Report
Peanut Genome Project
Ad hoc Committee Teleconference, 8/25/11

PARTICIPANTS

Lutz Froenicke	Peggy Ozias-Akins	David Bertioli (email)
Brian Scheffler	Richard F. Wilson	Scott Jackson

Wilson convened the teleconference to discuss items to optimize the BGI *Proposal for Arachis hypogaea Genome Project*. that was submitted by Xun Xu on 8/17/11. Chairperson Jackson led the discussion.

Discussion Items:

Genome Sequencing & Assembling Page 8, para 1 ….to the reference genome with 200 Recombinant Inbred Lines Need to specify genetic material to be sequenced. .
Suggested language:
… For peanut genome sequencing and assembly, the PGC will provide BGI with genomic DNA from *Arachis hypogea* cv. Tifrunner, cv. GT-C20 and 100 lines of a RIL population developed from Tifrunner x GT-C20. BGI will deep sequence the genome of Tifrunner (**60X**) using a whole genome shotgun approach plus BAC-by-BAC with HiSeq2000 technologies. The genome of GT-C20 will be sequenced using a whole genome shotgun approach to at least **60x**. Each of the 100 RILs will be sequenced to **3x** using a whole genome shotgun approach for SNP calling and alignment to the reference genome (cv Tifrunner). Note proposed additions.

Addition 1: There are only 165 RIL from Tifrunner x GT-C20. It was thought better to do ca 100 RIL from each from two difference matings. A 2nd RIL population from Guo was proposed:.
SunOleic 97R x NC94022 (198 RILs plus two parental lines). Tifrunner is a typical runner market type, C20 is a typical Spanish type, and SunOleic 97R is a typical runner, high oleic oil and susceptible to TSWV, and NC94022 is a selection from a cross of Virginia x hirsute type with excellent TSWV resistance.

Addition 2: High resolution genome maps of the A and B genomes of the cultivated peanut ancestors and the amphidiploid synthetic hybrid of A x B genomes species are being developed. UC-Davis is performing WGAS on 96 RILs form an AA genome progenitor species. A WGS approach on specific accessions of *A. duranensis* and *A. ipaensis* (BAC libraries exist) should be added to the BGI proposal.

Sample Requirements
Addition 1**:** DNA will be provided by **Ozias-Akins/Holbrook** for Tifrunner, GT-C20 and 100 RILs; and SunOleic 97R, NC94022 and 100 RILs. Action Note: 200 g FWT leaf tissues for BAC x BAC of Tifrunner = ca. 900 leafs, not feasible for 1 plant. Agreed to use seedlings traced to a single plant, grown in GH, genotype with same SSR panel used to confirm homozygousity of T x G RIL population. Tissue & DNA from Tifton as follows:

- Tifrunner, tissue for BAC x BAC plus DNA for WGS (60X), Library insert size: 180 bp, 500 bp, 800 bp, 2 kbp, 5 kbp, 10 kbp, 20 kbp, **40 kbp** (plus data depth [X}, Gb & number of libraries/insert size)
- GT-C20, DNA for WGS (60X). Library insert size: 180 bp, 500 bp, 800 bp, 2 kbp, 5 kbp, 10 kbp, 20 kbp, **40 kbp** (include data depth [X}, Gb & number of libraries/insert size)
- Tifrunner x GT-C20 RILs, DNA for WGS (3X). Library insert size: 500 bp (include data depth [X}, Gb & number of libraries/insert size)
- SunOleic 97R, DNA for WGS (10X). Library insert size: 500 bp
- NC94022, DNA for WGS (10X). Library insert size: 500 bp
- SunOleic 97R x NC94022 RILs, DNA for WGS (3x). Library insert size: 500 bp

Addition 2: **Ozias-Akins/Bertioli** will provide single-seed source DNA and tissues from two progenitor species, and DNA from diploid AA and BB RILs. Amphidiploids and the RILs derived from cultivated x amphidiploid cross will not be included at this time. The two diploid ancestors would be subject to WGS strategy proposed by BGI, but not the BAC-by-BAC approach. The sequencing of the AA and BB RILs is underway at UC-Davis. Seed source for single plants and DNA from Tifon.

- A genome progenitor, DNA for WGS (10X) Library insert size: 500 bp
- B genome progenitor, DNA for WGS (10X). Library insert size: 500 bp

Major Concerns

C1: US China MOST leaders need to be shown that the PGP is a collaborative research effort. This must be made clear for a successful outcome from the USDA meeting on September 8.

Suggestions to demonstrate collaboration with BGI:

- A PGP Technical Steering Committee to address current and future genome/informatics needs.
- An International PGC Annotation Group to interface with BGI for peanut genome annotation and the establishment of a controlled vocabulary nomenclature.
- A joint plan for long-term curation of the peanut genome sequence, updates on annotation, correction of assembly errors and incorporation of other relevant data
- A system for reporting research progress by each PGP collaborator on a quarterly basis

C2: Agreement is needed on when and the type of data that will be released by BGI to PGP collaborators. Pre-publication release is encouraged by federal granting agencies. Timeline for data release should be included in BGI proposal. Mechanisms should be put in place for interactive data transfer between BGI and other PGP collaborators.

Actions

A1: Jackson appointed a PGP Technical Steering Committee to interact with BGI and coordinate data/ material exchange among all PGP collaborators. Initial members are **Scheffler, D. Bertioli, Froenicke, Ozias-Akins, Guo, Wilson**.

A2: Jackson proposed the following language regarding data release policy for the BGI proposal.

> Data Release and Access Statement
> The Peanut Consortium works off the premise that rapid and public access to data is the ultimate goal, necessary to maintain openness across the scientific community and for the advancement of knowledge and peanut improvement. Therefore, the peanut genome data will be publicly accessible as soon as reasonably possible in order to benefit the entire community.
>
> Data release guidelines:
> Intermediate assemblies, routinely done in whole genome shotgun assemblies, will be released publicly, without restriction, except those contained in the Bermuda Agreement (e.g. exclusion of publication of large-scale analyses), via BGI's website, or on request. We anticipate at least two such intermediate releases prior to the final assembly.
>
> Final assembly: On completion of the final assembly, it will be released with all underlying data into the public domain. The genome assembly will be deposited at, but not limited to, the Legume Information System at NCGR (New Mexico) and NCBI. All ancillary data/information will be available without restriction from BGI upon request or via their website. This is to include resequencing of RIL lines, parents, mapping information, genomic libraries and all sequence data.

A3: Wilson proposed the following language adapted from current USDA NIFA guidance regarding the general PGC data release policy for the P&P.

> SHARING OF FINDINGS, DATA, AND OTHER PROJECT PRODUCTS
> Collaborating members of the PGP are expected to publish or otherwise make publicly available the results of work conducted except in cases where such disclosure would jeopardize proprietary information developed during the course of the project.
>
> GENOME MAP AND SEQUENCE DATA DISCLOSURE
> All investigators working under the auspices of the PGP are encouraged to collaborate and distribute peanut genome and sequence data via relevant worldwide web sites. Genome sequences, protein sequences, and genomic resources must be available to all segments of the scientific community, including industry and the international community. A reasonable charge is permissible for distribution. If accessibility differs between industry and the academic community, the differences must be clearly described in a plan of work.
>
> The PGC supports the currently accepted community standards (Bermuda and Ft. Lauderdale agreements; www.wellcome.ac.uk/assets/wtd003207.pdf) for rapid release of genome sequences following the current guidelines for quality assessment as described by the National Institutes of Health (NIH) National Human Genome Research Institute (NHGRI) at: www.genome.gov/10000923 and www.genome.gov/10001812).

- DNA sequence assemblies of 2kb or greater are to be deposited in a pre-existing public nucleotide sequence database (such as GenBank: www.ncbi.nlm.nih.gov) within 24 hours of generation.
- Sequences are to be deposited in a trace archive (such as the National Center for Biotechnology Information {NCBI} Trace Repository) within one week of production.
- Whole genome shotgun sequences are to be deposited in a trace archive (NCBI Trace Repository or Ensembl Trace Server) within one week of production.
- Whole genome assemblies are to be deposited in a public nucleotide sequence database as soon as possible after the assembled sequence has met a set of quality evaluation criteria.
- Other nucleotide sequences such as ESTs, full-length cDNA sequences, etc. must be submitted to a pre-existing public nucleotide sequence database (such as Genbank) within one month of production and quality assessment.
- Community resources (e.g., single nucleotide polymorphisms, SNP; haplotype maps, plant genetic stocks) whose primary utility will be for the broad scientific community will be made immediately available for free and unrestricted use by the scientific community as soon as the quality of these resources is verified.
- Products of PGP research that uses proprietary data or materials from other sources must be readily available without any restrictions to the users (no reach-through rights).
- PGP investigators that use genetic resources from outside the United States will seek information regarding any required prior informed consent from and benefit sharing with the appropriate host country authorities

Points for consideration by BGI:

Q1: Whole Genome Assembly page 9, para 1 ….generate ca. 100X paired-end 100/50 bp Solexa reads, …. (should 50 be 500 or 1500?)

P1: Demonstrating collaborative research interaction with BGI would be beneficial to discussion of including the PGP under the USDA China MOST (Ministry of Sceince & Technology) program. A first step would involve formation of a PGP Technical Steering committee for BGI and cooperating Chinese institutions.

P2: The number of SNPs discovered by GWS & BAC-ends of RIL populations may be insufficient to tag a high percentage of contigs BAC-Fish maps show large sections of DNA at end of chromosomes with no apparent markers. The SNP rate in *A. hypogaea* may be ca. 1.5/contig. A high density map may have low resolution.

P3: How will diploid/tetraploid gene-space data and SNP maps from diversity study be integrated/ provided for use in genome assembly?

Q2: A better description is needed on how the BAC and WGS assembly pipelines will be merged.

These notes represent a summary of the proceedings of the ad hoc committee on Aug 22 and Aug 25, 2011

Adjourned