

**Report**  
**Peanut Genome Consortium**  
**Executive Committee, 1/16/12**

**PARTICIPANTS**

|                    |                 |                  |               |
|--------------------|-----------------|------------------|---------------|
| Victor Nwosu       | Scott Jackson   | Mark Burow       | Arvind Bharti |
| Lutz Froenicke     | Corley Holbrook | Charles Chen     | Xin Liu       |
| Brian Scheffler    | David Bertioli  | Pat Donahue      | Wenbin Chen   |
| Peggy Ozias-Akins  | Soraya Bertioli | Jack Okamura     | Jianxin Ma    |
| Baozhu Guo         | Graeme Wright   | Howard Valentine | Andy Paterson |
| Rajeev Varshney    | Callum Bell     | Xun Xu           | Roy Scott     |
| Richard F. Wilson  | Howard Shapiro  | Steven Cannon    | Andrew Farmer |
| Richard Michelmore | Ray Schnell     | John Crow        | Jianping Wang |

Jackson convened the meeting and teleconference. A quorum was established. The agenda was approved. Minutes from 12/06/11 were approved.

**Discussion Items:**

**The BGI Proposal**

In the BGI response to the 10/15/11 proposal for the Arachis Genome Project, Xun Xu indicated that the cost of the project would be reduced from \$4.5M to \$2.8 M by pooling several BACs rather than individual BAC indexing while maintaining promised quality. Xu confirmed the total cost at \$2.8 M. The EC voted to accept this offer and requested the amendment in a revised proposal from BGI. It was proposed to conduct the project in two Phases instead of four steps (outlined on page 16 of the BGI proposal). The cost of Phase I would be \$840,000 USD; Phase II should be \$1,960,000 USD. Protocol for each Phase is outlined below under discussion of the MOU. **Xu agreed to revise the Timeline, Gantt Chart and Budget pages of the BGI proposal to reflect these changes; and provide Valentine with a final copy.**

DNA samples should be delivered to BGI by 1 April 2012. Phase I would commence upon receipt of acceptable DNA; Xu estimated completion of Phase 1 in about 3 months.

**Memorandum of Understanding between the Peanut Foundation and Beijing Genomics Institute**

Valentine presented a draft MOU for discussion. Advance copies were not available, each section was read to Xun Xu per phone. The following items were discussed.

- **Data release statements:** All agreed that data generated and held by BGI, NCGR or other designees would be made publically available without restriction. Prior to data deposition in federal bioinformatic resources such as NCBI, users are expected to refrain from patenting or copy writing information in the BGI, NCGR or private databases. **Crow and Xu agreed to discuss and draft language on versioning of pending data releases. USDA Representatives agreed with this action. The EC voted to amend the language in paragraph 7 of the MOU to reflect agreements on data release versioning, when available.**
- **Phase I: WGS sequencing of the diploids and tetraploid.** The project will begin with construction of a range of paired end and mate pair Illumina libraries (including 20 and 40 kbp mate-pair libraries). One line of each diploid and the tetraploid species will be sequenced to 100x; a second parental line, GT-C 20 of each species will be sequenced to 10x and each RIL will be sequenced to 3x. Assemblies of the WGS sequence, integration with genetic information and annotation will be included. Limited BAC sequencing will be included for the tetraploid to test the feasibility of the BAC-based approach for generating a high reference sequence of the tetraploid genome in phase II. This approach affords flexibility to move to newer technologies that ensure the generation of a high resolution reference sequence at the lowest cost. Xu agreed that Phase 1 provided a means to assess (phase I evaluation metrics) the utility of short reads or need for more BAC sequencing before moving into Phase II. In any event, Xu pleaded that BGI would use the best technology available.
- **Phase I Evaluation metrics:** For all three species (*A. duranensis*, *A. ipaensis*, Tifrunner/GT-C20): contig N50 >20kb, scaffold N50 > 300 kb. More than 75% of the genome should be captured in scaffolds greater than 2 kb. For the tetraploid, more than 80% of these scaffolds should be placed in chromosomal linkage groups and be genetically coherent. At least 95% of the transcriptome

should be represented in each assembly. At least 20 BAC clones should be sequenced for each species and shown to be at least 98% identical with the genome assemblies.

- Phase II: Based on the experience gained in phase I and consistent with the technologies prevailing at the time, the appropriate strategy will be implemented to generate a high quality reference genome sequence of cultivated peanut after approval and authorization from the Peanut Genome Consortium and the Peanut Foundation. The current strategy involves generating a 6x BAC library for cultivated tetraploid peanut and sequencing pools of ~5 BAC clones.
- Phase II Evaluation metrics: For the cultivated peanut: contig N50 >100kb, scaffold N50 > 2000 kb. Total number of scaffolds greater than 2 kb representing more than 95% of the genome should be less than 1,000. The single base error rate should be <1/100,000. More than 95% of these scaffolds should be placed in chromosomal linkage groups and be genetically coherent. At least 99% of the transcriptome should be represented in the assembly. The raw data sufficient to allow assembly using other algorithms will be provided in the form of FASTA files with base quality scores or FASTQ (before and after trimming), the paired-end and mate-pair information, data on the depth of base pair coverage at all positions within the pseudo-molecules, scaffolds and non-scaffolded contigs, and the sequences of the scaffolds and contig sequences.
- The following new payment schedule proposed by PGC was acceptable to Xu.
  - 30% prepayment to initiate the project
  - 20% payment shall be paid within 30 days after the sequencing is finished;
  - 20% payment shall be paid within 30 days after the assembly is finished;
  - 30% payment (the balance) shall be paid after evaluation and the genome data and analysis is uploaded to a bioinformatic resource for public release.
  - Payments shall be made to BGI US (INSERT ADDRESS).
- Good faith efforts will be made by all parties to fully execute the MOU by 29 February 2012.

A motion to approve the MOU with modifications presented herein was passed without opposition. **Xu agreed to provide a final revised BGI proposal including new payment schedule for inclusion as an addendum to the MOU. Shapiro and Valentine will request review of the final MOU by respective legal council.**

#### PGP Component Updates

*Component 2:* Micheltore and Froenicke have a preliminary high density map of 10 B-genome linkage groups. Algorithms are being tested on known data sets from lettuce; good progress reported. Order of work will proceed as: B-genome, A-genome, AB-genome (synthetic tetraploid), diversity panel from ICRISAT (no sequencing yet, still optimizing libraries; depth of sequencing will depend on degree of diversity), and DNA from the Chinese germplasm collection (receipt still anticipated). Scheffler will work on DNA from the USDA germplasm collection. Froenicke was scheduled to make the first Peanut Genome Initiative presentation at PAG. If available his presentation will be archived on PeanutBioscience.com.

*Component 3:* Ozias-Akins reported preliminary work for RNA-seq of Tifrunner.

*Component 4:* This component evaluates new sequencing technologies. One approach under consideration is utilization of long Pac Bio reads with an average of over 3 kb and up to 14kb that is attainable with their new chemistry with error correction using shorter Illumina reads. All work is very preliminary. Other advances in genome sequencing technology such as digital PCR were presented by Steven Quake in plenary sessions during PAG-XX

*Component 5:* Holbrook reported the status of plans for developing descriptors for phenotypic traits. Lists have been received from Brazil (in Portuguese). Varshney will request the list from Hari at ICRISAT. These efforts will be coordinated with SoyBase and the Plant Ontology Network for descriptors of similar traits in other species. The EC voted to establish a formal Phenotyping Task Force. Initial members include: **Holbrook (lead), Varshney, Bertoli, Guo, Ozias-Akins, Stalker, Sanders.** Holbrook has authority to recruit additional members as needed.

*Component 6:* Steven Cannon, USDA,ARS at Ames IA described working relations and his role as coordinator of the USDA ARS contract for LIS with NCGR. Crow, Farmer, Bharti, Callvin represented and reported for NCGR. Cannon provided a prioritized outline of work assignments for deliverables from bioinformatic resources in the PGP Action Plan. Detail was provided on the type of information that

would be needed to make a Peanut dB component of LIS useful to breeders (similar to SoyBase).

**Wilson will amend the Action Plan accordingly.** Cannon described the workforce needed to implement the plan (at minimum 2 FTE for cs. \$100K). USDA ARS Office of National Programs representatives expressed strong support for LIS; Crow reiterated that NCGR will stay invested in LIS for the long-term. The Peanut Foundation pledged to seek support for LIS. **Crow agreed to draft a briefing paper for Valentine on why the support is needed, what the peanut industry would get, and how much it would cost.**

The EC accepted the nomination of Steven Cannon, John Crow and Arvid Bharti for membership. **Wilson will develop and distribute an electronic ballot for this election.**

International Collaboration Updates: Varshney reported that ICRISAT is contributing to Components 2 and 5. Highlights include:

- DNA from a reference set of 300 genotypes representing diversity of the groundnut germplasm collection and one reference mapping population for the tetraploid groundnut (TAG24 x ICGV86031) has been sent to Micheltore. Phenotyping data on this material is available for GWAS.
- A consensus map with >900 SSR loci (for which sequence data is available) was constructed. in collaboration with several international labs, and is available to help integrate BAC-based sequence maps and genome assembly.
- DNA markers for late leaf spot resistance have been released.

**All International colleagues and PGC members were encouraged to send Wilson relevant news items on a timely and routine basis. Wilson will elaborate a NEWSLETTER section in PeanutBioscience.com**

Policies & Procedures: Accord on a broad general operating procedure for data access remained elusive. Caution was expressed by industry members regarding protection of data generated by the PGP prior to deposition in a federal bioinformatic resource such as NCBI. However, public policy makers argued against requirements for user registration or the use of shrink-wrap licenses for data access. No vote was taken. The issue was tabled until the US Supreme Court rules on whether genes found in nature are patentable.

Update on US-China MOST: As noted in the 12/6/11 Minutes, USDA and MOST delegations met in Las Cruces NM to continue diplomatic dialog relative to the inclusion of PGP under the US-China MOU on collaborative research. More meetings will be held in China and the US. The next meeting between the MOST and USDA delegates will be held in China in February 2012. Zhang and Chinese members of PGP are pushing the issue directly with MOST.

Other Business: Ozias-Akins proposed organizing a half-day symposium on PGP and related activities during the 2012 APRES meeting in Raleigh NC. The EC approved this action in principle, and requested additional information when the program was further developed.

Grateful appreciation was expressed to MARS, Inc. for providing lunch and refreshments after the EC meeting.

Next Meeting:

Peanut Foundation project review, March 29, 2012 in Atlanta GA; pending teleconference with EC American Peanut Research & Education Society annual meeting, July 10-12, 2012 in Raleigh NC

Adjourned