Switchgrass Research Group Meeting-Draft Notes

January 18, 2011

Introduction and Rationale for Meeting

The purpose of the meeting was to discuss the state of the art for switchgrass research, assess the current state of science and where it's going, and make steps to collaborate and organize research within the community so that critical gaps are avoided and research is completed more efficiently. Another key motivation for this meeting is to increase communication within the switchgrass community. The meeting was co-chaired by Jeff Bennetzen [University of Georgia and the DOE BioEnergy Science Center (BESC)] and Tom Juenger (University of Texas, Austin) and organized with Catherine M. Ronning of the U.S. Department of Energy Office of Biological and Environmental Research.

Jeff provided some perspective about how other plant research communities have coordinated research. He said, for example, the *Miscanthus* community is highly organized and has comprehensively covered needed research areas "from A to Z." This hasn't been done for switchgrass. Working together, the switchgrass group can identify and close research gaps and improve research efficiencies—this was the main motivation for the meeting. The people at this initial meeting represented the nation's switchgrass community. There was a sense that broader collaborations would be beneficial to switchgrass research.

What are the research gaps this community can address, and what are their priorities? When the maize research community organized, the funding agencies were pleased to hear what the community needed. What is critical is that the agencies got a consensus from the community and therefore listened to the community's needs and took them seriously. This isn't like a single investigator feathering his or her nest—it's the community speaking. *Arabidopsis* was the first community to organize around research needs. Other communities such as maize, sorghum, and soy have benefitted from this approach as well.

The incentive for forming a group like this and for making efforts to coordinate research and communicate is that if everyone keeps doing their own piece of switchgrass research, it will be hard to leverage funding from the agencies.

Because everybody in this room won't be interested in all issues, the group discussed creating national collaborative subgroups focused on specific research problems.

Summary of Action Items and Next Steps

Throughout the discussion, the group identified several action items summarized below. These include (1) sending a revised questionnaire to participants, (2) forming a switchgrass executive committee, (3) 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.

Sending Participants a Revised Version of the Pre-Meeting Questionnaire

Everyone will be asked to vote for what is at the top of their priority list and to provide annotation about their chosen priorities. Among the questions will be new ones such as:

- What are you going to do with the sequence?
- What more do you need from the sequence?
- Why and how can we provide that to you?
- How often should this group meet? Should it meet annually or semiannually?

Also, a participant said they'd like to see names associated with research areas in the pre-meeting materials and with answers to the charge questions. This material has now been distributed by Cathy Ronning (DOE BER).

Forming a Switchgrass Executive Committee

Also discussed was the idea of creating a switchgrass executive committee to represent the community and organize and communicate information within it.

The maize community's executive committee was discussed as an example. Jeff said that it is the committee's job to find out what that research community needs and to communicate to anyone who needs to hear about it. The committee's function is to organize and communicate, not to tell people what to do. For example, the maize executive committee conducts an annual survey to ascertain community issues. They also do write-ins and communicate this information more broadly outside the community. People in Washington listen. *The executive committee organizes information and communicates in a deliberate way.* In light of this, perhaps there is a better name for it than "executive committee" ("representative committee" was also used later in the discussion).

As far as the composition of the switchgrass executive committee, there is no money involved in serving on it. It is an honor to be elected to it by your peers. If the community identifies, for example, six research priorities, there should be representation on the committee in those six areas.

A revised pre-meeting questionnaire will be sent to the group asking them to identify their research priorities. Once responses are received, nominations for the executive committee will be made.

How big will this committee be? Additional people who weren't at the meeting can be suggested. It's OK to get redundant names.

Defining the Group's Scope and Expanding its Mailing List

Participants discussed the need to define the community's scope and the kinds of groups they should invite to join. This group's scope involves switchgrass genetics and genomics important for plant breeding. Although we aren't talking about, for example, agronomy and restoration, a lot of researchers in these areas are interested in our genetics work. Should we invite others?

In the maize community, the executive committee set the scope in the mission statement. It would have been good to have sent the mission statement to the entire maize community to get their approval, which would have represented a broader consensus than just the executive committee approving it.

This group is starting with mostly geneticists and breeders. Should we expand our mailing list to include groups like environmental researchers? Cathy Ronning of DOE said there is a larger mailing list available.

Some participants believed the question of "What are we going to do as a group," needed to be answered before the group decided who else is going to join them. Is the group going to have a scientific meeting? What would be the focus of that? Would it encompass everything about switchgrass or just genomics and genetics? Someone responded with: breeding, genetics, and genomics.

In answer to whether other groups should be included, one person said the group should add evolution, ecology, and physiology. If ecologists hear about this, they'll want to join. Jeff said the University of Georgia has 12 labs that do work as part of BESC and they could be invited .

Someone suggested this group's scope is from research to release—the whole spectrum. Cell biologists are contributing to switchgrass research. One person asked, "If the group becomes too big, will it be unwieldy?" A response was that the executive committee will do the grunt work.

Someone asked how big this group is envisioned to be. [Mike Casler responded] Two hundred to 300 people perhaps. This is about 80 labs. Fifty people from the DOE Great Lakes Bioenergy Research Center (GLBRC) alone could join. Then there is the restoration and conservation community. Many of these would be interested.

Jeff liked the idea of being inclusive. The executive committee itself will be small.

Someone [Christian Tobias] suggested the group should exclude deconstruction researchers who are thinking of switchgrass as a substrate to deconstruct. Jeff was OK with this.

Who the group invites versus who will attend is an iterative process. Over time, the composition of an organization may become fixed. In the interim, it could fluctuate a lot. In the next group mailing, it was suggested to ask the question of what groups to include. Attendees were encouraged to go home and think more about this issue.

When thinking about potential interested groups, there are the primary utilizers of the genomics and genetic sequences and there are the secondary users. There is also oversight. There are even people who don't want us to work on switchgrass at all. Someone [Gautam Sarath] said making the group too broad will dilute anything that anyone brings to the table, so start small. Another person [Mike Casler?] thought the group should invite everybody.

Evaluating the Idea of an Annual Switchgrass Meeting

It is clear from this meeting that things will move fast. In addition to an annual meeting, thinking of other ways to stay connected is good too (an annual meeting shouldn't be the sole avenue of communication within the group). A simple beginning of this would be to have a workshop and then have a meeting of the executive committee. We could emulate other groups in that they

meet periodically as part of a larger workshop or conference. Many in the group attend PAG anyway, so meeting here would not mean making an additional trip; it would just mean an extra day.

Someone said the group needs a meeting soon.

One idea was to plan an extended afternoon session where talks are limited to 10 minutes and there is time to have this type of discussion. The group should open the session up to anyone who wants to attend. Meeting for a whole day is challenging—the session needs to be relatively compact. If there is non-overlap in a researcher's interest level, then he or she wouldn't have to stick around.

Because a number of people don't come to PAG often, the question was asked whether anyone thought a stand-alone meeting at another place and time was a good idea. Someone mentioned the Noble Foundation because it is a centrally located nonprofit. The response was that it depends on how many people come. There are plenty of hotels in the area; attendees wouldn't have to stay at the lodge.

If such a meeting were to be organized, it would require support. Is this justified?

Sharlene Weatherwax (DOE BER) said that if the focus initially is to develop the genetic and genomic resources the community can use, it makes sense. She's thinking about the PAG or ASPB meetings. She said meeting at PAG initially is logical. We'll be talking about genomics a year from now. Further out, we'll have to re-evaluate. Agencies are always trying to reduce costs for people. Costs would be higher at a stand-alone meeting. Weigh this against a wiki that can be used for communicating. She said it would be up to the community to decide.

Someone noted that many places have weekend discounts. Maybe a stand-alone meeting in the Southeast wouldn't be so expensive. There are meeting options that are centralized.

There was agreement that an annual meeting is a good idea. The group needs to resolve what it will do next year. The idea of starting small and inexpensively was appreciated by the group. It would have been great to have heard talks from everyone at this meeting about one aspect of their research, so maybe this is something that can be planned for future meetings. We could slot times for six to eight talks. The group would have liked a presentation on the status of the genome, and it would have been nice to have a slide or two. *The group needs more background information on what everyone in the community is doing before it has a discussion.* The group will have discussions over the next few months and will make a decision that will work. Most everyone will be at PAG next year.

Creating a Wiki for Information Sharing

The group further discussed the value of creating a community resource such as a dynamic website or a wiki that fosters information exchange and communication, key goals for the community.

They will set up a real-time information-sharing site (e.g., a wiki). They talked about different options for finding someone to help with this. The community will set up and administer the wiki because it needs to come from them.

Cathy said once the notes from this meeting are finalized, she will send them around to everyone. They then can revise these notes on the wiki.

An Example Community Issue and Other Suggestions

An example community issue brought up by Jeff: He is uncomfortable that any transgenic switchgrass would go out that isn't male sterile. Another person said the companies don't ever guarantee 100 percent sterile. An example was given that there is a lot of acreage with domesticated loblolly pine which will soon lead to an extant loblolly pine population that will be domesticated only (none of this is transgenic). This is an example of what could happen.

A participant [Christian Tobias] suggested that it might be useful to have an analysis of research resource allocation. Over time, the balance of the resource allocation will change as needs change (e.g., shifting more resources between genotyping vs. phenotyping).

Someone asked whether anyone in the group was doing marker-assisted breeding. One person said that he is in the process of developing markers. Following marker development, he will map the traits.

Considering the Role of the Commercial Sector

Were any private companies invited to this meeting?

CERES was invited but was unable to attend. A couple of group members are collaborating with them. The maize community excluded commercial entities because these entities had their own voice. Maize commercial entities work with the community when needed.

In the role of communicating what the switchgrass community needs, would the commercial sector be a good partner? Looking at the maize model, things that the community wants may be things that the commercial entities *don't* want. Private companies could be involved in going to agencies and determine what should be going on in the public sector. People will take what the research community generates and uses, so having their input at a higher level makes sense.

Is this the way it is going to be? For some crops, the public sector will produce the seed. Right now it works both ways in switchgrass. There are public varieties, but the trend is toward privatized. What is the degree to which the group needs to support public and private as an option? It depends on the breeding program. The government and a group like this shouldn't regulate. It should be up to the breeders.

Should public breeders only breed species that aren't covered in the private sector? Every program in the private sector is a separate program. It is unclear how viable they are or how well funded they are. So switchgrass is now public breeding. It will be for more than 10 years.

People in this group should be sharing information equally. Currently, they don't share materials much. If they give each other seeds, material transfer agreements (MTAs) are needed because of regulations. The community shares things on a personal basis, but they do share info, and that's what they want to do as part of this effort to organize and communicate. *Information sharing is a key activity of this group.*

The private sector might not care about the societal benefits of growing switchgrass vs. *Miscanthus*. They don't care about diversity. Their goals are different from the group's. This is expected, but the group will take this into account when it hears anything from them. They won't tell other companies what their priority goals are.

The group would like to have the private sector here to hear their thinking, but they wouldn't want the private sector to dominate.

The decision to invite companies is not irreversible. They won't be offended if we tell them we have to go on our own. *The sense of this group was to leave their options open as to how long private companies could participate.* Private companies tend not to drive the conversation. If they steer the conversation, they are giving away information. Sometimes they'll speak up and say they already have something.

An example downside to working with the private sector: The maize community was promised EST libraries by the private sector, but it tried to tie conditions to the libraries. This caused a huge delay.

The Status of Switchgrass Sequencing and Discussion of Community Needs

The technical discussion on sequencing was framed by the question, "What will researchers do with the sequence?" The genome sequence *will* be completed. What then does the community need from it, and what are the issues pertaining to genome sequencing?

Dan Rokhsar from The DOE Joint Genome Institute (JGI) said we have two sources to support switchgrass sequencing at JGI: stimulus monies and DOE Bioenergy Research Center (BRC) sequencing. They are at the point where they have 8X coverage of 454 sequence and lots of coverage in Illumina sequence. That sounds like a lot, but as you know, switchgrass is a complicated outbred tetraploid. The data are publicly deposited. They are using the foxtail millet genome as an organizing substrate as a first pass for these data. They are mapping all switchgrass data onto foxtail millet. This will define localized collections of reads that come from the orthologs for localized assembly. They will have this organizing substrate in a couple of months of Alamo 13 that was chosen for sequencing.

In parallel, the JGI is working with several people to look at bona fide diploid switchgrass as a genomic model. This may provide useful sequence earlier because it is less complex. Dan will send out more regular updates to keep people apprised of where they are. This mailing list will help to let you know which data are available and the resources they are generating. Dan hopes we can assemble the gene space plus their surrounding genomic loci. He has talked with Christian Tobias and Tom Juenger about doing very dense marker sequence-based mapping. There will be two genes per locus, and they'll place them on analogous chromosomes. This will take some time. However, it will be less time than it takes to breed a new variety of switchgrass. This will take about a year.

A person in the group said that last year we talked about a linkage map approach to pull things together and you were hoping to get some things improved. The population you'd use would be AP13 x VS16 population. Now we have more than 300 plants in the greenhouse. There are 1200 markers, not the dark markers.

Dan said the conservative scenario we've thrown around is that if we can assemble each gene (each copy of a locus), we'd need 50,000 markers. We'd need 1X coverage of every member of your population. This would guarantee markers every several kilobase gene. If we want to map every gene in every experiment, we need a marker in every gene and this pushes us to brute force approaches. Given the interest and resources, we should proceed with this. It won't give an order, but it will allow you to know which genes are linked on different versions of chromosomes.

What about whole genome profiling? Maybe we'll use that in alfalfa and brassicas. There's a good separation based on sequencing.

Jeremy Schmutz and Jerry Jenkins at HudsonAlpha are doing the BAC end sequencing. This would be a nice resource at this scale.

The best we can do now is assume local colinearity with foxtail millet. They are not as closely related as sorghum and *Miscanthus* are to each other. There will be intermediate steps where we've separated out the subgenomes. Dan said it won't be perfect; maybe it will be version 2 of the genome.

How stable has *Panicum* been? Jeremy has information on colinearity of BACs. There are definitely several genes in a row that are in the same order.

It is hard to predict in this class. Pearl millet has an unstable lineage. *Panicum*'s genome is 550 megabases. Some of the original sequencing that JGI has done: 0.2 percent heterozygous sites. The allelic in ___ is said Dan. This is the genomic model that helps us avoid some of the issues.

A comparison between *P. hallei* and foxtail millet would be informative. Want to tie this into genes, QTLs. If you look at orthologous gene candidates….could lay current switchgrass data on top of foxtail millet orthologs regardless of whether there were rearrangements.

Orthologs can be lost due to polyploidy.

If you are after gene discovery with the sequence you've got now; the sequence data is there. How do you tell if an allele is on one homologue or the other? This is a massive problem for breeders who are doing gene discovery. This is important at this moment in time. They have to be able to do this by homologues.

Dan is curious, one of the steps would be to start skim sequencing the mapping population. He would like to get a priority from the group regarding this activity as to the work JGI needs to do.

[Note: The following notes are rough, but the question in bold and the ensuing discussion are important for efficiently using sequence and performing the extensive follow on research.]

The genome sequencing will get done. What are we going to do with it? *The sequence isn't enough. We need to know haplotypes for large stretches of chromosomes. For the gene discovery work and genome selection we're doing, this is what we need. Is this a lot of BAC end sequencing? This is one possibility, or could we split things apart to do haplotypes? It depends on the haploptype for each variation. We need to determine which homologue it came from.*

To calculate allele frequencies, for every cycle of selection, breeders have to know where it is to feed it into that spot for the predictive equation for discerning the breeding patterns. The only

way to know that is to consistently identify the two haplotypes that the sequence is falling in. We are doing bulk breeding, so we have a lot of haplotypes.

We are also going into diverse populations, and we don't know the genetic structure. There are lots of alleles. We need to factor these in for breeding.

[Yanqui Wu?] We know the homozygosity level is high….we examined 1400 homozygotes. This is a lowland tetraploid. The starting material was from the breeding population started in 1994 by his professor. He continued recurrent selection. About 60% are selfed seed (?).

Christian: We've got some similar resources. We've got a spontaneous dihaploid (?). There are other resources that are becoming available.

- You aren't getting segregation distortion?
- Compared to alfalfa there is still good rigor.
- One of the things mentioned in the list…a mapping population is the only one that makes great sense initially…either association or recombination mapping.
- Do we need a centralized process for doing this kind of thing? If we were to go out and say we wanted AP13 x VS16 to be mapped and phenotyped by everybody, would people jump on board? Yes, at specific loci.
- There is association mapping where they are related and a situation in breeding programs where they aren't
- Is this an "either/or" or an "and"? In maize, it is an "and." We also say we generate additional populations based on a density…
- How do you maintain the maize population? Answer: you need a lot of space. Someone keeps the seed. Does the seed exist for the switchgrass population?
- They are outbred. They could have phenotyping jamborees. These are maintained clonally. He is doing soil microbe stuff. Have to do exenic cultures. It is a nightmare to maintain for that study. They could request support for this type of activity. Have to determine if we'll be collaborative enough to share the info.
- *The genome sequence will get done. But we need to hear from you what we are going to do with it. What are you going to get done? The sequence isn't enough. Need to know haplotypes for large stretches of chromosomes. For the gene discovery work and genome selection we're doing, this is what we need. Is this a lot of BAC end sequencing? This is one possibility or could we split things apart to do haplotypes? It depends on the haploptype for each variation. Need to determine which homologue it came from.*
- *In order to calculate allele frequencies for every cycle of selection, breeders have to know where it is to feed it into that spot for the predictive equation for the breeding patterns. The only way to know that is to consistently identify the two haplotypes that the sequence is falling in.*
- *You are doing bulk breeding; you have a lot of haplotypes.*
- *We are going into diverse populations, and we don't know the genetic structure. There are lots of alleles. We need to factor these in for breeding.* We have a lot of parents. He isn't quibbling about Alamo 13. This sequence would get him started. He'd lose 100s of

thousands of SNPs out of millions of SNPs, but he'd catch a lot of SNPs and a lot would be useful.

- What is the number for divergence of the two homologues? Jeremy: 0.5 percent; 2.5 percent based on 3 X QTRs….how many years? He has no idea. He doesn't know the clock number. He's never seen this on a figure.
- Have close to 500,000 Sanger sequences
- Ability to turn shotgun reads into gene models is good. Getting them all into *cis* is the challenge. What is the LD? It is tiny. It's not even a gene.
- We need genomes for specific SNPs to track the right allele. This is the real problem.
- More general question: How are you going to use the genome sequence? There are a lot of next steps. They maybe don't all need to be done. They won't all be done with identical priority. If you can give us an idea of what things are of most value, that would accelerate us.
- If you have a good assembly for an initial background to align transcriptome sequences on. As long as a whole genome sequence can separate out the two, we are in good shape.
- If you have to hit every gene
- The other thing is mapping. Every map tells you which homologue you are on and you can separate out the physical
- If you do that, you want a big population
- You don't need a big population if you just want to know what homologue they are on.
- A1 A2 at 0.5% is nearly impossible to resolve. It is difficult to resolve with 100 bp reads (?).
- Christian Tobias: use Pac Bio sequencers
- Shouldn't they just collapse anyway? As long as A and B are distinct….
- A1 vs A2 is a hap map in a human genome sense.
- We want them to collapse.
- Jeremy: The assumption is for every genome that exists, there isn't a nicely partitioned ___. There will be regions that will not collapse.
- We don't need full genome coverage, not every gene. It just needs to be across all 18 linkage groups.
- Given the distinction between A and B….at some point get big enough pieces megabase size….600 megabase pieces. Can you map 600 pieces between the A and B genome?
- A and B are distinct.
- But I can't tell the difference.
- We don't care about that; we just want a marker for each.
- Jeremy: I don't know which is assigned to each.
- A GLBRC memeber: there are really 18 linkage groups. Forget homologues.

- .*missed some discussion here*
- It might even help for the assembly.
- Jeff: We need to resolve this. If it is so expensive to do it, then it can't be done. There are issues of feasibility vs. need. We can find out about new uses.
- GLBRC member: I mentioned this at lunch on Saturday, and it got shut down. The whole idea of running Christian's progenys and mapping them onto linkage groups. I hear Jeremy saying this isn't a big deal.
- The mapping probably isn't a big deal if ----
- Jeff: doing this right on each segregated individuals would have to be done to 100 X depth on 200 individuals. This is prohibitive. If you do it at less depth, it won't work so well.
- How difficult is it to link the BAC end sequences? The genes look compact. What is the gene space? It wouldn't be an enormous amount of resources. There are 200,000-300,000 pairs. The limiting factor is this won't be enough data to do it. We are trying to stick nearby reads together. We didn't. We got 400K. We have libraries and if this were a priority……come up with a scheme for markers localized in gene space that would help you organize As
- A5: We're doing transcriptomics. Is there a way to dovetail so this is useful? Will it be done in a way that will help?
- Jeremy: the other sets are tied up in the DOE BRCs. From the question of building a localized gene assembly, it would be nice to have enough to ___ the more the merrier.
- Should we send it to you? We have close to 60 million reads. He's understanding that…you need to have both. For some of these genes, you can find clear differences. It could be that you are able to find for certain things you get a and b, but not A1 and A10.
- A GLBRC member said he could do that for one gene but needs to do it on more.
- Send Jeremy, Dan or Kerrie an email.
- Can still do the bar coding and 6-cutter with the same number of samples. He has no experience on how to get insight into it.
- What is the perfect genotype for your sequencing? It would be an inbred line.
- Alamo 13 is what all the sequencing has been done in and this is where all the resources are. Are you asking to generate all of these resources in the inbred line? It would have been nice 2 years ago.
- Thinking about other populations and genotypes, this is up to each individual to worry about. As the technology gets cheaper, we can do it ourselves. Alamo is close enough and is closer than *P. hallei* or foxtail millet. It is very far away from what he is trying to breed.
- Jeff: we won't start a new source of sequencing in switchgrass in the next few weeks. If we'd known there were 75 lines available…BESC group didn't know inbreds were available. We will have 800 switchgrass lines resequenced 5 years from now, but can't do

it now. To start anew would be slower to get a useful product in the next year or two. This isn't off the table long term, however.