies. In 2011, the inaugural SGEC was elected by the community (http://www.switchgrassgenomics.org/exec_committee.shtml [verified 28 Oct. 2011]).

As the switchgrass genetics and genomics community has grown so have germplasm exploration and collecting efforts. Most of these collections have yet to be deposited in the GRIN system, the USDA-ARS system that is the only nationwide repository of publicly available germplasm. The USDA-NRCS Plant Materials Centers have been responsible for collecting switchgrass from much of its range and many of these collections have been deposited into GRIN for public distribution. Many hundreds of additional collections have been made, in the form of either live tillers or seeds, which could be deposited into GRIN, making them accessible to the entire community. The SGEC will take up the challenge of coordinating accessibility of these accessions, balancing the need for wide community access with the need to avoid overloading the workload of the GRIN switchgrass curator.

CONCLUSIONS AND FUTURE OUTLOOK

Current efforts to develop switchgrass into a dedicated energy crop and to unlock its physiological and genomic secrets are hampered by several research needs that have yet to be fulfilled. Community priorities will be identified and articulated by the Switchgrass Genetics Executive Committee. Included in the needs likely to be identified as high priority by the community are

- A completed reference genome including separation of homoeologous genomes and advanced annotation. The information needs to be accessible to all community members.
- A central database that serves as an entry to the genome sequence and genome resources, archives genetic information such as mutant stocks, and quantitative trait loci.
- Cost-effective, high-density genotyping platforms.
- Mapping populations and association panels with high-density genotype information.
- Transcriptomics, proteomics, and metabolomics tools and information for key genotypes.
- Phylogenomic databases to predict switchgrass gene function through comparative genomic analyses of grasses and other species.
- Genotype-independent methodologies to clonally archive and propagate heterozygous individuals.
- Facile transformation across genotypes.
- Deletion, silencing, and insertion mutant collections to facilitate validation of gene function.
- Germplasm collection, characterization, and preservation.
- Systems biology computational tools to facilitate prediction of gene and pathway function from diverse datasets (Lee et al., 2011).

With the anticipated advances in genome sequencing, tool development, genetics, and breeding, a solid foundation for advanced breeding, genomics, genetics, ecology, and physiology will soon be available. These tools will facilitate a new era of bioenergy crop development and production.

Acknowledgments

The studies conducted by the U.S. Department of Energy Joint Genome Institute, the BioEnergy Science Center, and the Joint Bioenergy Institute are supported by the Office of Science of the U.S. Department of Energy under Contract Numbers DE-AC02-05CH11231, DE-AC05-00OR22725, and DE-AC02-05CH11231, respectively. This research was also supported by a grant from the NIFA Plant Feedstock Genomics for Bioenergy Program (#2010-04195) to P.C.R. We thank D. Rokhsar, L. Bartley, R. Sharma, and M. Sharma for helpful discussions.

References

- Cao, P., K.-H. Jung, and P.C. Ronald. 2010. A survey of databases for analysis of plant cell wall-related enzymes. *Bioenergy Res.* 3:108–114. doi:10.1007/s12155-010-9082-6
- Casler, M.D., and E.C. Brummer. 2008. Theoretical expected genetic gains for among-and-within-family selection methods in perennial forage crops. *Crop Sci.* 48:890–902. doi:10.2135/cropsci2007.09.0499
- Casler, M.D., K.P. Vogel, C.M. Taliaferro, N.J. Ehlke, J.D. Berdahl, E.C. Brummer, R.L. Kallenbach, C.P. West, and R.B. Mitchell. 2007. Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Sci.* 47:2249–2260. doi:10.2135/cropsci2006.12.0780
- Cosart, T., A. Beja-Pereira, S. Chen, S.B. Ng, J. Shendure, and G. Luikart. 2011. Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics* 12:347–354. doi:10.1186/1471-2164-12-347

- 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 Gen.* 3:130–141. doi:10.3835/plantgenome2010.04.0010
- Dai, S., P. Zheng, P. Marmey, S. Zhang, W. Tian, S. Chen, R.N. Beachy, and C. Fauquet. 2001. Comparative analysis of transgenic rice plants obtained by *Agrobacterium*-mediated transformation and particle bombardment. *Mol. Breed.* 7:25–33. doi:10.1023/A:1009687511633
- Dixon, R.A., J.H. Bouton, B. Narasimhamoorthy, M. Saha, Z.-Y. Wang, and G.D. May. 2007. Beyond structural genomics for plant science. *Adv. Agron.* 95:77–161. doi:10.1016/S0065-2113(07)95002-6
- Fu, C., J.R. Mielenz, X. Xiao, Y. Ge, C. Hamilton, M. Rodriguez, F. Chen, M. Foston, A. Ragauskas, J. Bouton, R.A. Dixon, and Z.-Y. Wang. 2011a. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA* 108:3803–3808. doi:10.1073/pnas.1100310108
- Fu, C., X. Xiao, Y. Xi, Y. Ge, F. Chen, J. Bouton, R.A. Dixon, and Z.-Y. Wang. 2011b. Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. *Bioenergy Res.* 4:153–164. doi:10.1007/s12155-010-9109-z
- Ge, Y., C. Fu, H. Bhandari, J. Bouton, E.C. Brummer, and Z.-Y. Wang. 2011. Pollen viability and longevity of switchgrass (*Panicum virgatum* L.). *Crop Sci.* 51:2698–2705. doi:10.2135/cropsci2011.01.0057
- Gunter, L.E., G.A. Tuscan, and S.D. Wullschlegler. 1996. Diversity of switchgrass based on RAPD markers. *Crop Sci.* 36:1017–1022. doi:10.2135/cropsci1996.0011183X003600040034x
- Gupta, S.D., and B.V. Conger. 1999. Somatic embryogenesis and plant regeneration from suspension cultures of switchgrass. *Crop Sci.* 39:243–247. doi:10.2135/cropsci1999.0011183X003900010037x
- Hu, T., S. Metz, C. Chay, H.P. Zhou, N. Biest, G. Chen, M. Cheng, X. Feng, M. Radionenko, F. Lu, and J. Fry. 2003. *Agrobacterium*-mediated large-scale transformation of wheat (*Triticum aestivum* L.) using glyphosate selection. *Plant Cell Rep.* 21:1010–1019. doi:10.1007/s00299-003-0617-6
- Jung, K.-H., P. Cao, Y.-S. Seo, C. Dardick, and P.C. Ronald. 2010. The rice kinase phylogenomics database: A guide for systematic analysis of the rice kinase super-family. *Trends Plant Sci.* 15:595–599. doi:10.1016/j.tplants.2010.08.004
- Lee, I., Y.S. Seo, D. Boutz, S. Craig, K. Nozue, C.J. Park, D. Cotrane, T. Oh, E. Marcotte, and P.C. Ronald. 2011. Genetic dissection of the biotic stress response using a genome scale gene network for rice. *Proc. Natl. Acad. Sci. USA* doi:10.1073/pnas.1110384108.
- Li, R., and R. Qu. 2011. High throughput *Agrobacterium*-mediated switchgrass transformation. *Biomass Bioenergy* 35:1046–1054. doi:10.1016/j.biombioe.2010.11.025
- Martínez-Reyna, J., and K.P. Vogel. 2002. Incompatibility systems in switchgrass. *Crop Sci.* 42:1800–1805. doi:10.2135/cropsci2002.1800
- Missaoui, A.M., A.H. Paterson, and J.H. Bouton. 2005. Investigation of the genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. *Theor. Appl. Genet.* 110:1372–1383. doi:10.1007/s00122-005-1935-6
- Missaoui, A.M., A.H. Paterson, and J.H. Bouton. 2006. Molecular markers for the classification of switchgrass (*Panicum virgatum* L.) germplasm and to assess genetic diversity in three synthetic switchgrass populations. *Genet. Resour. Crop Evol.* 53:1291–1302. doi:10.1007/s10722-005-3878-9
- Narasimhamoorthy, B., M.C. Saha, T. Swaller, and J.H. Bouton. 2008. Genetic diversity in switchgrass collections assessed by EST-SSR markers. *BioEnergy Res.* 1:136–146. doi:10.1007/s12155-008-9011-0
- Okada, M., C. Lanzatella, M.C. Saha, J. Bouton, R. Wu, and C.M. Tobias. 2010. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics* 185:745–760. doi:10.1534/genetics.110.113910
- Perrin, R., K. Vogel, M. Schmer, and R. Mitchell. 2008. Farm-scale production cost of switchgrass for biomass. *BioEnergy Res.* 1:91–97. doi:10.1007/s12155-008-9005-y
- Richards, H.A., V.A. Rudas, H. Sun, J.K. McDaniel, Z. Tomaszewski, and B.V. Conger. 2001. Construction of a GFP-BAR plasmid and its use for switchgrass transformation. *Plant Cell Rep.* 20:48–54. doi:10.1007/s002990000274

- Saathoff, A.J., G. Sarath, E.K. Chow, B.S. Dien, and C.M. Tobias. 2011. Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. *PLoS ONE* 6:e16416. doi:10.1371/journal.pone.0016416
- Saha, M.C., D. Serba, H.S. Bhandari, L. Escamilla-Treviño, J.H. Bouton, E.C. Brummer, and K.M. Devos. 2011. Developing resources for molecular breeding of switchgrass. Abstract W084. *Plant Animal Gen.* XIX:15–19.
- Sanderson, M.A., P.R. Adler, A.A. Boateng, M.D. Casler, and G. Sarath. 2007. Switchgrass as a biofuels feedstock in the USA. *Can. J. Plant Sci.* 86:1315–1325. doi:10.4141/P06-136
- Saski, C.A., Z. Li, F.A. Feltus, and H. Luo. 2011. New genomic resources for switchgrass: A BAC library and comparative analysis of homoeologous genomic regions harboring bioenergy traits. *BMC Genomics* 12:369. doi:10.1186/1471-2164-12-369
- Somleva, M., K. Snell, J. Beaulieu, O. Peoples, B. Garrison, and N. Patterson. 2008. Production of polyhydroxybutyrate in switchgrass, a value-added co-product in an important lignocellulosic biomass crop. *Plant Biotechnol. J.* 6:663–678. doi:10.1111/j.1467-7652.2008.00350.x
- Somleva, M.N., Z. Tomaszewski, and B.V. Conger. 2002. *Agrobacterium*-mediated genetic transformation of switchgrass. *Crop Sci.* 42:2080–2087. doi:10.2135/cropsci2002.2080
- Strauss, S.H., D.L. Kershen, J.H. Bouton, T.P. Redick, H. Tan, and R.A. Sedjo. 2010. Far-reaching deleterious impacts of regulations on research and environmental studies of recombinant DNA-modified perennial biofuel crops in the United States. *Bioscience* 60:729–741. doi:10.1525/bio.2010.60.9.10
- Tobias, C.M., P. Twigg, D.M. Hayden, K.P. Vogel, R.M. Mitchell, G.R. Lazo, E.K. Chow, and G. Sarath. 2005. Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass. *Theor. Appl. Genet.* 111:956–964. doi:10.1007/s00122-005-0030-3
- Tobias, C.M., P. Twigg, D.M. Hayden, K.P. Vogel, R.M. Mitchell, G.R. Lazo, E.K. Chow, and G. Sarath. 2008. Comparative genomics in switchgrass using 61585 high-quality expressed sequence tags. *Plant Gen.* 1:111–124. doi:10.3835/plantgenome2008.08.0003
- Vogel, K.P. 2004. Switchgrass. p. 561–588. *In* L.E. Moser et al. (ed.) *Warm-season (C4) grasses*. ASA, CSSA, SSSA, Madison, WI.
- Wang, Z.-Y., and Y. Ge. 2006. Recent advances in genetic transformation of forage and turf grasses. *In Vitro Cell. Dev. Biol. Plant* 42:1–18.
- Xi, Y., C. Fu, Y. Ge, R. Nandakumar, H. Hisano, J. Bouton, and Z.-Y. Wang. 2009. *Agrobacterium*-mediated transformation of switchgrass and inheritance of the transgenes. *Bioenerg. Res.* 2:275–283. doi:10.1007/s12155-009-9049-7
- Xu, B., L.L. Escamilla-Treviño, N. Sathitsuksanoh, Z. Shen, H. Shen, Y.H.P. Zhang, R.A. Dixon, and B. Zhao. 2011. Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for bio-fuel production. *New Phytol.* 192:611–625. doi:10.1111/j.1469-8137.2011.03830.x
- Young, H.A., B.J. Hernlem, A.L. Anderton, C.L. Lanzatella, and C.M. Tobias. 2010. Dhaploid stocks of switchgrass isolated by a screening approach. *BioEnergy Res.* 3:305–313. doi:10.1007/s12155-010-9081-7
- 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. Zalapa, A.R. Jakubowski, D.L. Price, A. Acharya, Y. Wei, E.C. Brummer, S.M. Kaeppler, and M.D. Casler. 2011a. 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.
- Zhang, Y., J. Zalapa, A.R. Jakubowski, D.L. Price, A. Acharya, Y. Wei, E.C. Brummer, S.M. Kaeppler, and M.D. Casler. 2011b. Natural hybrids and gene flow between upland and lowland switchgrass. *Crop Sci.* 51:2626–2641.