Minutes
Peanut Genome Consortium Meeting, Washington DC
12/9/13

PARTICIPANTS

Peggy Ozias-Akins
Baozhu Guo
Corley Holbrook
Richard F. Wilson
Charles Chen
Howard Valentine
Tom Stalker
Noelle Barkley
George Birdsong
Darlene Cowart
Tom Isleib
Mark Kline
W. Pierce
Lutz Froenicke
Sven Patrick
Dan Ward
Jiujiang Yu
Amme Deloranio
Michael Davis
Soraya Bertioli
David Bertioli
Steven Cannon
Roy Bechtold
John Takash
Jim Elder
Stan Fletcher
David Hoisington
Andrew Farmer
Brian Scheffler
Dan O'Conner
Rich Michelmore
Mark Burow
Scott Jackson

Jackson convened the meeting and welcomed those attending. The agenda was approved. Minutes from 7/08/13 were approved. These documents are posted on *peanutbioscience.com.*

REPORTS TO STAKEHOLDERS

Genome Sequencing & Assembly

FROENICKE reported diploid sequencing is completed and efforts are directed toward genome assembly. Assemblies of diploid A- and B-genomes are very good quality. In each case, scaffolds cover the whole genome with less than 1 million contigs. *A. duranensis* genome size is 1.2 Gb, N50 for the A-genome is 1.2 million bp; *A. ipaensis* genome size is 1.7 Gb, N50 for the B-genome is 5 million bp. The BGI assembly was validated with sequence data, mis-joins have been identified and removed. The assembly has been split into the component sub-genomes. Annotation is in process and will facilitate development of models for pseudomolecules by end of January. Scaffolds also cover the whole *A. hypogaea* genome assembly with over 4 million contigs. Genome size is 2.9 Gb, but N50 for the cultivated tetraploid genome is less than 70 thousand bp. In lieu of longer reads which would facilitate tetraploid assembly, the diploid assemblies will be used to help split out the component subgenomes.

High Density Diploid Maps

FROENICKE performed genome-wide association studies (GWAS) on diploid mapping populations provided by BERTIOLIs'. Scaffolds covering 90% of the genomes were mapped to linkage groups. Over 3 million high-quality SNPs were found in each diploid species (*A. duranensis, A. magna, A. ipaensis*). BERTIOLIs' refined the genetic maps for the A-, B- and AB genomes for ordering. Genotyping on current consensus maps have been constructed with SSR and Illumina derived SNP markers (ca. 400 SNP for A-genome; 500 SNP for B-genome). Resolution is good for diploids, but more markers (larger RIL populations) are needed. BUROW offered 1500 F2 lines from (*A. cardenasii x A. duranensis*) to help enhance map density. SCHEFFLER raised discussion of segregation distortion (favored inheritance of alleles from one parent) in populations among different species. The nature of the apparent distortion was attributed to marker segregation rather than linkage disequilibrium (LD). Effects appear to be stronger in AxB populations, least strong in AxA populations; stronger in cultivated AABB x diploid, least strong in synthetic x cultivated tetraploid populations. This problem might be overcome with greater recombination. Still, maps in hand plus synteny data are sufficient to proceed with the first publication.

High Density Tetraploid Maps

FROENICKE initiated GWAS with: ICRISAT reference set resources (with multiple season phenotypic data for 51 traits), the Chinese mini core diversity panels, and the USDA mini core collection. Genotype x Sequencing (GBS) which provides reduced representation libraries covering the whole genome is an efficient means to genotype these resources. However, results from a 96 sample multiplex of the ICRISAT diversity panel and a 48 sample multiplex of the Chinese diversity panel (paired end; 300-500 bp size selection) were disappointing. Several different protocols were evaluated (Tassel/Uneak, STACKS, RAD-seq), RAD-seq gave best performance yielding ca.12.000 candidate SNPs & PAVs and ca. 1300 SNPs after removal of A/B genome variants. More (about 15,000 SNP) are needed. SNP discovery in *A. ipaensis* and *A. duranensis* reference assemblies filtered against Tifrunner SNP produced

only 2,300 SNPs (low compared to 82,000 differentially filtered SNPs in lettuce). A 5-fold increase in number of reads (10 x 10 million) may help overcome the low nucleotide variation in the tetraploid genome. However, a targeted enrichment (exome capture) approach has proven to be cost effective in difficult genomes such as potato and loblolly pine. FROENICKE will use Moleculo™ long reads (tetraploid), transcriptome assemblies, RNA-seq and EST data to develop bait libraries for exome capture to survey sequences in diversity panel, reference and mini core collections. (it's a cliff hanger).

Gene Expression Studies

Ozias-Akins reported on the status of gene expression analyses in 24 libraries (8 tissues in triplicate) of the Tifrunner transcriptome. All vegetative tissue libraries have been sequenced. 5 of these libraries have extreme depth (100 to 200 million reads each) which will facilitate high quality differential gene expression analysis. MiSeq (2 x 300) also produced 500 bp merged reads from tetraploid transcriptome libraries. Data appears sufficient to proceed with assembly. **CANNON will convene a teleconference with Ozias-Akins, Scheffler, Jackson, Burow for coordination.** An assembly pipeline has been developed and evaluated. A- and B-transcriptomes split out from the tetraploid evenly with no gross overexpression. The B-transcriptome aligned closely with the B-genome. Incomplete transcripts (8700 *A duranensis*-derived, 7700 *A. ipaensis*-derived) were filtered out. Remaining transcripts were screened against non-coding regions. Alignment was noted for 3800 B-transcripts and 2600 A-transcripts that were associated with 10 elements. Expression of this set of transcripts was often clustered in specific transcriptome libraries. Filtered assemblies were produced via genome guided homolog specific (GGH), CLCBio™, SoapDenovo™ or Trinity™ software. GGH assemblies gave the best performance, scoring about 80% of Gmax protein hits.

Phenotyping

Holbrook updated group on activities with 16 RIL populations. 8 populations were increased in 2013, the remainder in 2014. Four populations were phenotyped (C1799, C1801, C1798 and C1803). Birdsong Inc graciously stored all seed for these populations in -18C freezers. **A seed-inventory will be updated by Holbrook and posted on www.PeanutBioscience.com/**. Examples of results were reported for:

- C1799: a F6:8 RIL population of 286 lines from Tifrunner x NC3033. The population was genotyped with 121 polymorphic SSR primer sets. Transgressive segregation was observed for early leaf spot and TSWV resistance/tolerance. Phenotypic data was taken on these RIL. Significant differences were found for: pod weight, kernel weight, pod volume, pod filling and pod density, maturity, seed color and transpiration (indicator of drought tolerance).
- C1799 RILs also were genotyped for resistance to white mold (S. rolfsii). Lines were genotyped by Ozias- Akins with 380 polymorphic SSR markers newly developed from transcript sequences, and 2000 publically available SSRs with high polymorphism.
- C1801 (Florida 07 x SPT-06-06) was genotyped with 79 polymorphic SSR. Segregates were evident for early leaf spot resistance. Initial QTL maps were developed for ELS and TSWV.

Remaining populations will be phenotyped and mapped in 2014 for CBR, LLS, TSWV, WM & ELS.

Bioinformatic Resources

CANNON reported on progress and goals for PeanutBase.org. The website will combine the strengths of SoyBase and LIS to provide: 1. convenient access to data sets (maps, transcriptome, SNPs, RNA-seq, genome assembly, annotations, etc), 2. links to external information, 3. biological information for context, integrated maps, 5. QTL and trait information, 6. assistance with genome assembly and annotation (gene order between peanut and soybean; gene predictions from transcripts), and 8. Training, outreach, and coordination (development of a GCP Integrated Breeding Platform). CANNON also has initiated ordering of pseudo molecule (chromosome) assemblies of the diploid genomes. Early observations predict 10 chromosomes with about 250 scaffolds each (2726 scaffolds placed on *A. ipaensis* linkage groups; 2496 scaffolds place on *A. duranensis* linkage groups). Software and filtering protocols developed at Ames, IA will be used to make placements in large pericentromeric regions. Data could be available in 1-2 months and will be reviewed by Michelmore. Synteny plots between diploid with soybean and Phaseolus linkage groups will be very useful in aligning scaffolds and developing finished models of pseudo molecules.

Action Items:

- Spiral Genetics – collaboration in progress, waiting on scaffold assignments
- Report of the ad hoc publication committee: FROENICKE updated the publication action plan. Several areas will wait for pseudomolecule models before finalizing sections. Decisions on publication logistics will be discussed at PAG XXII in January 2014.
- VALENTINE presented draft language for sharing PCP generated data posted on or accessed through PeanutBase.org. Revised language will be sent to Jack Okamura at USDA ARS ONP for information.

OTHER BUSINESS:

- AAGB-2014 will be held at Savannah GA in November 2014.
- PGC and PGP ad hoc publications committees will meet MONDAY January 13, 2014 from 10:30 am to 2:00 pm EPT in the Dover room at the Town & Country Hotel during PAG-XXII in San Diego. AV and telecommunication facilities will be provided.

NEXT MEETING:

MONDAY January 13, 2014 from 10:30 am to 2:00 pm EPT in the Dover room at the Town & Country Hotel during PAG-XXII in San Diego

ADJOURNED