

340  
/1421  
2016

**Title:** Leveraging genomic data and tools from botanical models in peanut improvement  
**Funding Year:** 2016

**Investigator:** Andrew Paterson, Regents Professor & Head, Plant Genome Mapping Lab (Dept. #398), Univ. GA, Athens. Ph. 706-742-8658, email [paterson@plantbio.uga.edu](mailto:paterson@plantbio.uga.edu)

**Layman's Summary:** Most plants share a largely-common set of genes, and gene functions change only very slowly. Using information from well-studied botanical models, the functions of perhaps half of the estimated 30,000 peanut genes might be deduced, *once we understand the gene-by-gene relationships among these genomes*. Using methods that we pioneered (Paterson et al., 2010; Tang et al., 2008), our objective is to 'align' key segments and chromosomes of peanut to those of such models, providing the means to begin to deduce functions of peanut genes based on knowledge from well-studied plants.

**Project Purpose:** The greatest challenge now facing the peanut community is no longer genome sequencing *per se* but the conversion of sequence to actionable knowledge. For example, completion of the *Arabidopsis thaliana* sequence was quickly followed by inception of the NSF 2010 project, which has greatly increased knowledge about functions of *Arabidopsis* genes but at a cost near US\$200 million. This project will position the peanut community to leverage the collected knowledge of gene functions in many additional plants, jump-starting progress toward understanding the functions of peanut genes at very modest cost.

**Experimental Plans and Methods:** While many investigations show parallels of chromosome 'macrostructure' among well-studied legume genomes, as well as between these genomes and those of botanical models, data for investigating these potentially important parallels in peanut are just beginning to achieve critical mass. The required data for peanut are now emerging from several sources including (a) sequencing of carefully-chosen segments of the peanut genome that contain particular genes of interest (disease resistance genes, allergen-encoding genes, nodulation-related genes), (b) physical mapping including end-sequencing of large-insert BAC clones of peanut DNA and hybridization of sequence-tagged site DNA markers to the same BACs, (c) enriched genetic mapping, and (d) whole genome sequencing.

Using methods that we pioneered (Tang et al., 2008), we will utilize these emerging peanut data to begin to 'align' the chromosomes of peanut to those of sequenced legumes and botanical models, providing the means to begin to locate key peanut genes and deduce their functions based on knowledge from better-studied plants.

**Progress.** We have engaged in two approaches to accomplish the proposed goal of 'aligning' the chromosomes of peanut to those of sequenced legumes and botanical models, providing the means to begin to locate key peanut genes and deduce their functions based on knowledge from better-studied plants.

1. *Participation in annotation and analysis of a peanut genome assembly* (Chen et al 2016, below) – We participated in the assembly and annotation of a draft genome of the peanut A-genome progenitor, *Arachis duranensis*, and identification of its 50,324 protein coding gene models. Patterns of gene duplication suggest the peanut lineage has been affected by at least

three polyploidizations since the origin of eudicots. Re-sequencing of synthetic *Arachis* tetraploids reveals extensive gene conversion in only 3 seed to seed generations since their formation by human hands, indicating that this process begins virtually immediately following polyploid formation. Expansion of some specific gene families suggests roles in the unusual subterranean fructification of *Arachis*. For example, the S1Fa-like transcription factor family has 126 *Arachis* members versus no more than 5 in other examined plant species, and is more highly expressed in roots and etiolated seedlings than green leaves. The *A. duranensis* genome provides a major source of candidate genes for fructification, oil biosynthesis and allergens, expanding knowledge of under-studied areas of plant biology and human health impacts of plants, informing peanut genetic improvement and aiding deeper sequencing of *Arachis* diversity.

*2. Alignment of 10 sequenced legume genomes to clarify relationships and provide for a cross-legume research platform* (Wang et al, in review, below)-- By performing comparative and hierarchical genomic analysis, with grape, medicago, and common bean genomes as outgroups, we produced a hierarchical and event-related alignment of the legume genomes, deconvoluted layers of the aligned genomic regions corresponding to ancestral polyploidizations and speciations, and deciphered gene-colinearity-supported paralogs derived from these events. We also illustrated genomic fractionation after the polyploidizations. Though gene losses were mostly random, being largely but not fully described by a geometric distribution, we showed that polyploidization contributed divergently to copy number variation of gene families, with genes related to macromolecular complexes, membrane functions and other classes significantly retained, while those related to biological regulation were preferentially deleted or transposed. Based on genome-scale homologs inferred above, we showed significantly divergent evolutionary rates among legumes, and by performing rate correction, re-dated major evolutionary events. We also inferred ancestral genome contents at major evolutionary nodes. The present effort laid a solid foundation for further genomic exploration in the legume research community and beyond, with the required information stored in the newly constructed Legume Comparative Genomics Research Platform ([www.legumegrp.org](http://www.legumegrp.org)).

At present, we are building a database to accelerate the identification and study of plant resistance genes ('R genes') in entire genomes, including a tool for gene family visualization that expedites study of these usually complex gene families. Peanut is a key test data set and will be an early application.

### ***Citations resulting from this work:***

Xiaoping Chen, Hongjie Li, Manish K. Pandey, Qingli Yang, Xiyin Wang, Vanika Garg, Haifen Li, Xiaoyuan Chi, Dadakhalandar Doddamani, Yanbin Hong, Hari D. Upadhyaya, Hui Guo, Aamir W. Khan, Fanghe Zhu, Xiaoyan Zhang, Lijuan Pan, Gary J. Pierce, Guiyuan Zhou, Katta AVS Krishnamohan, Mingna Chen, Ni Zhong, Gaurav Agarwal, Shuanzhu Li, Annapurna Chitikineni, Guoqiang Zhang, Shivali Sharma, Na Chen, Haiyan Liu, Pasupuleti Janila, Shaoxiong Li, Min Wang, Tong Wang, Jie Sun, Xingyu Li, Chunyan Li, Mian Wang, Lina Yu, Shijie Wen, Sube Singh, Zhen Yang, Jinming Zhao, Chushu Zhang, Yue Yu, Jie Bi, Xiaojun Zhang, Zhongjian Liu\*, Andrew H. Paterson\*, Shuping Wang\*, Xuanqiang Liang\*, Rajeev K. Varshney\*, Shanlin Yu\* 2016. Draft Genome of the Peanut A-genome Progenitor (*Arachis*

*duranensis*) Provides Insights into Geocarpy, Oil Biosynthesis and Allergens. (\* co-corresponding authors). Proceedings of the National Academy of Sciences of the USA 113 (24): 6785–6790.

Jinpeng Wang, Pengchuan Sun, Yuxian Li, Ruiyan Xia, Yinzhe Liu, Xuelian Ma, Nanshan Yang, Sangrong Sun, Jigao Yu, Tianyu Lei, Beibei Jiao, Yue Xing, Li Wang, Zhenyi Wang, Weina Ge, Lan Zhang, Xiaoming Song, Dianchuan Jin, Yuxin Pan, Xiaojian Liu, Tao Liu, Jinshuai Sun, Jiaxiang Yu, Jun Qin, Meng-chen Zhang, Andrew H. Paterson, Xiyin Wang 201#. Hierarchical and event-related alignment of 10 legume genomes establishes a genus-level comparative genomics platform and implies an autotetraploid ancestor of soybean (in review).