

# The Peanut Genome Consortium

## ‘Creating a Better Future through Global Food Security’

### Background:

The Peanut Genome Consortium (PGC) is an extension of the International Peanut Genome Initiative (IPGI), and is embodied by a coalition of international scientists and stakeholders engaged in the Peanut Genome Project (PGP).

The *International Strategic Plan of the Peanut Genome Initiative 2012-2016* documents IPGI research priorities. PGC governance is referenced in *Policies & Procedures*. These files are posted at <http://www.peanutbioscience.com/>.

The PGP includes genome distinguishing gene expression studies, characterization of germplasm diversity, a genome wide association study (GWAS), a phenotyping component, high density genome maps of the cultured variety & of wild peanut crosses, plus the generation of a reference sequence.

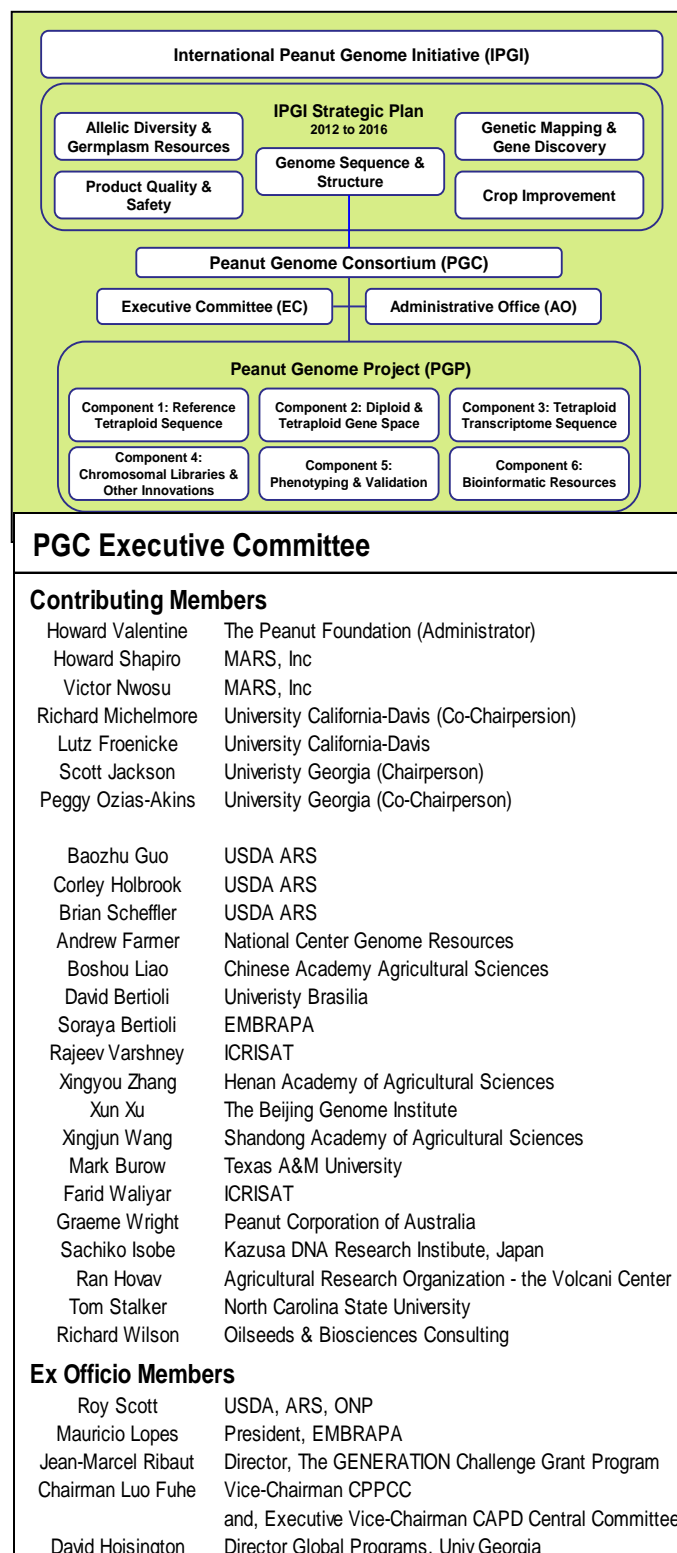
Specific PGP goals include: 1) development of a high quality chromosome scale draft of a tetraploid (cultivated species) as the reference genome sequence, 2) high throughput genome and transcriptome characterization of tetraploid, amphidiploid and diploid (progenitor species) germplasm, 3) phenotypic trait association with mapped genetic markers, and 4) interactive bioinformatic resources for data curation and analysis. These efforts will enable molecular breeding approaches for enhancing peanut yielding ability, optimizing resistance to diseases and insects, tolerance to environmental stresses, and improved quality traits. This technology will promote peanut crop competitiveness and enhance grower’s profitability in an environmentally sustainable manner.

**Issue:** Whole genome sequence (WGS) of peanuts is needed to expedite elite variety selection & to help ensure global food security & safety.

### PGP Milestones & Deliverables

*Generation of a high quality reference genome sequence of cultivated peanut that is anchored to chromosomal linkage groups (PGP Component 1)*

- WGS of the cv Tifrunner library with sequence depth of at least 300x
- WGS of the GT-C20, with sequence depth of 60x
- WGS of 100 RILs from Tifrunner x GT-C20 at 1x
- WGS of 100 RILs from SunOleic97R x NC94022
- A high-resolution genome assembly of each haplotype present in the 2 genomes of the allotetraploid.
- Raw data sufficient to allow genome assembly
- Physical and genetic coordinates of the scaffolds and contigs in chromosomal linkage groups based on analysis of the segregation data from the RILs.
- Data on genome assembly quality (euchromatic and gene region coverage with sequencing depth)
- Genome annotation results (repeat analysis; protein-coding genes including gene structure prediction and gene function annotation; non-coding microRNA, tRNA, rRNA and other ncRNA; transposons and tandem repeats)
- Comparative genomics and evolution analysis (chromosome structure variation detection of specific genome regions; fast evolutionary gene detection; synteny blocks; and gene family analysis)
- Validated genome assembly with a linear order of the contigs in chromosomal linkage groups.



### *Genome mapping and allelic analysis through GWAS (PGP Component 2)*

- High resolution genome maps of A and B genomes of the cultivated peanut ancestors and the amphidiploid synthetic hybrid of A x B genomes species.
- SNP maps correlated with the variation captured in the diversity panels and germplasm collections.
- GWAS studies of the agricultural traits phenotyped on the ICRISAT genetic diversity panel.

### *Catalog expressed genes and profile gene expression in cultivated peanut (PGP Component 3)*

- Expression profiles of genes that mediate resistance to diseases and pests, such as: tomato spotted wilt virus (TSWV), leaf spot (early - *Cercospora arachidicola*; late - *Cercosporidium personatum*), rust (*Puccinia arachidis*), white mold (*Sclerotium rolfsii*), nematode (*Meloidogyne arenaria*), and pre-harvest aflatoxin contamination (*Aspergillus flavus*)
- Expression profiles of genes that mediate tolerance to abiotic stresses, such as: drought, temperature (cold, heat), and nutrient deficiency
- A peanut gene atlas which includes the diploid genome origin of genes plus a comprehensive list of all expressed soybean genes, alternative splice products, identification of co-regulated genes and gene networks.

### *Evaluation of emerging technologies for genome sequencing and characterization (PGP Component 4)*

- Direct sequencing of individual chromosomes using the Pacific Biosciences platform for single molecule real-time analysis
- Strobe sequencing for scaffolding contigs and assigning haplotypes in heterozygous and tetraploid genomes.

### *Phenotypic validation of gene predictions (PGP Component 5)*

- DNA markers that contribute to the assembly and annotation of the peanut genome
- DNA markers that can be used in pre-breeding for disease and pest resistance including TSWV, Early & Late Leaf Spot, CBR, nematodes, PAC, drought.
- DNA markers that can be used in pre-breeding for quality traits including seed fatty acid composition, flavor quality, nutritional benefits, and other seed composition traits
- DNA markers for peanut yielding ability and other traits

### *Development of bioinformatic resources for peanut genome data (PGP Component 6)*

- A PGP Informatics Steering Committee to address current and future informatics needs.
- An International PGC Annotation Group to interface with BGI for peanut genome annotation and the establishment of a controlled vocabulary nomenclature.
- A peanut genomic database that facilitates navigation from maps to genes to traits
- An integrated database including available genetic stocks, mutants and germplasm collections
- A HapMap browser that connects the sequence to polymorphisms for traits of interest
- Ability to map RNA-seq and Sanger reads from expression data onto QTL data
- Integration of genome sequence with physical, genetics and transcriptome maps
- Molecular tools for the identification of candidate genes underlying QTLs.
- Integration of plant trait and phenotypic data with genetic maps and other genetic data.
- A plan for long-term curation of the peanut genome sequence, updates on annotation, correction of assembly errors and incorporation of other relevant data

### **Summation:**

Cultivated peanut poses the most difficult challenge that has been attempted in crop genome sequencing. Genome size is nearly equal to humans, and is distinguished by two sets of chromosomes from two different *Arachis* species.

Peanut varietal development worldwide is totally a function of the public research sector. DNA sequence assisted breeding methods are essential for timely increases in crop productivity & quality to ensure global food security.

The PGC is well positioned to deliver a reference peanut genome sequence within 2 to 3 years by virtue of multilateral multi-disciplinary cooperative research relations with world class scientists, organizations and institutions in the U.S., China, Brazil, India, Japan, W. Africa, Australia, and Israel that possess the latest advances in genomic technology.