

Switchgrass Research Group: Progress Report

January 2012

Introduction

Switchgrass shows tremendous promise as a candidate feedstock for the U.S. biofuel industry. Its capability for broad geographical adaptation and high yields on marginal lands with minimal inputs bolster the argument that switchgrass will make a significant contribution toward current and future demands for an efficient and sustainable biomass feedstock source. However, only a handful of laboratories are studying this organism, and switchgrass has little history of investigation for any use other than as a forage crop. As a possible biofuel crop, switchgrass can be considered primarily undomesticated with only limited selective breeding. A number of new research efforts now underway could speed up agronomic switchgrass improvement.

Over the last few years, opportunities provided by the Department of Energy (DOE), especially through the DOE Bioenergy Research Centers (BRCs), and the U.S. Department of Agriculture (USDA) have attracted more than a dozen new switchgrass research groups, more than tripling the number of basic researchers interested in this challenging organism. Some of these groups are performing their studies under an organized framework (e.g., the 13 switchgrass labs within the BioEnergy Science Center (BESC) at Oak Ridge National Laboratory), but there is little overall coordination. This lack of coordination has led to some duplication of efforts and, more importantly, to some potentially serious research gaps. Because no single project encompasses all that must be accomplished to provide optimal research and development for this organism, there is a real possibility that a bottleneck will be discovered that severely deters or even blocks the eventual genetic improvement and commercial development of this crop. Consequently, it was proposed that the switchgrass basic research community come together to discuss the coordination of genomic research on this crop. The first switchgrass research community meeting was held on January 18, 2011, at the Plant and Animal Genome (PAG) XIX meeting in San Diego, California. A follow-on workshop was held April 10, 2011, in Crystal City, Virginia.

Rationale

The purpose of these meetings was to assess the current state of switchgrass research, determine research gaps, and take steps to organize research and increase communication within the community to maximize efficiency, thereby leveraging the current and future resources being generated. In particular, this group will strive to ensure that all critical issues are considered and then formulate a strategy for continuity in communication, planning, and execution. The PAG meeting was co-chaired by *Jeff Bennetzen* (University of Georgia and DOE BESC) and *Tom Juenger* (University of Texas, Austin) and organized by *Cathy Ronning* of DOE's Office of Biological and Environmental Research (BER). This meeting was intended to provide a forum for initiating communication and collaboration among the switchgrass research community. A second meeting was held three months later—organized by *Cathy Ronning*, *Robin Buell*

(Michigan State University/DOE Great Lakes Bioenergy Research Center (GLBRC)), *Pam Ronald* (DOE Joint BioEnergy Institute (JBEI)), and *Michael Udvardi* (BESC)—to continue the dialogue with a specific emphasis on the roles of the three DOE BRCs, together with the DOE Joint Genome Institute (JGI). The anticipated outcome of these meetings is the integration of all research activities pertinent to the switchgrass genome sequencing effort into an efficient and effective plan that will enable the development of switchgrass as a biofuel crop.

1. PAG 2011 Meeting

Jeff Bennetzen convened the meeting by providing a perspective on how other plant research communities have coordinated research. He cited an example from the *Miscanthus* community, which is highly organized and has comprehensively defined needed research areas. Working together, the switchgrass community can also identify and close research gaps and improve research efficiencies, which is this research group's main goal.

Funding agencies require a sound and scientific rationale for supporting research, and this rationale must be provided by community consensus rather than by a single or select few investigators. *Arabidopsis* was the first community to organize around research needs in this manner, followed by the maize, sorghum, soybean, and other communities. These groups have all benefited from this approach, so there is precedence for forming a working group to coordinate research and facilitate communication among members, with the incentive of leveraging funding from federal agencies.

To focus discussion in the relatively short time period allotted, participants were asked to respond to several charge questions prior to the meeting. These questions were intended to provide the organizers with a sense of the current science, along with the community's future research plans and needs, particularly in anticipation of an annotated switchgrass genome sequence. Comments varied widely among the respondents, but there was general agreement that high-throughput screening methods, high-efficiency genetic transformation, and genome-based markers are needed to most effectively use the genetic resources becoming available for switchgrass improvements ([see Appendix on p. 11](#)).

Switchgrass Sequencing Status and Community Needs Discussion

The technical discussion on switchgrass genome sequencing was framed around the question of how researchers will use the sequence. JGI's *Dan Rokhsar* led a discussion on the current status of switchgrass genome sequencing and reasonable expectations for the future. As of January 2011, JGI had 8X coverage in 454 sequence and much higher coverage in Illumina sequence of the cultivar Alamo 13. This sequence coverage is considerable, but as an outbred tetraploid, it is not sufficient for a finished, fully assembled genome. All switchgrass data (which has been publicly deposited) is being mapped onto the foxtail millet genome. This genome is being used as an organizing substrate to define localized collections of reads for a first-pass assembly.

In parallel, JGI is collaborating with several researchers to investigate diploid switchgrass genotypes as a genomic model. Because it is less complex, diploid switchgrass (e.g., *Panicum hallii*) should enable assembly of the gene space plus the surrounding genomic loci and provide useful sequence much sooner than that of the tetraploid *P. virgatum*. *Tom Juenger* is planning very dense marker sequence-based mapping with this diploid genotype (anticipated to take place in early 2012).

Another approach is the construction of a dense genetic linkage map, which was viewed as a top priority by some group members. The AP13 x VS16 population, with more than 300 plants and 1,200 markers, is available now, but it was conservatively estimated that 50,000 markers are needed to guarantee markers every several kilobases. This approach may allow accurate parsing of genes into the separate A and B subgenomes, but would not likely provide highly accurate gene orders within chromosomes.

Given the interest and resources, it seems feasible to proceed with this approach. Additionally, separation of the two subgenomes should be possible by alignment with the foxtail millet genome, assuming local colinearity. Current switchgrass data could be aligned on top of the foxtail millet orthologs, regardless of the occurrence of rearrangements. A comparison between the diploid switchgrass *P. hallii* and foxtail millet would be informative and could be linked to genes and quantitative trait loci (QTLs).

Given that the genome sequence will be completed at least at some level, the larger question is how this sequence will be used. There are many possible approaches for assembly and analysis of the outbred, tetraploid switchgrass genome, but the optimal method should be defined by the end goal. For breeding purposes, broad genome coverage (rather than full coverage) would suffice; it is not necessary to know gene order within chromosomes, but a good subassembly that separates the two genomes would be essential.

Several genetic resources currently exist, including a highly homozygous lowland tetraploid population with simple sequence repeat (SSR) markers (*Yanqui Wu*, Oklahoma State University) and a dihaploid line (*Christian Tobias*, USDA-Agricultural Research Service (ARS)). Nearly 500,000 Sanger sequence reads exist and bacterial artificial chromosome (BAC) end sequencing is currently being provided by *Jeremy Schmutz* and *Jerry Jenkins* at HudsonAlpha. Other resources are becoming available as well. The ability to turn shotgun reads into gene models is good, but the challenge is in arranging them into larger haplotypes. Transcriptome data can be mapped onto genome sequence data as long as there is a good initial assembly. Again, the challenge will be tracking single-nucleotide polymorphisms (SNPs) to the right allele. The amount of sequencing required to resolve the A and B subgenomes, including the use of BAC end and transcript sequences and barcoding, was discussed at length. At least one researcher is developing markers, which will be mapped and then made available for marker-assisted breeding.

Issues to Consider/Resolve

Research Needs: Meeting participants were asked to consider their future priorities and needs, particularly regarding the switchgrass genome sequence, that will enhance their research productivity. Questions included:

- How will you use the genome sequence?
- What more do you need from the genome sequence?
- How can JGI best provide this?
- How often should the switchgrass research community meet?

Technical Issues: As switchgrass cultivar improvement proceeds, technical issues should be anticipated and considered before problems arise (e.g., potential invasiveness caused by the release of transgenic switchgrass that is not male sterile). The resource allocation balance will

change as needs change, and it was suggested that an allocation analysis be conducted on occasion to reassess these needs.

Commercial Sector: What role might the commercial sector play in the switchgrass research community? To determine the switchgrass community's needs, open meetings will be held and stakeholder input welcomed.

Summary of Action Items and Next Steps

Throughout the discussion, the group identified several action items, which are summarized below. These include (1) forming a switchgrass executive committee, (2) defining the group's scope and expanding its mailing list, (3) evaluating the idea of an annual switchgrass meeting, and (4) creating a wiki for information sharing.

1. Forming a Switchgrass Executive Committee

The creation of a switchgrass executive committee to represent the community and organize and communicate information within it was broadly accepted as a critical step. Members will be elected from within the switchgrass research community and will receive no monetary compensation for serving. The committee's size will be determined by the research priorities identified by the community, with representation from each identified research area.

The committee's job will be to identify the needs of the research community and communicate these needs to those in a position to take action. The committee's function is to organize and communicate, but not govern (e.g., the maize executive committee conducts an annual survey to ascertain community issues and communicates this information more broadly outside the community). The executive committee should organize information and communicate in a deliberate way, presenting a well-informed position that represents the research community's needs and interests to the funding agencies.

An election was held during the summer of 2011, and seven representatives are now serving on the executive committee (www.switchgrassgenomics.org/exec_committee.shtml). The committee is currently drafting a charter to clearly articulate the committee's goals and role in fostering switchgrass research.

2. Defining the Group's Scope and Expanding Its Mailing List

The researchers present at this meeting primarily represented the switchgrass breeding, genetics, and genomics areas. There was a sense that broader collaborations would be beneficial to switchgrass research. Because of the diverse research interests among the participants, the group discussed creating national collaborative subgroups focused on specific research problems.

It was recognized that scientists from many areas—including agronomy, evolution, ecology, physiology, cell biology, and restoration and conservation—are also contributing to switchgrass research and may be interested in participation. The group's scope should encompass the entire spectrum, from research to release; however, it was recognized that this inclusion may come at the expense of becoming too big and unwieldy. Potentially interested groups include primary and secondary users of genomic sequence and genetic data, as well as oversight.

Opinions on the ultimate size of the switchgrass working group varied among the meeting participants. Some felt that a smaller, more focused group would be the most effective, while others believed the group should be all inclusive. In general, group membership is expected to fluctuate as research priorities evolve.

A switchgrass mailing list was established during the summer of 2011 (<http://groups.google.com/group/switchgrass-genetics-and-genomics?hl=en>). Interested parties can sign up through a Google account or submit a request to be added at the switchgrass genomics website (www.switchgrassgenomics.org/contact.shtml).

3. Evaluating the Idea of an Annual Switchgrass Meeting

It was clear from the discussion that at a minimum annual meetings will be needed to continue the communication and coordination efforts begun here, but opinions varied on the need for more frequent gatherings. Meeting as part of a larger workshop or conference, such as this initial workshop at PAG, would be ideal for some but not all members of this group. The diversity of research interests represented indicates a similar diversity in conference attendance. An option is to have a stand-alone meeting at an outside location, such as the centrally located Noble Foundation, but this possibility needs to be weighed against the higher costs generally incurred by such an approach. Presentations from the various researchers should definitely be on the agenda. Such presentations would provide more background information for discussion as well as opportunities for collaboration and coordination. Options will need to be reevaluated as research and resource needs evolve.

4. Creating a Wiki for Information Sharing

A community resource such as a dynamic website or a wiki that fosters information exchange and helps elucidate key community goals would be extremely valuable. A real-time information-sharing site where meeting notes can be displayed and revised (e.g., a wiki), to be set up and administered by the community, was proposed. During the summer of 2011, a switchgrass genomics website was developed and hosted at www.switchgrassgenomics.org/index.shtml. This website is currently being maintained by *Robin Buell's* group at Michigan State University.

2. April 2011 Meeting

A second switchgrass community workshop was held on April 10, 2011, in Crystal City, Virginia, in conjunction with the joint meeting of the DOE Genomic Sciences Awardees/USDA-DOE Plant Feedstock Genomics for Bioenergy Awardees. With representatives from each of the three DOE BRCs and JGI present, as well as switchgrass researchers funded by the joint USDA-DOE program, this workshop focused on the roles of the BRCs in the switchgrass genome sequencing effort and their collaborations with JGI. BER's *Cathy Ronning* began the meeting with an overview of the switchgrass genome community organizing efforts, followed by USDA-ARS's *Christian Tobias* with a recap of the initial PAG meeting, and then presentations by representatives from each of the BRCs and JGI/HudsonAlpha. The meeting concluded with a discussion of needs and next steps to move the community forward.

Great Lakes Bioenergy Research Center

Mike Casler reported on GLBRC research, which focuses on the genetic improvement of switchgrass, translational genomics, and the utilization of genomic information from maize to design more efficient genetic improvement systems. This work includes the development of DNA marker selection systems for increasing biomass yield, decreasing lignin and etherified ferulates, and selection of late flowering survivors.

Near-term GLBRC goals are focused on developing new switchgrass varieties as efficiently as possible. Identifying the genes responsible for endogenous genetic variation in lignin and

understanding the regulatory factors involved is facilitating the effort to decrease lignin concentration in the plant. GLBRC is also developing late flowering switchgrass that does not produce seed in Wisconsin; selection is taking place in 2011 and 2012 for additional switchgrass populations.

Gains in biomass yield were made over the last 10 years, but this gain has leveled out. GLBRC is now utilizing genome sequences from both maize and switchgrass to design and test new and more efficient DNA marker selection systems, including SNP and SSR markers. Switchgrass mapping populations and association panels are in progress and will use marker-trait associations to develop improved varieties.

Results from the genetics and genomics analyses will feed into GLBRC research on increasing conversion efficiency. This work takes advantage of the endogenous variation for lignin that is present within various switchgrass accessions to select for plants with reduced lignification, thereby increasing microbial access to cell-wall carbohydrates. Again, knowledge of the regulatory genes responsible for lignin reduction will strongly impact this area of research.

GLBRC Genomics Needs: To fulfill the goal of improved switchgrass varieties, *Mike Casler* noted two critical items that will be required of the genome sequence: (1) ability to align sequences to a reference genome, and (2) unequivocal assignment of alleles to one of the two subgenomes. These achievements should allow the separation of true allelic variation as well as the ability to discriminate between subgenome duplication and tandem gene duplication.

Joint BioEnergy Institute

Pam Ronald reported on using rice as a model to obtain insights into switchgrass as well as investigate switchgrass directly. JBEI's research is focused on identifying BACs containing cell-wall and stress-related genes such as non-RD kinases, ethylene response transcription factors, glycosyltransferases, and glycosylhydrolases. Switchgrass BAC libraries were constructed to establish long-range connectivity among genomic sequences, obtain full-length sequences of coding and regulatory regions, differentiate homeologs and paralogs, and use in comparative analysis with other grass genomes. Libraries were constructed from the lowland cultivar Alamo, clone AP13, with 100K clones providing 16X coverage of the haploid genome. The HindIII/BstY1 libraries have low organellar contamination based on qPCR-based BAC library screening; pools, superpools, and filters for these libraries are available to the research community. Using rice-specific primers with a 10% success rate, they identified candidate switchgrass expressed sequence tags (ESTs), designed primers, and were able to extract the relevant BACs from switchgrass BAC pools.

From full-length sequences of 134 BACs, JBEI scientists were able to identify more than 300 cell-wall and stress-related genes, which were added to the switchgrass Affymetrix chip. In addition, HudsonAlpha/JGI has generated 360,000 BAC end sequences for a total of 7.2Mb for 47 randomly selected BACs. The repetitive sequence content, which accounts for approximately 26% of the sequence, was found to be mostly Class I transposable elements.

Five hundred switchgrass genes have been added to the Rice Phylogenomic Database to initiate a grass phylogenomic database, with a Gbrowse feature for switchgrass. Colinearity analysis with other grasses was conducted, resulting in the finding that switchgrass is most closely related to sorghum and maize, followed by rice and then *Brachypodium*.

JBEI Genomics Needs: *Pam Ronald* noted that a finished switchgrass genome will enable

- comprehensive and accurate estimation of cell-wall and stress-related genes,
- functional analysis of the aforementioned genes of interest,
- whole genome expression profiling of transgenic switchgrass expressing stress tolerance and altered lignin, and
- whole genome gene network analysis.

Additional JBEI needs include lines capable of self pollination, a pathosystem for *Botrytis* and *Puccinia*, and a publicly available database for access to newly generated switchgrass data.

Informatics needs include the establishment of a composite switchgrass network that is larger and more accurate than individual datasets. An example is RiceNet, a probabilistic functional gene network of rice containing 50 million data points and 23 types of datasets from five different species.

Bioenergy Science Center

BESC's goal, according to *Michael Udvardi*, is to develop improved switchgrass varieties with lower recalcitrance and higher ethanol yields from biomass while also retaining excellent agronomic traits. Towards this goal, BESC has developed several genetic and genomic tools. These include tools for gene expression and transgenics, clone libraries, gene sequence databases, genetic markers, mapping populations, and genetic maps. These resources have been used to characterize genetic diversity among switchgrass populations and map traits of interest. Work is ongoing to create mutagenized populations and construct physical maps.

BESC has several gene-centric projects as well, including the generation of ESTs from the switchgrass cultivars Alamo AP13 and Summer VS16. Using Newbler and PAVE, 120,000 unigenes were assembled from transcript and cDNA sequences. A switchgrass Affymetrix chip has been constructed with 120K probe sets. Additionally, a switchgrass Gene Expression Atlas is in progress with RNA-seq analysis in parallel. SNP discovery in AP13 vs. VS16 is ongoing.

BESC Genomics Needs: The genome sequence, according to *Michael Udvardi*, would provide the following resources, thereby facilitating BESC's work toward developing switchgrass cultivars with reduced recalcitrance:

- Complete set of genes for comprehensive studies, including high-quality annotation of coding sequences (CDSs), untranslated regions (UTRs), and promoters. (Question: How many contigs and how long do they need to be to achieve this goal?)
- Complete genome sequence with gene order and orientation, which would be extremely valuable. The alignment of switchgrass to foxtail millet is a good approximation but error prone. Fine mapping requires knowledge of precise marker location on the genome sequence.
- High-throughput marker system, requiring sequence from multiple genotypes for marker development.

Joint Genome Institute/HudsonAlpha

JGI's goal, according to *Jeremy Schmutz* (HudsonAlpha), is to produce an accurate reference sequence for switchgrass cultivar AP13. They aim to capture and order the gene space on each chromosome via mapping. Switchgrass genomic resources currently available from JGI/HudsonAlpha include BAC end and fosmid end sequences, 454 runs, and older Illumina datasets; sequencing of 221 BAC clones is in progress.

Jeremy Schmutz reported that the switchgrass A and B subgenomes differ by 2.5% and, inferring from analysis of 3' UTRs, alleles differ by 0.5%. Inspection of 24mers in Illumina data show that some regions on each subgenome are not shared, some regions are present in both A and B genomes, and some regions are shared between both.

The current Newbler assembly is composed of 8.3X linear 454 data and with N50 = 2.9kb, but for breeding efforts, knowledge of how much is collapsed among the subgenomes and how much is separated is critical. Based on Sanger ESTs and BAC end sequences, much of the gene space is separated in the 454 assembly. The proposed strategy is to separate haplotypes by sequencing BAC pools and then use these sequences to sort out the Illumina whole genome shotgun (WGS) sequence. A pooled BAC map could then be constructed to localize the subgenomes. Aligning 340 Mb of switchgrass assemblies to the foxtail millet genome usually, but not always, reveals the two subgenomes.

The challenge will be in taking this assembly from gene-scale contigs to chromosome-scale maps. One approach may be to use clone paired-end information. Colinearity is expected between the subgenomes and the genomes of foxtail millet and other grasses, but correctly linking alleles with subgenomes may be problematic. Large-scale linking and mapping will be needed to separate alleles into their respective subgenomes. An AP13 x VS16 mapping population is available that would be suitable for building a large binned map, and there are 40K contigs containing genes that need to be localized. Additional data and algorithms are needed to improve assembly, and even 1 Mb scaffolds would require several thousand markers to ensure proper placement.

A possible mapping scheme would be to identify informative SNPs in each contig through a brute force resequencing strategy. This strategy would involve deep shotgun sequencing of VS16 and aligning the reads to the AP13 genome sequence. Then, the F₁ offspring would be sequenced to 2X depth and loci mapped to chromosomes by cosegregation.

Another option involves the diploid species *P. hallii*. The 454 and Illumina-only assemblies are in progress, but these are highly fragmented and need additional resources. *P. hallii* is the same ploidy as foxtail millet and is highly inbred, so it is not directly useful for resolving the subgenomes in *P. virgatum*. However, since the two species shared an ancestor several million years ago, it might be useful for the local assembly and will be more straightforward than AP13.

Jeremy Schmutz highlighted several other potential resources that include

- Physical linkage: Laser capture microdissection (LCM), next-generation sequencing (NGS), BAC fingerprint map, and optical map.
- Genetic linkage: Develop array for 40K SNP makers and type them on a large population to build a dense genetic map.
- Ordering on *P. hallii* or other diploid *Panicum*.
- Additional sequencing resources that can be produced (new Illumina datasets, additional linear 454 coverage, low coverage Sanger sequencing, directed sequencing from combined BACs and WGS for QTLs and other regions of interest to the BRCs).

Next Steps: Moving the Community Forward

Shawn Kaeppler (University of Wisconsin) led a discussion to formulate concrete plans for moving this newly organized switchgrass community forward. Several reasons were articulated

for developing and implementing an action plan: Funding agencies have invested in switchgrass sequencing and want to see its maximum utilization; the research community wants to use genome sequence information for switchgrass improvement, but the best approach for achieving this goal has not been clearly defined; and funding agencies would like a process put in place that allows a unified switchgrass community voice to emerge.

With these reasons in mind, as well as the needs outlined in the previous sections, several action items were discussed and are outlined as follows.

1. Mobilize a Process to Elect a Switchgrass Executive Committee

Based on the ideas proposed and supported at the PAG meeting, an initial plan was put in place for electing motivated individuals to serve as an inclusive, unbiased voice of the community. To identify potential community members, request nominations, and hold the election, an ad hoc committee was composed of *Jeff Bennetzen*, *Christian Tobias*, *Pam Ronald* (who volunteered to initiate the process), and *Robin Buell* (who offered to make a Google group mailing list from the initial PAG mailing list and additional names from this meeting). *Christian Tobias* was appointed to organize the executive committee.

Criteria for the Switchgrass Genetics Executive Committee (SGEC) were defined. The elected committee will be composed of approximately six members who serve staggered, three-year terms, with term length to be determined by vote in the initial election. The elected committee will draft bylaws to be approved by community vote, identify community research priorities and needs, and organize regular research meetings.

As mentioned previously, seven representatives were elected in the summer of 2011 to serve on the SGEC (www.switchgrassgenomics.org/exec_committee.shtml). The committee is developing a charter and holding regular bimonthly teleconferences. A meeting at PAG in January 2012 has been arranged.

2. Develop and Maintain a Community Website

The community website will be a basic portal for information dissemination and communication among members of the switchgrass research community and provide access to public sequence data and genetic resources. *Robin Buell* volunteered to oversee and maintain the website through a neutral URL (www.switchgrassgenomics.org). PDF files of previously mentioned presentations have been posted on this site. Through this website, the SGEC is working toward improving information dissemination to the entire community.

3. Produce a White Paper on Switchgrass Genetics/Genomics and Future Needs

A white paper or review and perspectives paper that documents the current state of switchgrass genetics, genomics, and the potential uses of these resources, published in a refereed journal, would provide a valuable outreach component for the community. This effort, which was initiated at the April 2011 meeting and involved the collaboration of several switchgrass research community members, resulted in a recent publication in *The Plant Genome* (Casler, M. D., C. M. Tobias, S. M. Kaeppler, C. R. Buell, Z.-Y. Wang, P. Cao, J. Schmutz, and P. Ronald. 2011. "The Switchgrass Genome: Tools and Strategies," *The Plant Genome* 4(3), 273–82). (Reference link: www.crops.org/publications/tpg/abstracts/4/3/273.)

4. Identify Priorities and Collaborations to Improve Currently Available Sequence

Since obtaining and maintaining a reliable seed source for selected accessions is quite difficult for switchgrass, a standard protocol for confirming the genotype of accessions is needed. This information is critical for the correct interpretation of genetics and 'omics studies.

Cryopreservation is successful in some species, and a switchgrass study has been initiated at the Noble Foundation.

Switchgrass breeders have clearly stated that for switchgrass genomics resources to be effectively used in their programs they must be able to separate alleles and discriminate between subgenomes. The question of how many progeny would be needed to separate out the linkage groups elicited discussion and varied opinions. One idea put forth, to create a nested association population, would require large amounts of field work.

There should be a process in place to track and facilitate progress in the genome sequencing effort. As part of this process, several priority activities were suggested including assembly, subgenome identification, annotation, exome assembl(ies), genotyping resources, and expression resources. Other priorities will surely be identified as sequencing and research progresses.

Summary

Since the first organizing meeting in January 2010, a switchgrass community website and mailing list have been established, the SGEC has been elected, and at least two meetings are being planned for 2012. A switchgrass review and perspectives paper has been published. The switchgrass genome sequencing effort is progressing well, with annotated contigs and localized alignments to foxtail millet. There is an effort to reach out to switchgrass researchers from other disciplines such as ecology to develop a broad, well-rounded community. Importantly, the community now has a structure and organization that will enable the best possible science from the switchgrass genome sequence.

APPENDIX

Plant and Animal Genome XIX Meeting

January 2011

Responses to Charge Questions

Participants were given seven charge questions prior to the organizational meeting on the development of switchgrass as a study system and target for improvement as a bioenergy crop. This meeting was held on January 18, 2011, at the Plant and Animal Genome XIX meeting in San Diego, California. Responses to the charge questions were collated and compiled into relevant categories, described here, although there was significant overlap and synergy between several of the responses.

1. What type of research is your laboratory doing currently on switchgrass?

Basic biology and **molecular biology** provide the foundation, tools, and resources for other research areas. Fundamental studies on seed dormancy, germination, enzymology, and cell-wall metabolism, including cell-wall traits affecting conversion into biofuels, are ongoing in several laboratories. Transformation of switchgrass with rice gene(s) to improve deconstructability is under development, as is the generation of switchgrass bacterial artificial chromosome (BAC) libraries that will be useful for analysis of the genome sequence ([see responses to charge question 6 on pp. 15–16](#)).

Breeding efforts continue and use a combination of both conventional field-based and molecular techniques. Both upland and lowland types are of interest, and methods employed for characterization include molecular markers, progeny testing methods, and recurrent selection. Traits of interest include improved germination, quality, biomass yield, disease resistance, and reduced lignification. The development of new, improved cultivars is the ultimate goal, and methods to attain this goal include the development of experimental synthetics, inbred lines for hybrid cultivar development, and haploids.

Breeding strategies are intimately associated with **genetics**. Respondents reported work on genetic mapping (association mapping, linkage mapping, and quantitative trait locus (QTL) mapping) for selected agronomic traits related to biomass yield. Development of molecular markers such as simple sequence repeat markers (SSRs) and single-nucleotide polymorphisms (SNPs) are critical for mapping and diversity studies. Respondents also reported on the development of heterotic gene pools and hybrids and on cytogenetic studies (e.g., fluorescence *in situ* hybridization (FISH) and karyotyping).

The rapidly expanding field of **genomics** is evident in the types of current research reported by the respondents. Genome-wide selection and candidate gene analysis is being performed in many laboratories to identify and validate genes involved in lignification, cell-wall metabolism, defense response, and many other bioenergy-relevant traits. Several genomics technologies are being employed (transcriptomics, metabolomics, and genomics) as well as comparative phylogenomics to identify switchgrass homologs of key genes from rice and other more well-

studied grass species. Genome structure and gene organization, reduced representation sequencing for genotyping, identification and characterization of microRNA populations, and simulation modeling were also noted.

Another interesting area of switchgrass study is **evolution and adaptation**. Research on the origins of polyploidy, ecotypes, and regional diversification are being conducted, and issues of regional and local adaptation to abiotic stresses (including drought tolerance, winter hardiness, and switchgrass-insect interactions) are being addressed. The development of tools and resources for studying local adaptation in the diploid switchgrass relative *Panicum hallii* is ongoing, as well as modeling yield performance under future predicted climate scenarios.

2. Results to date?

In the area of **basic and molecular biology**, respondents reported increasing understanding of the molecular mechanisms controlling seed dormancy and germination. The genes and proteins important to germination, plant development, and lignification have been cloned and characterized. A real-time fluorescent assay for caffeic acid O-methyl-transferase (COMT; a key enzyme in the lignin biosynthetic pathway) has been developed. Switchgrass genotypes that differ in cell-wall properties were characterized for cell-wall accessibility and ethanol production. Cinnamyl alcohol dehydrogenase (CAD) proteins, which function in lignin biosynthesis and plant stress response, have been biochemically characterized, and a CAD-silenced switchgrass was developed. Progress was reported in the construction of two high-quality switchgrass BAC libraries, each containing 101,376 clones with an average insert size of 144 and 110 Kb, respectively; pools and superpools have been generated for both.

Switchgrass **breeders** reported results on diversity, population development, and cultivar releases. Biomass yields were found to vary widely between sites, and some cultivars had better establishment across these different marginal sites. There were also significant differences in disease resistance and biomass yield between ecotypes. Eight regional gene pools for cultivar deployment have been identified, and improved populations have been developed. Biomass yield was increased by approximately 10% between 2000 and 2006, a new lowland switchgrass—‘Cimarron (SL 93 2001-1)’—was released in 2008, and an array of 19 new synthetics has been created and tested in more than 10 locations in Oklahoma, Texas, Arkansas, and Kansas.

Several new **genetics** resources have been developed and established by the switchgrass research community. A large set of SSR markers was developed and published in 2010, and new mapping populations from *P. virgatum*, including a QTL mapping population, are being developed. An expressed sequence tag (EST)-SSR map of switchgrass has been constructed. QTL and association mapping projects are ongoing. Primary and secondary centers of switchgrass diversity and two primary lowland and three upland gene pools have been identified. Microsatellite (SSR) diversity analysis has been conducted.

In the **genomics** arena, efforts to identify candidate bioenergy-relevant genes are underway. A quantitative polymerase chain reaction (PCR)-based high-efficiency screening method has been established to identify switchgrass orthologs of key rice genes. One hundred eighty-three BACs carrying potential cell-wall-related sequences have been identified, and most have been fully sequenced; annotation and analysis are in progress. Forty-seven random BAC clones have been analyzed for repetitive regions and gene organization and integrated with existing rice databases of glycosyltransferases, glycosylhydrolases, and kinases. Transcriptome analysis of ‘Summer’

crowns and rhizomes using 454-FLX is complete; assembly and annotation is in progress and is expected to be available in 2011. Restriction-site associated DNA (RAD) sequencing and RNA-seq methods are being developed for *P. virgatum* and *P. hallii* to study population structure and conduct genetic mapping and gene-expression studies in these species. ESTs are being sequenced by 454 from normalized *P. hallii* cDNA libraries constructed of various tissues and environmental conditions. In collaboration with the Department of Energy's Joint Genome Institute, 454 and Illumina short sequences are being generated from this species for genomic tool development. Over 50 million small RNA reads have been generated from different switchgrass tissues subjected to various stresses, and 20 families of conserved miRNAs, some of which are novel and switchgrass specific, have been identified. The chloroplast sequences of 'Summer' and 'Kanlow' cultivars are available.

The effect of latitude and longitude on **adaptation** and agronomic traits has been quantified. *P. virgatum* clonal replicates and the "Albany" mapping population ('Alamo' x 'Kanlow' F₁) are planted under rainout shelters. Physiological and transcriptome data will be collected during the summer of 2011 to determine drought tolerance and related parameters.

3. Do you collaborate, and if so, with whom?

Collaborations were documented with 39 institutions throughout the United States, including academic, government, and private:

Affymetrix, Santa Clara, CA
Appalachian State University, Boone, NC
Ceres, Inc., Thousand Oaks, CA
Clemson University Genomics Institute, Clemson, SC
Cornell University, Ithaca, NY
Creighton University, Omaha, NE
Ernst Conservation Seeds, Meadville, PA
HudsonAlpha, Huntsville, AL
Joint Genome Institute, Walnut Creek, CA
Kansas State University, Manhattan, KS
Michigan State University, East Lansing, MI
Oklahoma State University, Stillwater, OK
Oklahoma State University Biobased Energy Center, Stillwater, OK
Rutgers University, New Brunswick, NJ
Samuel Roberts Noble Foundation, Ardmore, OK
Sandia National Laboratories, Albuquerque, NM
South Dakota State University, Brookings, SD
Texas A&M University, College Station, TX
U.S. Department of Agriculture (USDA)-Agricultural Research Service (ARS),
Albany, California
USDA-ARS, Ithaca, NY
USDA-ARS, Kingsville, TX
USDA-ARS, Lincoln, NE
USDA-ARS, Madison, WI

USDA-ARS, Peoria, IL
USDA-ARS, St. Paul, MN
USDA-ARS, Temple, TX
USDA-ARS, University Park, PA
University of Arkansas, Little Rock, AR
University of California, Berkeley, CA
University of California, Davis, CA
University of Georgia, Athens, GA
University of Minnesota, St. Paul, MN
University of Nebraska, Kearney, NE
University of Nebraska, Lincoln, NE
University of Nebraska, Omaha, NE
University of Oklahoma, Norman, OK
University of San Diego, San Diego, CA
University of Texas, Austin, TX
University of Wisconsin, Madison, WI

4. What are your goals, and when do you hope to achieve these goals?

The **breeders** contended that cultivar improvement is always needed and is ongoing. They visualized combining traits and using marker-assisted breeding and haploids to develop new cultivars, with a commercial switchgrass hybrid being developed by 2020. An antibody-based protein chip for phenotyping, the development of markers and diagnostics to phenotype cell-wall traits in germplasm, and the identification of conserved and novel microRNAs were seen as short- to midterm goals for **genetics** research.

The generation of basic **genomics** knowledge for the improvement of switchgrass as a feedstock for biofuels is the underlying theme, and integration of translational genomics into functional genomics for traits of relevance to bioenergy, such as the regulation of cell-wall metabolism, follows. Development of a pipeline to scaffold Illumina data onto available 454 and EST resources is a short-term goal, with eventual data incorporation into gene regulatory networks and functionality studies. Genome-wide selection should be implemented soon, followed by gene discovery, SNP development, and marker selection for reduced lignin. Cloning and characterization of other key genes and proteins is anticipated within the next 3 to 5 years. Using computational and proteomics tools, identification and characterization of genes involved in traits such as defense response and stress tolerance will occur over a longer timeframe.

Improved understanding of the **evolution** of modern switchgrass is seen as a feasible goal within 1 to 2 years, and understanding the control of winter hardiness and mechanisms of resistance to potential insect pests should be accomplished within the next 3 to 5 years. Other midterm goals include functional validation of individual disease-resistance genes and the identification of microRNAs associated with drought, salinity, cold, high temperature, and nutrient stress. Understanding of how genetic variation impacts **adaptation** in natural populations should ultimately lead to the breeding of agronomically suitable cultivars.

5. What are your future plans for switchgrass research?

Future plans for **basic and molecular biology** research in switchgrass include the examination of lateral root formation through the analysis of gene expression and cell-wall content and the transformation of switchgrass with promising genes from other grass species. Potential disease-resistance genes will be tested in rice and used to engineer disease-resistant switchgrass. Further insights into traits impacting the conversion of switchgrass biomass into biofuels are anticipated.

In the area of **breeding and genetics**, the goal is to translate basic research into improving switchgrass as a bioenergy crop. High-throughput phenotyping and diagnostics are being developed to quickly and efficiently screen lines for traits such as cell-wall properties, biofuel yield and quality, stress tolerance, nitrogen-use efficiency, and reduced lignification. Development of molecular markers for diversity studies and their direct application to breeding programs is planned. These markers are needed for QTL-based, marker-assisted selection; association and linkage mapping; genome-wide selection; and genotype-by-sequencing (GBS) strategies and to provide empirical data on linkage disequilibrium in breeding material and on the practicality of association mapping. Development of hybrid and synthetic cultivars using field-based breeding and molecular tools such as molecular markers will be pursued, as well as the development of heterotic gene pools.

Genomics analysis will be used to provide insights into bioenergy-relevant traits, and ultimately traits that impact the conversion of switchgrass biomass into fuels. Analysis of gene expression and genome structure are ongoing and will continue. Candidate gene analysis and association mapping with SNPs in these candidate genes, for traits such as the regulation of cell-wall biosynthesis and lignin concentration, are planned. Genome-wide selection, coupled with simulation modeling and methodology development, will facilitate breeding efforts. Post-transcriptional gene regulation controlling biomass production and stress tolerance characteristics will be investigated further. Continuing annotation of the switchgrass genome and the enrichment of existing phylomic databases such as the rice glycosyltransferase, glycosylhydrolase, and kinase databases with switchgrass data will enhance genomics tools available to switchgrass researchers.

Studies on **evolution and environmental adaptation**, such as polyploidy, ecotypes, regional diversification, and stress studies, will continue. Also of interest is identifying potential disease resistance candidate genes, testing these genes in rice, and using validated genes to engineer switchgrass lines resistant to the potential pathogens *Puccinia* and *Botrytis*.

6. What research needs or gaps do you see that, if filled, would facilitate your work and that of the research community as a whole?

Significant research needs and gaps were identified in all topic areas. For **basic and molecular biology**, work is needed on basic reproductive biology, including self-incompatibility, pollen-longevity, post-zygotic compatibility, interploidal crosses, and other hybridization barriers. An efficient transformation method for switchgrass is essential to further work in improving switchgrass cultivars using transgenics. As an essential sideline to these efforts, practical regulations to enable research on transgenics, such as gene flow studies, are needed.

For **breeding and genetics** efforts to move forward, an annotated genome sequence will be needed as well as increased bioinformatic efforts to translate genomic and transcriptomic data into forms useful for breeding. Systematic germplasm collection and characterization representing the genetic diversity of the species/genus needs to be established, in coordination

with USDA's National Plant Germplasm System (NPGS). There is currently no efficient way to screen germplasm. High-throughput phenotyping methods for biofuel quality traits, drought tolerance, nitrogen-use efficiency, and other traits are greatly needed for genomic selection and would dramatically improve the speed of the breeding process. These phenotyping methods will require an increased capacity for field trials that will accommodate the large mapping populations needed to assess phenotypic response in a multi-environment field trial network. A cultivar testing network managed by a public agency could be established. Also needed are high-throughput SNP analysis platforms, the ability to map markers or QTLs to specific chromosome homologs, and methods to analyze polyploidy and gene expression in tetraploid switchgrass.

First and foremost in the **genomics** arena, calls for a full-scale, reliable reference genome sequence or genome representation were voiced. Among other things, a switchgrass reference sequence would allow GBS strategies and mapping and annotation of smRNA reads and targets. Resequencing and gene space assembly efforts could then extend to other switchgrass cultivars and related species such as *P. hallii* for comparative studies. A gene expression atlas or gene catalog for switchgrass would facilitate the translation of results from other organisms. Badly needed are bioinformatic efforts to translate genomic and transcriptomic data into forms useful for breeding such as the correspondence of traits to sequence information, which is required for genomic selection.

A multi-environment field trial network is needed to study genotype-environment interactions as well as the **evolution and adaptation** of traits to a changing environment. A pathosystem should be established for profiling the potential switchgrass pathogens *Puccinia* and *Botrytis*. As noted above, gene flow studies are needed to enable responsible research on transgenic switchgrass, and practical regulations need to be instituted.

7. What community tools or resources (e.g., website, database, newsgroup, mailing lists, annual meeting, stock center) would facilitate your work or the growth of the switchgrass community?

All respondents agreed that **website and database accessibility** are critical components to facilitate switchgrass research. Forward thinking about this issue will be necessary, especially on how best to develop web-based systems for handling the deluge of newly generated data (e.g., genomics, transcriptomics, and proteomics) that provide relatively easy access to all researchers. Some of the specific types of databases and analyses that would be useful are a SNP database, tools to perform comparative analysis of switchgrass genes with other more well-studied grasses, data storage for first draft and finished annotations of genome and full-length BAC sequences (in collaboration with HudsonAlpha), and Gbrowse. This aggregate genomics database/website, parts of which are in the development phase, should integrate all relevant phenotypic, genotypic, expression, and sequence information such as "SoyBase", and be made easily accessible. This site should also be a place to share genotypic and phenotypic data from multi-environment field trial networks, similar to the Barley CAP project "The Hordeum Toolbox" (<http://hordeumtoolbox.org/>).

In general, the respondents felt that a **newsgroup or mailing list** would be helpful for sharing plant material and protocols, coordinating field work, and providing updates on resources. A web-based forum would be useful for collecting, cataloging, distributing, and communicating research information, similar to the forum in use by the U.S. wheat and barley scab initiative

(<http://scabusa.org/home>). However, at least one respondent indicated newsgroups or mailing lists may not be that useful.

While it was generally felt that holding regular switchgrass community **meetings** would be beneficial, preferences on how often to hold these meetings varied from annual to biennial to just a single switchgrass breeding/genetics/genomics conference to get groups organized and avoid duplicative efforts. A two-day, biennial meeting on switchgrass genetics/genomics was proposed, with the first day for brief presentations and a second day for discussions within full or sub-groups to identify critical needs and formulate appropriate action plans. Others felt a yearly meeting would be most useful, possibly at the Plant and Animal Genome conference. Ultimately, a more inclusive and extensive switchgrass meeting would be ideal. In any case, all deliberations should be made freely available to effectively identify and formulate collaborations. The key will be to (1) set up a dialog to help deploy research findings into translatable breeding gains, and (2) understand needs along the conversion chain (harvesting, storage, transport, and conversion).

Germplasm is already well-curated through NPGS, but a **stock center** for other types of switchgrass germplasm such as mutants would be useful.