340/1288
2014

Title: Leveraging genomic data and tools from botanical models in peanut improvement
Funding Year: 2014
Investigator: Andrew Paterson, Regents Professor & Head, Plant Genome Mapping Lab (Dept. #398), Univ. GA, Athens. Ph. 706-742-8658, email paterson@plantbio.uga.edu
Funds requested: $20,000.
Location where research will be performed: Plant Genome Mapping Lab (Dept. #398), Univ. GA, Athens.
New or Continuing: Continuing, 2nd year.

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.

Hypothesis and Objectives: While peanut still lacks a genome sequence (as of the start date), several other legumes and botanical models have now been sequenced, and we are near completion of the first 'physical map' for peanut, a framework for assembling the peanut genetic blueprint in the near future. Our hypothesis is that alignment of the peanut genome to those of other well-studied plants will permit us to deduce the identities and possible functions of perhaps half of the estimated 30,000 peanut genes.

Experimental Plans and Methods: While a growing number of 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, and (c) enriched genetic mapping.

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 to date.

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); A manuscript describing the sequencing of ~617 kb from the genome of cultivated peanut and 215 kb from a wild relative including three *Arah1*, one *Arah2*, eight *Arah3* and two *Arah6* gene family members has been published. To assign polarity to differences between homoeologous regions in peanut, we used as outgroups the single orthologous regions in *Medicago*, *Lotus*, common bean, chickpea and pigeonpea, which diverged from peanut about 50 million years ago (mya) and have not undergone subsequent polyploidy. These regions were also compared with orthologs in many additional dicot plant species, to help clarify the timing of evolutionary events. The lack of conservation of allergenic epitopes between species, and the fact that many different proteins can be allergenic, makes the identification of allergens across species by comparative studies difficult. The peanut allergen genes are interspersed with low-copy genes and transposable elements. Phylogenetic analyses revealed lineage-specific expansion and loss of low-copy genes between species and homoeologs. *Arah1* syntenic regions are conserved in soybean, pigeonpea, tomato, grape, *Lotus* and *Arabidopsis* while *Arah3* syntenic regions show genome rearrangements. We infer that tandem and segmental duplications led to the establishment of the *Arah3* gene family. Our analysis indicates differences in conserved motifs in allergen proteins and in the promoter regions of the allergen encoding genes. Phylogenetic analysis and genomic organization studies provide new insights into the evolution of the major peanut allergen encoding 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; Three genome assemblies for peanut progenitors have been released recently. We have full access to one, however publication of analyses of the other two are forbidden until the sequencing team publishes their primary manuscript (http://peanutbase.org/data_usage_agree_aradu). Accordingly, we are focusing on an initial assembly of the sequence that is available for us to publish on, *A. duranensis*, and will look at other assemblies when analyses can be published. Efforts to increase contiguity (the length of un-interrupted sequence), so as to increase the ability to identify 'orthologous' (corresponding genes between divergent taxa, have been completed. We identified and provided to the sequencing team BAC clones that were terminal to their 'contigs' (contiguous groupings) based on our physical map. These clones were sequenced in a pooled-sample strategy by the sequencing team. We have now completed (and submitted for publication) study of the extent to which duplicated genes from the respective peanut subgenomes have recombined with one another since polyploid formation, potentially doubling the dosage of specific alleles and providing new variation that we have shown in other polyploid crops (cotton, Paterson et al 2012 and Guo et al 2014) to contribute to the novel features that distinguish crops from their wild relatives.

(c) enriched genetic mapping. At present, data from this approach is only available from the two genome assemblies for which access is restricted. This analysis will be deferred until it can be published.