Notes from Switchgrass Genomics Meeting

Crystal City, VA April 10, 2011

Cathy Ronning gave an overview of the switchgrass genome community organizing efforts. The PAG meeting was organized to see what the community needs from the JGI effort and the DOE wanted to keep the lines of communication between/among the community and JGI.

Christian Tobias: Gave a recap of the PAG switchgrass meetings.

Reported on the meeting at PAG. It was a small setting with program managers present—there were observers from USDA, National Program Staff along with people invited as having interests in switchgrass including the BRCs.

Prior to the meeting they had to answer a set of charge questions---such as who you are, what are you doing, what does the switchgrass community need? All hit on the same theme---that of a need for a website for information clearinghouse, place to communicate on the web that would be easier for people to stay in touch.

People expressed a need for a genome sequence--realizing that not having a fully annotated genome sequence would be helpful as a scaffold for marker assisted selection.

Discussed having a follow-on meeting, a desire to have that at PAG either before or at a site in San Diego away from the conference. Another possibility was having it at the Noble foundation.

However, nothing was decided and we need to keep the momentum going as we can not wait another year for a switchgrass meeting.

It did not appear that there was a clear set of action items from the meeting; there was plans for the JGI to sequence the mapping population.

What followed then was a presentation by each of the BRCs, then that of JGI. Note: these notes are abbreviated as a lot of data was presented by the BRCs and JGI. The intent is that the presentations will be available on the to-be-developed switchgrass genomics/genetics community website.

Mike Casler (GLBRC):

GLBRC working on genetic improvement of switchgrass; translational genomics; utilize genomic information from maize to design more efficient genetic improvement systems

-Development of DNA marker selection systems

-increasing biomass yield

-decreasing lignin and etherified ferulates

-selection for late flowering surviours

Near term goals:

-utilize maize and switchgrass sequence to design and test efficient DNA marker selection systems

-decrease lignin concentration and gain and understanding of the regulatory factors; identify genes responsibility for endogenous genetic variation in lignin

-develop late flowering switchgrass

-Focused on developing new varieties as efficiently as possible.

-Switchgrass mapping populations and association panel; SNP/SSR markers and marker-trait associations for improved varieties

-Gains have been made in yield in the last 10 years but there has been a leveling out of the gain; reaching the limit of the technology

-Also working on conversion efficiency; reduce lignification to increase microbial access to cell wall CHO; utilizing endogenous variation for lignin; would like to know what regulatory genes are responsible for the reductions in lignin

-Selecting for late flowering plants that do not produce seed in Wisconsin; similar selection protocols to take place in 2011 and 2012 for additional switchgrass populations.

-Also working on big blue stem; have an accession that has >50% more biomass

-Genomics needs for GLBRC

-ability to align sequences to a reference genome

-need unequivocal assignment of alleles to one of the two subgenomes

-need to separate true allelic variation from different alleles on different subgenomes

---

Pam Ronald (JBEI):

JBEI: using rice as a model to get insights into switchgrass; also looking at switchgrass directly

-Looking at BAC libraries of switchgrass; to establish long range connectivity among genomic sequences; obtain full length sequences of coding and regulatory regions; differentiate homeologs and paralogs; compare to other grass genomes

-Lowland cultivar Alamo, clone AP13; 100K clones; 16X haploid genome; HindIII and BstY1 library; low organellar contamination; qPCR based BAC library screening; pools, superpools, and filters are available for these libraries

-Goal at JBEI is to identify BACs with cell wall and stress related genes including non-RD kinases, ethylene response TFs, glycosyltransferases, glycosylhdrolases; Used rice specific primers and had a success rate of 10%; then took rice genes and got candidate switchgrass ESTs and then designed primers and fish out BACs from pools.

Got full length sequence of 134 BACs, did repeat masking, gene prediction, gene annotation, integrating gene models in phylomic DBs; were able to identify >300 cell wall/stress related genes); added these genes to the affy chip

-HudsonAlpha/JGI has generated 360k BAC end sequences

-In total, generated 7.2Mb for 47 randomly selected BACs

-Annotated repetitive sequence content in BACs (occupies 26% of the sequence)-most are Class I TEs

-Grass phylogenomic DB: Rice Phylogenomic DB added in Arabidopsis and now adding switchgrass (500 genes have been added); have a GBrowse for switchgrass

-Done collinearity with other grasses. Switchgrass most closely related to sorghum/maize, then rice, then *Brachypodium*.

Resources Needed:

-lines capable of self pollination

-pathosystem for *Botrytis* and *Puccinia*

-publicly available DB for access to newly generated switchgrass data

-Finished Whole Genome Switchgrass will enable:

-comprehensive and accurate estimation of cell wall and stress related genes in switchgrass

-functional analysis of genes of interest (cell wall, abiotic stress)

-whole genome expression profiling of transgenic switchgrass expression stress tolerance and altered lignin

-whole genome gene network

Establishment of a composite network that is larger and more accurate than individual datasets; RiceNet-probablistic functional gene network of rice-50M data points, 23 types of datasets from 5 different species; have one network involved in stress response-validation of 5 genes were validated through transgenic work.

Michael Udvardi (BESC):

Goal to improve ethanol yields from biomass due to lower recalcitrance; improved varieties

-Necessary genetic and genomic tools

-clone libraries

-sequence genes and DBs

-expression tools

-transgenic tools

-genetic markers, mapping populations and genetic maps

-characterized genetic diversity

-mapped traits of interest

(mutatgenized populations

(physical maps and genome sequence)

All but the last 2 were developed by BESC in the last 3 years but the last 2 need work.

Gene Centric research:

ESTs: Alamo AP13 and Summer VS16 ESTs

Assembled Transcripts/cDNA: 120K unigenes from Newbler and PAVE; 70% putative FL cDNA, Annotated, Reference for RNA-seq

Affymetrix chip with 120K probe sets

Gene Expression Atlas in progress

RNA-seq in parallel

SNPs AP13 vs VS16 on-going

<u>Need from genome sequence:</u>

-Complete set of genes for comprehensive studies

-Complete, HQ CDS, UTR, promoters

-Question-how many and length of contigs do we need to get the genes?

-Need sequence from multiple genotypes for marker development

-Fine mapping requires knowledge of precise location of markers on the genome sequence

-Complete genome sequence with gene order and orientation extremely valuable

-Alignment of switchgrass to foxtail millet is a good approximation but is error prone

-Need a high throughput marker system (no statement of which platform)

Jeremy Schmutz-Alpha Hudson-JGI

Goal-produce an accurate reference sequence for AP13

-Capture the gene space, order the gene space on chromosomes via mapping

Subgenomes differ by 2.5%; alleles differ by 0.5% (looking at 3'UTRs); looked at 24mers in Illumina data-have unshared regions of the subgenomes, have A/B genomes, and have shared regions of genome.

Have BAC end sequences, fosmid end sequences, 454 runs, Illumina data (older datasets); 221 BAC clones in progress

Did a Newbler assembly of 8.3X linear 454 data; N50= 2.9kb

How much is collapsed among the subgenomes? How much is separated? Based on Sanger ESTs, BES a good amount of the gene space is separated in the 454 assembly.

Sequencing BAC pools to separate the haplotypes. Then use these to sort out the Illumina WGS. Looks promising. Could make a pooled BAC map to localize subgenomes.

Could place 340 Mp of switchgrass assemblies to foxtail millet

Have alignment of the Newbler contigs to the foxtail millet genome, can see the 2 subgenomes most of the time, but not all of the time.

Challenge: from gene scale contigs to chromosome scale maps

-Can use clone paired end info

-Expect colinearity between subgenomes and foxtail and grasses but which copy goes with which subgenome? Need large scale linking and mapping to separate into subgenomes

-Have an AP13 x VS16 mapping population suitable for building a large binned map

-Have 40K contigs that contain genes that need to be localized

-Additional data, algorithms to improve assembly but even 1 Mb scaffolds would require several thousand markers

Possible mapping scheme:

-Identify informative SNPs in each contig

-Brute force resequencing strategy. Deeply shotgun VS16; align reads to AP13; then sequence the F1 offspring (2X depth); map loci to chromosomes by cosegregation;

-Another option: use diploid *Panicum hallii*; 454 and Illumina only assemblies in progress, highly fragmented, needs additional resources; same ploidy as foxtail millet; do not directly useful for separating subgenomes; it is highly inbred so it may not be useful for resolving subgenomes; shared an ancestor with *P. virgatum* several million years ago, it might be useful for the local assembly; will be more straightforward than AP13.

Other potential resources:

-Physical linkage: laser capture chromosomes (LCM), NGS BAC fingerprint map, 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 QTL and other regions of interest to the BRCs)

Shawn Kaeppler-Led a discussion on what we need to move the community forward

Next steps-we need concrete plans to move forward

Reasons to make and implement an action plan:

-funding agenices have invested in switchgrass sequencing and want to see maximum utilization

-funding agencies would like a process for switchgrass community voice

-strong desire for utilize information for switchgrass improvement

Next steps:

-Mobilize a process to elect a switchgrass executive committee

Potential process-goal is to elect motivated individuals to serve as inclusive, unbiased voice of community; appoint an ad hoc committee to identify potential community, request nominations and hold election

-6 members with staggered 3 year terms, term-length determined by number in initial election

-elected committee will draft bylaws, identify community needs, organize research meetings

This was supported at the PAG meeting. A group of 3 people could get this started: *Jeff Bennetzen, Christian Tobias, Pam Ronald*. The goal would be to have something going in the next month.

There is a mailing list that Cathy and Jeff have, we can add people from tonight's meeting; Robin can make a google group mailing list; we could advertise this through professional organizations such as GrainGenes; *Christian is in charge of the Executive Committee organizing.*

One job of the committee will be to see what the community needs

-Community website and listserve

Robin-can take care of the logistics of the website through a neutral URL

It will be a basic portal for information dissemination

These will go on the web. *Robin will ask the presenters to send a PDF of their presentations to put on the web.*

-Produce a white paper on switchgrass genetics and genomics and future needs

Document the current state of switchgrass genetics, genomics and potential uses; publish in refereed journal

Pam Ronald has thought about a basic paper on the resources (sequence based), *Mike Casler* could add the genetic resources; describe the current state of switchgrass and where we could go in the future. Could be a review and perspectives paper. Could be accompanied by a few primary research articles. Robin comments that while nice this is a hard reality to achieve, keep it limited to the white paper to make sure it happens this year. Target date: mid-June. Could do this by having people send in sections. Will need a single person to be the manuscript coordinator.

-Identify priorities and collaborations to improve currently available sequence

Shawn would like to have more selected good accessions with seed sources. It is not as easy to do as one would think. It grows, drops seed in the pots, if you are not careful, you are not sure what you are growing.

Need standard protocol for confirming genotype of accessions, e.g. AP13.

Cryopreservation has been done in other species; someone at the Noble is starting a project.

Sequencing:

How many progeny are needed to separate out the linkage groups? 200?

Jeff states that you should plan on having mapping populations of n=1000 in preparation of cheap genomes; Christian disagrees as with just 2 parents you don't have enough capability to identify QTL;

Should we make a nested association population? It would not be that much work to get the seed, it is the placement of them in the field that is a lot of work.

Could have a web access meeting in a few months.