Diversity and distribution of single feature polymorphisms in food crop legumes.

R. Varma Penmetsa[1]; Andrew D. Farmer[2]; Noelia Carrasquilla-Garcia[1]; Birinchi K. Sarma[1,3]; Subhojit Datta[1,4]; Sally L. Vail[5]; Bert Vandenberg[5]; Jimmy Woodward[2]; Gregory D. May[2]; George E. Bruening[1]; Rajeev K. Varshney[6]; Douglas R. Cook

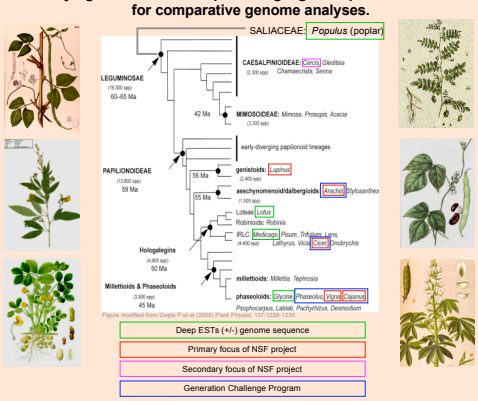
[1]Department of Plant Pathology, University of California-Davis, CA, USA; [2]National Center for Genome Research, Santa Fe, NM, USA; [3]Current address: Department of Mycology & Plant Pathology, Banaras Hindu University, Varanasi, India; [4]Current address: Indian Institute of Pulses Research, Kanpur, India; [5]Applied Genomics Laboratory, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India; [6]Department of Plant Sciences, University of Saskatchewan, Saskatoon, Canada

ABSTRACT

Using a set of 1440 low degeneracy oligonucleotide pairs developed for comparative mapping across crop legumes, we amplified and Sanger sequenced gene orthologs from the food legumes chickpea (*Cicer arietinum*), lentil (*Lens culinaris*), pigeonpea (*Cajanus cajan*), cowpea (*Vigna unguiculata*), common bean (*Phaseolus vulgaris*) and peanut (*Arachis hypogaea*). 1043 to 1251 high quality sequences were obtained and aligned for 2-4 parents of mapping populations in each species. We validated 22,827 single nucleotide (SNPs) and indel polymorphisms, within 503 to 691 orthologous genes in each crop legume at an average rate of 6.6 SNPs/locus. High throughput Goldengate OPA assays are being used for genotyping to generate genetic maps in each crop, and comparative maps between crops.

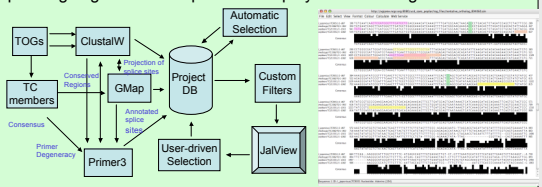
Additionally we are working to develop protocols to define the diversity and distribution of SNPs in the large (>10K accessions) germplasm collections in each food legume via next generation sequencing technologies. This will establish high-throughput genotyping systems for marker-assisted selection, and characterize the genetic structure of crop species in terms of allele frequency, haplotype structure and linkage disequilibrium. Results from these ongoing studies and implications for association mapping and allele mining will be described.

Phylogenetic relationships among legume species selected for comparative genome analyses.



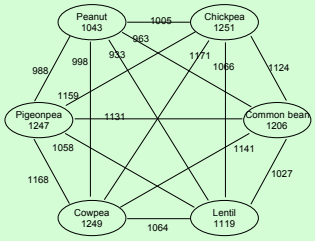
Step 1: Start with Quackenbush's Eukaryotic Gene Ortholog set of 72K best hit clusters plants, animals, fungi and protists

Step 2: Database of best hit homologs, intron position, and intron-spanning oligonucleotide primers. Display on web using JalView.



Database schema and JalView representation of multiple sequence alignment with PCR primers and splice junctions highlighted.

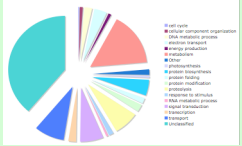
Step 3: Amplify, sequence and identify polymorphisms in high value parents and across germplasm.



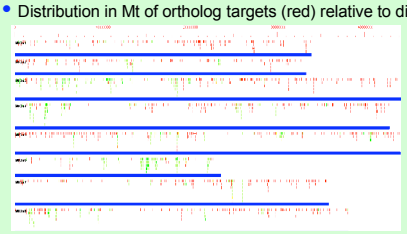
Summary of sequencing of orthologs by species

species	# of COSs with high quality sequence	total high quality sequence (Mbp)
Chickpea	1324	1.180
Cowpea	1396	1.080
Pigeonpea	1372	1.144
Common Bean	1369	1.139
Lupin	1367	1.304
Peanut *	1404	1.177

- Biological Process Ontology of the 1369 Orthologous Markers

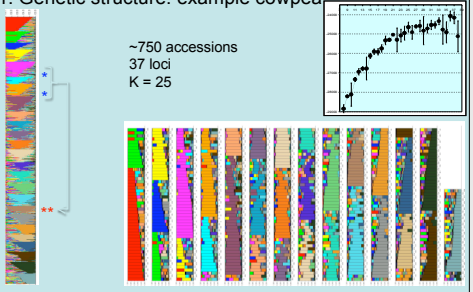


- Distribution in Mt of ortholog targets (red) relative to disease resistance genes (green)

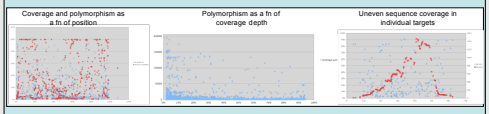


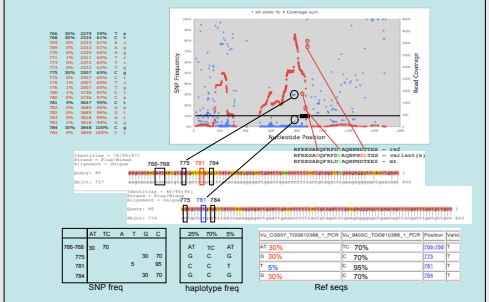
SNP analyses in germpalsm

1. Genetic structure: example cowpea



2. Next Generation discovery: 120K SNP/1369 genes





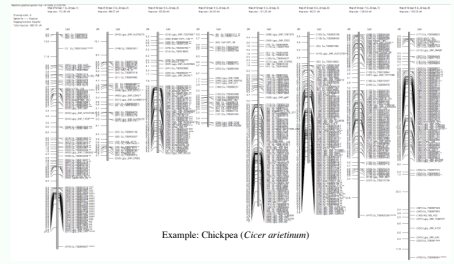
3. Analysis of ortholog SNP in germplasm - Illumina

Allele distribution in 11 mapping genotypes of common bean.

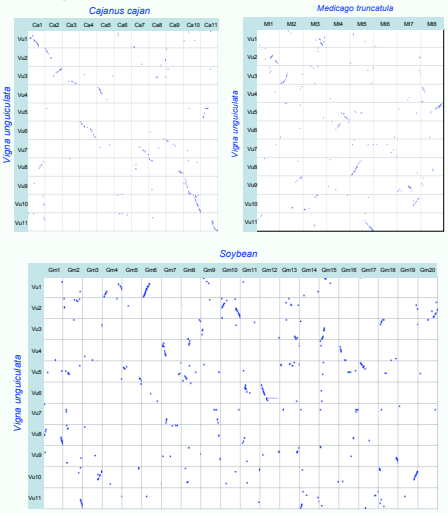
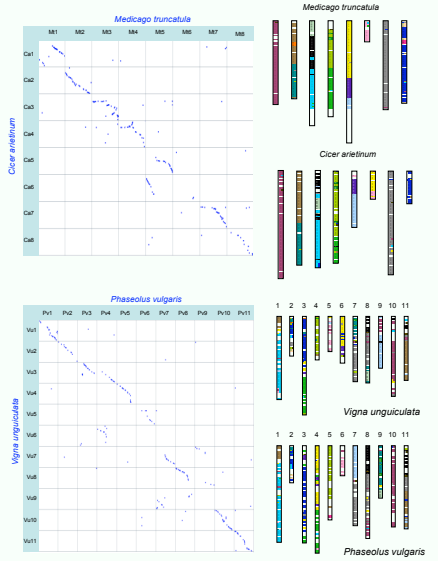
768-SNP OPA assay identifies polymorphism and distinguishes gene pools. Blue=BAT93 allele; tan=JALO allele.

SNP genotyping and linkage map analysis

1. Individual species maps



2. Binary dot-plot comparisons



3. Distribution of ancestral linkage blocks