

**Report**  
**Peanut Genome Consortium annual meeting @ PAG XXI**  
**1/14/2013**

**PARTICIPANTS**

Victor Nwosu	Andrew Farmer	Nong Chen
Lutz Froenicke	Robert Henry	Rajeev Varshney
Brian Scheffler	Sachiko Isobe	Richard Michelmore
Peggy Ozias-Akins	Peter Bretting	Dan O'Connor
Baozhu Guo	Jack Okamura	Soraya Bertoli
Corley Holbrook	Paxton Payton	David Bertoli
Mark Burow	Beth Grabau	Liu Xin
Richard F. Wilson	Ran Hovav	Sarwar Azam
Scott Jackson	Anna Claudia Guerra Araujo	Rao Kottapalli
Steven Cannon	Xun Xe	

Jackson convened the meeting at 10:45 am PST, welcomed members attending or via teleconference and guests. A quorum was established. The agenda was approved. Minutes from 10/24/12 and 11/1/12 were approved. These documents are posted on [www.peanutbioscience.com](http://www.peanutbioscience.com).

**Reports:**

**CY2013 RESEARCH PROPOSALS:** Wilson reported on the review process and the status of pending research proposals for CY2013. A call for proposals was issued on 11/12/12. Since funding was provided by the US peanut industry, a U.S. PI was a primary requirement for each proposal. A peer panel chaired by Dr. Randy Shoemaker (USDA, ARS, Ames IA) met 12/19/12 in Atlanta GA or via teleconference to evaluate and rank 20 proposals. Four well known scientists served as peers. Representatives of producer, sheller, and manufacturer segments of the U.S. peanut industry plus Valentine and Wilson observed the review. Proposals were evaluated for relevance to the PGP Action Plan, merit, scientific ability, and potential impact. The CY2013 budget was \$1,790,000. \$1,110,000 was set aside to cover contractual payments to BGI. \$680,000 was available for funding accepted proposals. 11 proposals were recommended for funding. The recommendations were approved by The Peanut Foundation Board on 12/21/12. The American Peanut Council Board approved the recommendations for funding on 12/29/12. Letters of notification will be sent to all applicants pending review of the progress report from BGI at the PGC meeting during PAG XXI on 1/14/13.

**PROGRESS on GENOME SEQUENCING & ASSEMBLY.** Liu Xin reported for BGI via teleconference. Xun Xu and Nong Chen attended. Performance measures for Phase I are:

1. Sequencing and assembly of the tetraploid (Tifrunner) genome (60×)
2. Resequencing three other tetraploid stkains (GT-C 20, SunOleic 97R, NC94022) (10×)
3. Resequencing of two RIL populations (3×)
4. Sequencing and assembly of the diploid A (*A. duranensis*) genome (100×)
5. Sequencing and assembly of the diploid B (*A. ipaensis*) genome (100×)

**Current Progress**

1. *Sequencing and assembly of the reference tetraploid genome (Tifrunner):* Sequencing libraries with insert sizes ranging from 250 bp to 40 kb were constructed, and sequenced to more than 110×. The genome assembly was finished. Kmer analyses suggested the genome size to be about 2.65 Gb.
2. *Resequencing of three other tetraploids (GT-C20, SunOleic 97R, and NC94066):* One 500 bp insert library was constructed for each line and sequenced to about 12X coverage.
3. *Resequencing of two RIL populations:* One 500bp insert library was constructed for each of two RIL populations. Individual RILs were sequenced to an average sequencing depth of 3.6×. More RILs were requested to improve resolution. **Guo provided BGI with additional RILs.** Next steps involve construction of the genetic map for the tetraploid genome. There are now a total of 137 RILs for the S-population and 113 RILs for the T-population for resequencing.
4. *Sequencing and assembly of A and B genomes:* Insert libraries (250bp, 500bp, 2 kb, 5 kb) have been constructed and should be sequenced by the end of January 2013. More seed is needed to construct 10kb, 20kb, and 40kb insert libraries for both diploids. Arrangements are being made to ship the requested seed.

Due to the apparent complexity of the tetraploid genome assembly, it may be necessary to use a BAC x BAC approach in Phase II. However, BGI has used fosmid sequencing in other species which yielded N50 of 300bp. Improvements in BGI SOAP denovo software may also help close microgaps in the genome assembly. Application of MolecuLo™ and other new technologies was discussed (see below). BGI submitted a formal written report to the PGC, and provided a pdf file with a summary report. The pdf will be posted on [www.peanutbioscience.com](http://www.peanutbioscience.com). **BGI agreed to share data with PGC members ASAP.**

BGI plans for completion of Phase I

1. *Sequencing and assembly of the reference tetraploid genome:* a) Finish gap filling; obtain the final genome assembly of WGS before Feb 1, 2013; b) Use RIL populations to construct a genetic map (43 more RILs for SunOleic 97R x NC94022 and 118 RILs from Tifrunner x GT-C20 were received and will be sequenced soon)
2. *Resequencing of three other tetraploid GT-C20, SunOleic 97R, and NC94022:* Map the final tetraploid assembly and identify variations of these three lines before Feb 14, 2013.
3. *Resequencing of two RIL populations:* After obtaining the final tetraploid assembly, map the reads of each individual to identify variations and construct the genetic map.
4. *Sequencing and assembly of A genome:* a) Finish sequencing the short insert size libraries (250 bp, 500 bp, 2 kb and 5 kb) which have already been constructed, before 20 Feb, 2013. b) Finish construction and sequencing of 10 kb, 20 kb and 40 kb libraries after receiving the DNA. c) Finish the genome assembly and later analysis.
5. *Sequencing and assembly of B genome:* a) Finish sequencing the short insert size libraries (250 bp, 500 bp, 2 kb and 5 kb) which have already been constructed, before 20 Feb, 2013. b) Finish construction and sequencing of 10 kb, 20 kb and 40 kb libraries after receiving the DNA. c) Finish the genome assembly and later analysis.

Michelmore suggested a two month delay to allow inspection of MolecuLo™ data before committing to a course of action for Phase II. BGI agreed. If this technology works on the 40kb insert libraries it should help bring the assembly together, and based on analysis of GC skew should help determine how much genetic information would be gained compared to a BAC approach. High polymorphism in A- and B-genome RILs would also help bring the assembly together. D. Bertioli reported the development of three sets of RILs from A- B-, and AB-genome crosses. Those sent to Michelmore were F5 to F6. New populations are mostly F4. It was agreed that the current set of 100 lines should be supplemented with additional RILs from the synthetic amphidiploid populations to help increase the quality of the reference genome sequence. **Bertioli agreed to advance the synthetic tetraploid RILs and coordinate actions with Michelmore.**

**Action Item: Jackson will organize a teleconference in three months, possibly in April, to discuss results of MolecuLo™ and other new technologies. The outcome will help determine the course of action taken in Phase II.**

ULTRA DENSE MAP OF ALLELIC VARIATION: Michelmore and Froenicke reported progress toward gene space characterization of the following genetic resources:

- 1) Genotype by Illumina™ sequencing of RIL populations from progenitor species by Bertioli.
  - A genome: *A. duranensis*, *A. stenosperma*
  - B genome: *A. ipaensis*, *A. magna*
  - Synthetic allotetraploid: *A. hypogaea* (Runner) x (*A. ipaensis* x *A. duranensis*)

After trimming, the A and B genome gene space size was estimated at 900Mb. N50 for gene space assemblies was 13K (A-genome), 7.5K (B-genome; where >4K is good. GC% for both was about 35%. QC was very high. Gene space annotation was supported by use of a Maker genome annotation pipeline for data from an *A. duranensis* transcriptome, *A. hypogaea* ESTs, plant protein database, Repeat-Masker data base on eudicotyledons, and peanut retrotransposons. Analyses predicted about 35K genes. A-genome and B-genome transcriptomes also are being assembled. An advanced genome mapping pipeline is being tested with new data. **Results appear**

**good enough to proceed ASAP with publication of ultra dense genetic maps of the A- and B-genomes. Michelmore will share these data with BGI.**

Michelmore and Froenicke also are testing MolecuLo™ technology on diploid genomes. This technology is a highly regarded new process for aligning 10kb sheared fragments in near perfect order. It consists of proprietary methods for single molecule amplification, highly-parallel library prep, and algorithms for long fragment reconstruction. In practice, a client prepares and sends DNA to MolecuLo™ for library construction, the libraries are sequenced by the client, MolecuLo™ assembles data sets. If this technology works for peanut, it would substantially reduce the need for an elaborate BAC x BAC approach in Phase II. **Michelmore will share these data with BGI.** In addition BGI is evaluating methods developed by Complete Genomics™ for assembling longer reads. This may also help reduce need for BAC x BAC derived assemblies.

- 2) SNP identification and LD analysis will be conducted in timely order for: a) 192 phenotyped tetraploid peanut RILs segregating for drought tolerance and foliar diseases (ICRISAT); b) 325 phenotyped/genotyped diversity panel accessions from ICRISAT (R. Varshney); c) 298 Chinese Mini-Core collection accessions from Oil Crops Research Institute, Chinese Acad. Agric Sci., Wuhan, China; and d) 112 USDA Mini-Core accessions from the USDA Peanut Germplasm Collection at Griffin GA.

**DATA SHARING SYSTEMS:** BGI discussed the need to agree on means and methods for data release. BGI agreed to release tetraploid assembly data ASAP. Michelmore agreed to share diploid data with BGI. All agreed that raw data should be protected, as stated in the PGC Policies & Procedures manual. Cannon suggested the possible utility of iRODS™, an Integrated Rule-Oriented Data System for transfer of large datasets including the full genome and resequencing data. The iRODS system was developed to support of data grids, digital libraries, persistent archives, and real-time data systems. Chen stated that BGI has not used iRODS. Michelmore suggested that GBrowse might be adequate for diploid data sets, but not for tetraploid data. Cannon also suggested iPlant, an NSF funded project that supports the Integrated Breeders Platform (which includes peanut). However, the project is up for renewal with a possible decision by summer 2013. **Cannon agreed to lead a study group on methods for handling large data sets.**

**GENE EXPRESSION.** Scheffler reported with Ozias-Akins on plans to conduct gene expression analyses of the Tifrunner transcriptome. Twenty-four libraries (triplicate libraries from 8 tissues) have been sequenced with 8 to 10 M reads per sample of high quality total RNA. Experiments are being conducted to determine an optimal read depth with paired ends. Cannon noted that several transcriptome data sets were being developed. **Cannon, Ozias-Akins and Froenicke agreed to coordinate handling and sharing of transcriptome data. Michelmore agreed to send Scheffler A- and B-genome transcriptome data for annotation.**

**PHENOTYPING:** Holbrook reported on recommendations from a webinar on 11/20/12 organized by Trushar Shah and Xavier Delaney (ICRISAT). A recording of the teleconference is available at:

<https://dl.dropbox.com/u/102827068/2012-11-20%2008.34%20Introduction%20of%20the%20IBP%20to%20the%20Peanuts%20team.wmv>

The group recognized the existence of a wide range of ontologies for crop and animal species, including: GO, NCBI, FMA, CARO, PaTO, PRO, SO, RNAO, ChEBI, CO. All agreed a common terminology was needed with a controlled vocabulary. The best option appears to be the Integrated Breeding Platform.

<https://www.integratedbreeding.net/>

The Integrated Breeding Platform was built by the Texas Advanced Computing Center and is hosted by the iPlant Collaborative. Information available on the IBP includes a Trait Directory and Crop Database collected in the course of routine breeding activities. This information is compiled and published by International Research Centers of the CGIAR. The GCP has proposed a general strategy for managing this type of crop information. Peanut is included in the IBP. However, the IBP is written in visual Basic program language which will not interface with existing peanut databases. **Cannon agreed to work with**

**Shah (ICRISAT) to rewrite the IBP in Squel. Holbrook and Ozias-Akins agreed to search the literature for improved peanut trait & ontology descriptions for peanut.**

EXPERIMENTAL SEQUENCING TECHNOLOGIES. See discussion of MolecuLo™ technology.

AAGB-2013. Nwosu and Guo reported that the technical and social programs for the meeting are being developed. Deadlines for registration, call for papers and housing are available via [www.peanutbioscience.com](http://www.peanutbioscience.com). The meeting will be held at the Crown Plaza hotel in Zhengzhou PRC on June 17 to 21, 2013. About 150 attendees are expected with about 21 U.S. and 6 from ICRISAT. Foreign travel requests should be initiated ASAP. Guo will work with the General Chair, Dr. Zhang to make Red Letters for visas available upon request.

PARALLEL WORKGROUPS IN CHINA. A written progress report was received from the Chinese Peanut Genome Project launched jointly by Shandong Sheng-Feng Seeds Ltd, Guangdong Academy of Ag. Sci., Shandong Academy of Ag. Sci. and MacroGen. The Chinese Genome Project members include chief scientific officer Li Hong-Jie from Sheng-Feng Seeds, Drs. Liang Xuanqiang and Chen Xiaoping from Guangdong Academy of Ag. Sci., Director Yu Shanlin from Shandong Academy of Ag. Sci., Dr. Rajeev Kumar Varshney from Genome Center of ICRISAT, CIO Hwanseok Rhee of MacroGen, ChangHoon Kim in charge of MacroGen-de novo Bioinformatics, and Jong-So Kim of MacroGen NGS Chief Officer. The full report will be posted on [www.peanutbioscience.com](http://www.peanutbioscience.com). Briefly the group expects to complete the sequence of an A-genome progenitor in January 2013, and publish the results in June 2013. Jackson stated the opportunity still exists for cooperation between groups, and that the PGP will proceed according to established PGC guidelines.

RESEARCH ACCOMPLISHMENTS DATABASE. Wilson stated the need to keep stakeholders informed of research progress toward the expectations of the Peanut Genomic Research Action Plan. Although reporting requirements for funded PGP projects are spelled out in Policies & Procedures, communication with stakeholders would likely be on a more frequent basis. Wilson suggested it would be helpful to develop a research accomplishments database that was relevant to the IPGI Strategic Plan and the PGP Action Plan. Such an asset would be useful in producing accomplishment reports as well as program assessment. **Wilson agreed to lead an ad hoc workgroup to explore this option..**

INTERNATIONAL COLLABORATIONS. Beth Grabau reported that VPI will host a USDA FAS sponsored workshop for foreign bioregulators to help harmonize U.S. and foreign interests in regulatory approvals for genetically engineered crops. Robert Henry reported that he has acquired leadership of the public peanut research program in Australia. He stated that his program will observe the covenants of the PGC Policies & Procedures regarding data sharing and publication of results. Anna Claudia Guerra Araujo expressed the same convictions on behalf of the South American delegation to PGC. Nwosu raised a point of order regarding unauthorized communications from the PGC. A policy was adopted to require discussion and sanction by the PGC chairs of communications that could impact PGC operations, agreements or collaborations. All attendees were reminded that Drs. Varshney, Guo and Xun would speak during the BGI symposium at PAG XXI.

POLICIES & PROCEDURES. Version 5.9 of the P&P is now posted on [www.peanutbioscience.com](http://www.peanutbioscience.com)

OTHER BUSINESS. None

NEXT MEETINGS.

March 13, 2013, American Peanut Council Spring meeting, Atlanta GA

Mid April, 2013, PGC Teleconference, Next Steps for Phase II

9-11 July 2013, American Peanut Research Education Society, Young Harris, GA

17-21 June 13, AAGB-2013, Zhengzhou, PRC

ADJOURNED