final

Translational Genomics + molecular breeding approaches for accelerating and enhancing

Title: Enhancing Genomic Resources for Peanut

Principal Investigator

Steven J. Knapp
Institute of Plant Breeding, Genetics, and Genomics
Crop and Soil Sciences Department
111 Riverbend Road
The University of Georgia
Athens, Georgia, 30602
TEL: 706-542-4021
E-MAIL: sjknapp@uga.edu

Objectives and Accomplishments

Our research focuses on enhancing the infrastructure for genomics and marker-assisted breeding research in cultivated (tetraploid) peanut (*Arachis hypogaea* L.) through high-density genetic mapping in A- and B-genome diploid species, comparative mapping and macrosynteny analysis between diploid and tetraploid species, and the large-scale development of simple sequence repeat (SSR) and single-nucleotide polymorphism (SNP) marker resources for cultivated peanut. The application of next-generation DNA sequencing and single nucleotide polymorphism (SNP) genotyping technologies are playing an important in our research. We specifically proposed: (a) producing, annotating, and assembling short- and long-read ESTs (SR- and LR-ESTs, respectively) from normalized cDNAs isolated from developing seeds of diploid and tetraploid lines; (b) gaining insights into phylogeographic, genetic, and nucleotide diversity in *A. duranensis* (AA), *A. batizocoi* (BB), and *A. hypogaea* (AABB); (c) identifying single nucleotide polymorphisms (SNPs) in diploid and tetraploid populations; (d) developing intraspecific A- and B-genome diploid mapping populations; and (e) comparative genetic mapping of the A- and B-genomes and identification of chromosomal rearrangements between the A- and B-genomes by genotyping SSR and SNP markers in intraspecific *A. duranensis* and *A. batizocoi* F_2 populations.

We developed normalized cDNA libraries from root and developing seed RNAs isolated from *A. duranensis* (the A-genome progenitor), *A. ipaensis* (the B-genome progenitor), and multiple *A. hypogaea* cultivars (New Mexico Valencia A, NC12C, Tifrunner, GT-C20 and A13), produced 35,291 long-read (Sanger) ESTs for *A. duranensis* and 32,787 long-read ESTs for *A. ipaensis*, produced 1,000,000 short-read (454-FLX) ESTs for two *A. duranensis* ecotypes (DUR2 and DUR25), produced 304,215 454-FLX ESTs for two *A. hypogaea* cultivars (New Mexico Valencia A and NC12C), built tetraploid and diploid-tetraploid transcript assemblies (using Sanger and 454-FLX ESTs), and developed a searchable public EST database for peanut (http://www.peanut.uga.edu/).

The tetraploid peanut transcript assembly (101,132 unigenes) was mined for SSRs. We identified 7,413 perfect repeats, developed and screened 2,134 EST-SSR markers for amplification and length polymorphisms among the parents of two diploid and two tetraploid mapping populations, and developed 1,557 high-quality EST-SSR markers for peanut. Of the

2,134 EST-SSR primers tested 577 of the EST-SSR primers tested (27%) either failed to produce amplicons or produced complex amplicons or amplicons longer than 1,200 bp (the limit for our SSR genotyping platform). While 80% were polymorphic among the parents of diploid mapping populations, only 16% were polymorphic among the parents of the tetraploid mapping populations. Collaborators at ICRISAT, Hyderabad, India (Rajeev Varsheny, Rupakula Aruna, and David Hoisington) are currently screening a diverse panel of 24 cultivars for EST-SSR marker polymorphisms. From our previous screening of genomic survey sequence (GSS) SSR markers, we anticipate at least 30% of the EST-SSR markers to be polymorphic among elite and exotic tetraploid peanut lines.

We developed three intraspecific diploid mapping populations, two for *A. duranensis* (DUR2 x DUR25 and DUR25 x DUR35) and one for *A. batizocoi* (BAT9 x BAT12), in addition to producing a synthetic AABB allotetraploid from an *A. duranensis* x *A. batizocoi* hybrid (DUR37 x BAT8). We produced 1,000 to 1,600 F_2 seeds per population, isolated DNA from 94 DUR2 x DUR25, and are currently isolating DNA from 94 BAT9 x BAT12 F_2 progeny. We are presently wrapping up SSR and EST-SSR marker genotyping in the DUR2 x DUR25 F_2 population, have genotyped and mapped approximately 300 SSR markers so far, and anticipate having 600 SSR markers mapped by the end of the year. We identified 11 linkage groups which almost certainly represent the 11 chromosomes in the haploid genome of *A. duranensis* ($2n = 2x = 22$). We anticipate having 300 EST-SSR marker loci mapped in the *A. batizocoi* F_2 population and a minimum of 800 EST-SSR marker loci mapped in the A- and B-genomes by March, 2009, of which at least 200 will be mapped in both species.

We developed a SNP discovery pipeline and mined the *A. duranensis* transcript assembly for SNPs between the parents of the DUR2 x DUR25 mapping population. Using stringent criteria, we identified 8,478 SNPs in 3,922 unigenes. To facilitate the selection of candidate SNPs for designing and building Illumina GoldenGate SNP genotyping arrays, putative intron positions were predicted by aligning *Arachis* contigs with *Arabidopsis* and *Medicago* genomic DNA sequences identified by BLAST analyses. SNPs within 60 bp of a putative intron were eliminated, thereby reducing the collection of candidate SNPs to 6,789 in 3,264 unigenes; however, through reanalysis with less stringent criteria in consultation with Illumina (a shorter flanking sequence window), we identified 7,900 SNPs in 3,500 unigenes. Our transcript assembly, ESTs, and candidate SNPs databases have been submitted to Illumina for bioinformatic analyses and assigning SNP genotyping quality scores. Once the latter have been received, we will select 3,072 SNPs for the development of two 1,536-SNP arrays and proceed with SNP genotyping and mapping in the DUR2 x DUR25 F_2 mapping population. We anticipate validating and mapping 2,150-2,458 (70-80%) of the SNPs tested. Hence, between EST-SSR and SNP markers, we anticipate mapping 3,000-3,400 transcribed loci in the A-genome and upwards of 3,300-3,700 transcribed loci in the A- and B-genomes.

One of the 1,536-SNP arrays will include SNPs identified by allele resequencing in a subset of 768 conserved orthologous sequences identified by collaborators at UC-Davis (Varma Penmetsa and Douglas Cook). This array will be used for genotyping SNPs in an interspecific *A. duranensis* x *A. stenosperma* F_2 population developed by a collaborators at EMBRAPA and Catholic University (David Bertioli).

We mined the tetraploid EST database for NBS-LRR cDNA sequences, developed 403 SSCP markers for 318 partial NBS-LRR genomic or cDNA sequences identified in the EST database and GenBank, screened 556 SSR and 403 SSCP markers for polymorphisms between nematode susceptible and resistant lines, and mapped a nematode resistance gene (*Rma*) and a wild genomic DNA segment dragged with *Rma* in tetraploid RIL mapping populations (NemaTAM x WS14 and Tifguard x Gregory) developed by collaborators at the University of Georgia and USDA-ARS, Tifton, Georgia (Peggy Ozias-Akins, Juliet Chu, and Corley Holbrook). Several SSR and NBS-LRR loci mapped in the introgressed segment and were tightly linked with *Rma*. Genetic mapping was limited to NemaTAM x WS14 because recombination was completely suppressed in the genomic segment spanning the wild introgression in Tifguard x Gregory (no recombinant were observed in this population). To locate the introgression in the peanut genome and gain deeper insights into the organization and genetic distances of loci spanning the wild introgression, the aforementioned SSR and SSCP marker loci were mapped in the DUR2 x DUR25 F_2 mapping population. These loci spanned 32 cM, a distance 10-fold greater than the same loci in the NemaTAM x WS14 F_2 population (3 cM); hence, the *A. cardenasii* introgression spanned at least one-third of a chromosome in two wild introgression lines (NemaTAM and Tifguard). Moreover, recombination appears to be strongly suppressed in segregating populations developed from crosses between cultivars and NILs carrying the *A. cardenasii* introgression. We developed a NemaTAM fosmid library and are currently shotgun sequencing two fosmid clones to isolate full-length sequences for NBS-LRR encoding genes linked to *Rma*. Our collaborators at Tifton are planning functional analyses of NBS-LRR genes identified as candidates for *Rma*.

We are currently testing a next-generation sequencing strategy for simultaneous SNP discovery and mapping in collaboration with Floragenex, Eugene, Oregon (Eric Johnson and Rick Knipper). This strategy entails reducing genomic DNA complexity by capturing genomic DNA fragments produced by restriction-enzyme digestion, shearing, and length selection. We are testing this strategy in a diploid F_2 population (DUR2 x DUR25) and two tetraploid recombinant inbred line populations (TAG24 x ICGV 86031 and ICGS 76 x CSMG 84-1). The latter were developed by collaborators at ICRISAT, Hyderabad, India (Rajeev Varsheny, Rupakula Aruna, and David Hoisington) and are part of an ongoing Generation Challenge Program collaboration. Genetic mapping of the aforementioned EST-SSR markers is underway in these populations at ICRISAT and will supply the tetraploid information needed for needed macrosynteny analyses with A- and B-genome species, *Medicago*, *Glycine*, and other legumes.

Finally, we are collaborating on the development of 16 mapping populations for tetraploid peanut with colleagues at NCSU (Tom Isleib, Susanna Milla-Lewis, and Tom Stalker), the University of Florida (Barry Tillman, Mario Gallo, and Brad Barbazuk), USDA-ARS, Tifton, Georgia and the University of Georgia (Peggy Ozias-Akins and Corley Holbrook), and Texas Tech University (Mark Burow). These populations sample the spectrum of genetic diversity found in tetraploid peanut, were developed crossing two elite parents with eight elite and exotic parents, and will ultimately supply a collection of structured recombinant inbred line (RIL) mapping populations for comparative mapping and meta-analyses of phenotypic and quantitative trait loci.

Broad Impacts

Our research has significantly enhanced the infrastructure for forward genetics analysis, genomics, and marker-assisted breeding in peanut. DNA sequences, DNA markers, the EST database, and other databases and resources have been shared with collaborators in the US and abroad, have been catalysts for several collaborations, and are accelerating and enhancing the application of genomics approaches in peanut.

Publications

Varshney, R.K., D.J. Bertioli, M.C. Moretzsohn, K. Ravi, V. Vadez, L. Krishnamurthy, R. Arunal, S.N. Nigam, B.J. Moss, K. Seetha, G. He, S.J. Knapp, and D.A. Hoisington. The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. (in press).

Presentations

Nagy, E.D., Y. Chu, Y. Li, W.B. Dong, P. Timper, P. Ozias-Akins, C.C. Holbrook, B. Rosen, D. Cook, S.J. Knapp. Discovery of NBS-LRR encoding genes linked to a dominant nematode resistance gene (*Rma*) in groundnut lines carrying a genomic segment introgressed from a wild diploid donor

. Plant and Animal Genome Meeting, San Diego, California, January, 2009.

Taylor, C.A., S. Tang, E. Bachlava, A. Farmer, S. Ayyampalayam, J. Huntley, G.D. May, and S.J. Knapp. SNP Discovery by massively parallel transcriptome resequencing in sunflower and development of a bioinformatic pipeline and database for mining and displaying SNPs in next-generation sequence assemblies. Plant and Animal Genome Meeting, San Diego, California, January, 2009.

Knapp, S.J. SNP discovery, validation, and mapping in groundnut. Generation Challenge Program Annual Meeting, September, 2008, Bangkok, Thailand (Invited Presentation).

Nagy E, Chu Y, Li Y, Dong WB, Timper P, Ozias-Akins P, Holbrook CC, Radwan O, Rosen B, Cook D, Knapp SJ. Discovery and genetic mapping of NBS-LRR encoding resistance gene candidates linked to a root knot nematode resistance gene (*Rma*) introgressed from a wild diploid donor in groundnut. Generation Challenge Program Annual Meeting, September, 2008, Bangkok, Thailand.

Khanal S, Tang S, Nagy E, Guo Y, Li Y, Beilinson V, San Miguel P, Guo B, Nielsen N, Stalker T, Cordonnier-Pratt MM, Pratt LH, Johnson VE, Taylor CA, Wiley GB, Macmil SL, Roe B, Ravi K, Naidu G, Hoisington D, Varshney R, Knapp SJ. EST-SSR marker resources for groundnut. Generation Challenge Program Annual Meeting, September, 2008, Bangkok, Thailand.

Knapp, S.J. Genomics resource development for peanut. American Peanut Council Annual Meeting, Washington DC, February, 2008.

Knapp, S.J. Enhancing the infrastructure for comparative and translational genomics in peanut. USDA CSREES Plant Genome Program Symposium, Plant and Animal Genome Meeting, San Diego, California, January, 2008.

Khanal S, Tang S, Beilinson V, San Miguel P, Guo B, Nielsen N, Stalker HT, Pratt MMC, Pratt L, Knapp SJ. ESTs are a rich source of polymorphic SSRs for genomics and molecular

breeding applications in peanut. Plant and Animal Genome Meeting, San Diego, California, January, 2008.
Knapp, S.J., The US Peanut Genomics Intitiative, American Peanut Council Peanut Genomics Initiative Strategic Planning Meeting, November, 2007.
Knapp, SJ. Enhancing the infrastructure for translational genomics in peanut. UC-Davis NSF Comparative Legume Genomics Workshop, August, 2007, Skamania Lodge, Washington.
Ma W, Li, Y, Guo B, Culbreath AK, Milla-Lewis S, Tallury S, Holbrook CC, Isleib T, Stalker HT, Tang S, Knapp SJ. Simple sequence repeat polymorphisms in peanut. Plant and Animal Genome Meeting, January, 2007, San Diego, California.

Community Resources

Thus far, our primary contributions have been the development of community resources. We have, as noted earlier, shared unpublished EST-SSR markers, EST and DNA sequences, transcript assemblies, and the EST database and other databases with collaborators in the US and abroad. We have shared the diploid mapping populations and are participating in the development of a large number of tetraploid mapping populations through a community wide collaboration in the US. The mapping information we are generating has been shared with collaborators to speed genetic mapping in tetraploid populations.

Training

Shunxue Tang (Research Professor)
Yufang Guo (Postdoctoral Research Associate)
Ervin Nagy (Postdoctoral Research Associate)
Wenshi Ma (Postdoctoral Research Associate)
Nelly Khalilian (Research Professional II)
Jennifer Wood (Research Professional II)
Sameer Khanal (M.S. Student)
Yan Li (M.S. Student)
Sarah Hafner (Undergraduate Student)
Rebecca Okashah (Undergraduate Student)
Alison Richbourg (Undergraduate Student)
Bryan Bardash (Undergraduate Student)

Collaborations

Varma Penmetsa and Douglas Cook (UC-Davis).
David Bertioli (EMBRAPA and Catholic University)
Peggy Ozias-Akins, Juliet Chu, and Albert Culbreath (The University of Georgia)
Corley Holbrook (USDA-ARS, Tifton, Georgia (Corley Holbrook)
Tom Isleib, Susanna Milla-Lewis, and Tom Stalker (North Carolina State University)
Barry Tillman, Mario Gallo, and Brad Barbazuk (The University of Florida)
Mark Burow (Texas Tech University)
Rajeev Varsheny, Rupakula Aruna, and David Hoisington (ICRISAT, Hyderabad, India)

Gregory May and Andrew Farmer (National Center for Genomic Resources, Santa Fe, New Mexico)